NZHTA TECHNICAL BRIEF
January 2007
Volume 6 Number 1

Surveillance of Women at High Risk of Breast Cancer

Emma Davidson
Sarah Hancock

New Zealand
Health Technology Assessment
Department of Public Health and General Practice
Christchurch School of Medicine and Health Sciences
Christchurch, New Zealand

NEW ZEALAND HEALTH TECHNOLOGY ASSESSMENT (NZHTA)

Department of Public Health and General Practice
Christchurch School of Medicine and Health Sciences
Christchurch, New Zealand

Surveillance of Women at High Risk of Breast Cancer

Emma Davidson
Sarah Hancock

NZHTA TECHNICAL BRIEF
January 2007 Volume 6 Number 1

This report should be referenced as follows:

Davidson, E. Hancock, S. Surveillance of Women at High Risk of Breast Cancer. *NZHTA Report* 2007 **6(1)**.

Titles in this series can be found on the NZHTA website:
http://nzhta.chmeds.ac.nz/publications/

2007 New Zealand Health Technology Assessment (NZHTA)

ISBN 1-877235-95-4 (Print)
ISBN 1-877235-96-2 (Web)
ISSN 1174-5142

CONTRIBUTION BY AUTHORS

This report was authored by Mrs Sarah Hancock and Dr Emma Davidson, who conducted the critical appraisals and prepared the report.

ACKNOWLEDGEMENTS

Dr Rob Weir (Director) peer reviewed the final draft. Ms Susan Bidwell (Information Specialist) developed and undertook the search strategy and coordinated retrieval of documents. Mrs Kay Hodgson assisted with retrieval of documents. Miss Catherine Turnbull (Administrator) provided document formatting. Carol Webb provided sub-editing.

We acknowledge the contribution of Professor Anne Richardson, Chair of Public Health, Christchurch School of Medicine and Health Sciences, who undertook an external peer review of a late draft and provided valuable comments on the report.

We would like to extend grateful thanks to our colleagues from the International Network of Agencies for Health Technology Assessment (INAHTA), Bert Boer, Marli van Amsterdam and M Sanchez van Ginkel of the Dutch Health Insurance Board (CVZ) for their help in supplying a number of useful reports, and Professor Ard den Heeten of the Dutch National Expert and Training Centre for Breast Cancer Screening for sending us the thesis by Rian Rijnsburger.

The Canterbury Medical Library assisted with the retrieval of articles.

NZHTA is a Research Unit of the University of Otago funded under contract to the New Zealand Ministry of Health.

This report was commissioned by Madeleine Wall, Clinical Leader BreastScreen Aotearoa, National Screening Unit of New Zealand's Ministry of Health.

DISCLAIMER

New Zealand Health Technology Assessment (NZHTA) takes great care to ensure the information supplied within the project timeframe is accurate, but neither NZHTA, the University of Otago, nor the contributors involved can accept responsibility for any errors or omissions. The reader should always consult the original database from which each abstract is derived, along with the original articles, before making decisions based on a document or abstract. All responsibility for action based on any information in this report rests with the reader. NZHTA and the University of Otago accept no liability for any loss of whatever kind, or damage, arising from reliance in whole or part, by any person, corporate or natural, on the contents of this report. This document is not intended as personal health advice. People seeking individual medical advice are referred to their physician. The views expressed in this report are those of NZHTA and do not necessarily represent those of the University of Otago or the New Zealand Ministry of Health.

This review was commissioned by Madeleine Wall on behalf of the New Zealand Ministry of Health. NZHTA is a Research Unit of the University of Otago and is funded under contract by the New Zealand Ministry of Health.

COPYRIGHT

This work is copyright. Apart from any use as permitted under the *Copyright Act 1994* no part may be reproduced by any process without written permission from New Zealand Health Technology Assessment. Requests and inquiries concerning reproduction and rights should be directed to the Director, New Zealand Health Technology Assessment, Christchurch School of Medicine and Health Sciences, P O Box 4345, Christchurch, New Zealand.

CONTACT DETAILS

New Zealand Health Technology Assessment (NZHTA)
Department of Public Health and General Practice
Christchurch School of Medicine and Health Sciences
PO Box 4345
Christchurch
New Zealand
Tel: +64 3 364 3696 Fax: +64 3 364 3697

Email: nzhta@chmeds.ac.nz
Web Site: http://nzhta.chmeds.ac.nz/

EXECUTIVE SUMMARY

Objective

This systematic review identified and appraised the international evidence for surveillance of women at high risk of breast cancer. The accuracy and health outcome of the following modalities of surveillance were assessed in comparison to normal care: mammography (XRM), ultrasound (US) and magnetic resonance imaging (MRI).

Data sources

Medline, Embase and Current Contents were searched for both primary studies and systematic reviews/meta-analyses. In New Zealand, databases were accessed from the National Bibliographic Database, Ministry of Health website and library, university and medical library catalogues and the NZHTA in-house collection. Relevant publications referenced in material obtained in the course of research on the topic were also identified. Searches were limited to English language material from 1996 to June 2006 inclusive. Full information on the data sources and search strategies is given in Chapter 2 and Appendices 1 and 2.

Selection criteria

Studies were included if they were of women at a high risk of breast cancer and had at least 20 participants who underwent XRM, US or MRI, or combinations of these interventions. High-risk women were defined as those with a family history of breast cancer, including women with and without known genetic mutations which predispose to breast cancer. The estimated lifetime risk of the women in each study was documented, where possible, as was the method by which this risk was calculated. The comparison was stipulated as usual care, including clinical breast examination (CBE), or a single test if a combination was being evaluated and the outcomes were predetermined as measures of test performance (sensitivity, specificity positive and negative predictive values) or health outcome (breast cancer related mortality, cancer detection rate, tumour stage, node status, and interval cancers). Systematic reviews were preferentially included and observational designs, for the time period covered by key systematic reviews, were only included if the systematic reviews were not of adequate quality.

Excluded studies included non-systematic reviews, letters, editorials, expert opinion articles, superseded publications, conference proceedings, comments and articles published in abstract form. A technical exclusion related to studies of ultrasonography which used water baths or frequency probes with a resolution of less than 7.5 MHz.

Of more than 2780 articles identified by the search strategy, 156 articles were retrieved as full text from which a final group of 34 primary data papers and four systematic reviews were identified as eligible for appraisal and inclusion in the review.

Data extraction and synthesis

A systematic method of literature searching, selection and appraisal was employed in the preparation of this report. Level of evidence was assigned using the National Health and Medical Research Council levels of evidence (2000). However, in areas where high levels of evidence did not exist, an alternative system was applied which is outlined in the methods section.

Studies were appraised based on study design and evidence tables were developed describing the key aspects and limitations of each study included in the review.

Key results and conclusions

The following conclusions are based on the current evidence available from this report's critical appraisal of literature published on the surveillance of women at high risk of breast cancer.

Accuracy and efficacy of mammographic surveillance in women at high risk of breast cancer

One systematic review and 24 primary studies were identified which looked at the accuracy and efficacy of mammographic surveillance in women at high risk of breast cancer. There was no evidence from randomised controlled trials. Considerable heterogeneity was found between the studies in terms of the surveillance conducted, the level of risk of the participants, the age groups included and the inclusion or exclusion of women with a past history of breast cancer. The studies were frequently limited by the small number of participants and the relatively few tumours that arose during the study period. The heterogeneity between studies prevented any meta-analysis of the results.

There were three principal conclusions from these studies. The first was that mammographic surveillance had a higher cancer detection rate and was more accurate than surveillance with clinical breast examination alone. Two studies disagreed with this finding, but the results of these studies were unreliable due to the method of analysis and the small sample size. The accuracy (sensitivity) of mammography was shown to decrease as the risk status of women under surveillance increased.

The second conclusion was that cancer detection rates from the population receiving mammographic surveillance were equivalent to, or greater than, those of established breast screening programmes (BSPs) for women of all risk groups over the age of 50 years. Eight studies demonstrated similar rates of detection: six to the British BSP, one to the Italian BSP and one to the Dutch BSP. The theory behind this comparison was that if surveillance in women at high risk of breast cancer detected cancers at an equivalent rate to established BSPs then surveillance should be equally acceptable to adopt. However, these populations are not directly comparable and this assumption does not consider the potential harms of conducting mammographic surveillance in younger, high-risk women.

The third conclusion was that the characteristics of tumours detected in the population receiving mammographic surveillance were more favourable than those of tumours arising sporadically in women not under surveillance. Evidence from two studies demonstrated a significantly higher proportion of *in situ* tumours in the surveillance population compared to the population without surveillance. However, this could potentially represent over-diagnosis, of lesions that would never have been diagnosed in the women's lifetime without surveillance, rather than early diagnosis. One study also provided evidence of a significantly higher proportion of tumours with a good prognostic index in the surveillance group compared with the population without surveillance. The assumption behind this comparison is that detecting tumours at an earlier stage can lead to early treatment and that this may translate to a decrease in mortality. However, the natural history of tumours in high-risk women, and their response to treatment, may differ from tumours in women at average risk, i.e. over 50 years, in whom early detection and treatment has been proven to reduce mortality. There were very few studies presenting evidence on the outcomes of survival or mortality. In interpreting these studies, consideration needs to be given to whether any demonstrated survival advantage may be a product of lead-time or length bias. Evidence from one study suggested a significant decrease in mortality associated with the surveillance of women at high risk of breast cancer. However, this evidence was unreliable due to the short period of follow-up and the small numbers involved in the study.

Accuracy and efficacy of ultrasound surveillance in women at high risk of breast cancer

Nine studies were identified of relevance to the accuracy and efficacy of ultrasound surveillance of women at high risk of breast cancer. There was no evidence from randomised controlled trials (RCTs). Four studies compared ultrasound to clinical breast examination and all of them compared ultrasound to mammography. The evidence shows ultrasound surveillance to be more accurate and effective in detecting early breast cancer in women at high risk than clinical breast examination alone. However, the sensitivity of ultrasound surveillance was still relatively low, suggesting that ultrasound and clinical breast examination, individually or in combination, are not adequate for the surveillance of women at high risk of breast cancer.

The evidence showed ultrasound surveillance to have an equivalent sensitivity to mammographic surveillance in women at high risk of breast cancer. The sensitivity of ultrasound, like mammography, was shown to decrease as the risk status of women under surveillance increased. There were more false positives generated by ultrasound surveillance than mammographic surveillance. This would lead to anxiety and a higher rate of invasive investigations. Due to this, ultrasound may remain a diagnostic tool and other modalities of surveillance, if available and affordable, may be required in women at high

risk of breast cancer. There was no evidence on the outcomes of survival or mortality related to ultrasound, whether alone or in combination with mammography, in the surveillance of women at high risk of breast cancer.

Accuracy and efficacy of MRI surveillance in women at high risk of breast cancer

Two systematic reviews and 10 primary studies were identified of relevance to the accuracy and efficacy of MRI surveillance of women at high risk of breast cancer. There was no evidence from randomised controlled trials. Four studies compared MRI to clinical breast examination and all of them compared MRI to mammography. The evidence shows surveillance with MRI is superior to clinical breast examination for the early detection of breast cancer in women at high risk. Surveillance with MRI also appears to be associated with substantially higher sensitivity than mammography in detecting cancers in women at high risk of breast cancer. The difference in sensitivity between MRI and mammography is particularly pronounced in the women at highest risk, because they are genetic mutation carriers, in the studies where this has been examined. However, all of these results are based on a relatively small numbers of cancers detected and should be interpreted with some caution.

The specificity of MRI was relatively high, although in most cases lower than that for mammography. The higher number of false positives has implications for resource use and unnecessary anxiety of those undergoing surveillance. However, there appeared to be a learning effect demonstrated whereby those reading the MRI scans become more skilled as a result of increased experience and the availability of previous films for comparison, resulting in a decrease in false-positive results over the course of the study period.

Two studies examined MRI alone compared to MRI combined with mammography. The sensitivities were high in both studies, suggesting that each surveillance regimen is efficacious for the early detection of tumours in women at high risk of breast cancer. However, the evidence was inconclusive regarding combined surveillance regimen, and whether this offers any additional benefit over surveillance with MRI alone.

There was no evidence on the outcomes of survival or mortality related to MRI alone or in combination with mammography in the surveillance of women at high risk of breast cancer.

Accuracy and efficacy of combination surveillance in women at high risk of breast cancer

One systematic review and four primary studies were identified of relevance to the accuracy and efficacy of combination surveillance of women at high risk of breast cancer. There was no evidence from randomised controlled trials. The evidence shows that surveillance with MRI in women at high risk of breast cancer is superior to mammography, ultrasound or mammography and ultrasound combined. It is reinforced that MRI appears to be especially advantageous in women at the highest risk (mutation carriers) as, unlike mammography and ultrasound, and their combination, the sensitivity of MRI does not decrease with increased risk status. However, as discussed for all studies of MRI, these results are based on a very small number of cancers detected and this reduces their reliability.

The higher number of false-positive examinations in surveillance strategies using MRI is demonstrated in one of these studies. As discussed previously, breast surveillance with MRI is still early in its development and as radiologists gain experience and increase the number of breast MRIs they are reading, the number of false positives may decrease substantially. Further research is required to determine whether this will be the case.

In conclusion, MRI alone or in combination with other surveillance modalities appears to be a promising strategy for the surveillance of women at high risk of breast cancer. However, there is no evidence currently to suggest that such surveillance will necessarily translate to a decrease in mortality among this population. More research with larger numbers of participants and longer follow-up is required to truly assess the performance of MRI and combination strategies for the surveillance of women at high risk of breast cancer. In addition to its accuracy, MRI has the advantage of not using ionising radiation. The drawbacks of MRI are primarily related to the potential harm of false-positive diagnoses, cost and availability. If the introduction of a surveillance strategy for women at high risk of breast cancer with MRI was to be contemplated, a more complete assessment would need to be carried out. This should include the potential benefit from surveillance versus the potential physical and

psychological harm caused by the test, diagnostic procedures and treatment; the health care system being capable of supporting all the necessary elements of the surveillance pathway, including diagnosis, follow-up and evaluation; consideration of social and ethical issues and consideration of cost-benefit issues.

Search strategy

MeSH headings

Medical subject headings (MeSH) used were: breast neoplasms, mammography, ultrasonography, magnetic resonance imaging, predictive value of tests, sensitivity and specificity and genetic predisposition to disease.

Additional key words

These were supplemented with free text to expand these concepts and ensure full coverage of variations in vocabulary and indexing between databases. Fuller details of the search terms are given in the main body of the report and the complete search strategies are described in Appendix 1.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS AND ACRONYMS

95% CI	-	95 percent confidence interval
AJCC	-	American Joint Committee on Cancer
AUC	-	area under the curve
BRCA1	-	breast cancer gene 1
BRCA2	-	breast cancer gene 2
BIRADS	-	breast imaging reporting and data system
BSE	-	breast self examination
BSO	-	bilateral salpingo-oophorectomy
BSP	-	breast screening programme
CBE	-	clinical breast examination
CI	-	confidence intervals
DCIS	-	ductal carcinoma *in situ*
DNA	-	deoxyribonucleic acid
ESR	-	Environmental Science and Research
FBC	-	familial breast cancer
FNA	-	fine needle aspiration
HBC	-	hereditary breast cancer
HBOC	-	hereditary breast and ovarian cancer
HRT	-	hormone replacement therapy
IUCC	-	International Union Against Cancer
INAHTA	-	The International Network of Agencies for Health Technology Assessment
LCIS	-	lobular carcinoma *in situ*
MeSH	-	medical subject headings
MHz	-	megahertz
MOH	-	Ministry of Health (NZ)
MRI	-	magnetic resonance imaging
NHSBSP	-	National Health Service Breast Screening Programme (UK)

NICE	-	National Institute for Clinical Excellence (UK)
NPI	-	Nottingham Prognostic Index
NPV	-	negative predictive value
NSU	-	National Screening Unit (NZ)
NZ	-	New Zealand
NZHTA	-	New Zealand Health Technology Assessment
OBSP	-	Ontario Breast Screening Programme (Canada)
OCP	-	oral contraceptive pill
OR	-	odds ratio
PICO		Population. Intervention. Comparison. Outcome
PPV	-	positive predictive value
PTEN	-	phosphatase and tensin homolog
RCT	-	randomised controlled trial
ROC	-	receiver operator characteristic
RR	-	relative risk
SD	-	standard deviation
Se	-	sensitivity
Sp	-	specificity
TNM	-	tumour node metastasis
TP53	-	tumour protein 53
UK	-	United Kingdom
USA	-	United States of America
US	-	ultrasound
USS	-	ultrasound scan
XRM	-	x-ray mammography

GLOSSARY

Applicability (synonyms: external validity, generalisability, relevance, transferability) – The degree to which the results of an observation, study or review hold true in other settings.

Asymptomatic – Asymptomatic people are those who do *not* have one or more symptoms, eg skin changes, which may be due to a disease, eg breast cancer.

Attrition bias – Systematic differences between comparison groups in withdrawals or exclusions of participants from the results of a study. For example, patients may drop-out of a study because of the side effects of the intervention. Excluding those patients from the analysis could result in an overestimate of the effectiveness of the intervention.

Biannual – Something that happens twice a year

Biennial – Something that happens every two years

Bias – Deviation of results or inferences from the truth, or process leading to such deviation. Any trend in the collection, analysis, interpretation, publication or review of data that leads to conclusions that are systematically different from the truth.

Biopsy – In a breast biopsy, a small sample of breast tissue is removed and examined under a microscope as an aid to diagnosis.

Blinding (synonym: masking) – Keeping secret group assignment, eg to treatment or control, from the study participants or investigators. Blinding is used to protect against the possibility that knowledge of assignment may affect patient response to treatment, provider behaviours, i.e. performance bias, or outcome assessment, i.e. detection bias such as radiologists' interpretation of mammograms being biased by knowledge of a woman's risk status for breast cancer. Blinding is not always practical, eg when comparing surgery to drug treatment. The importance of blinding depends on how objective the outcome measure is.

Cancer – A general term for a large number of diseases that all display uncontrolled growth and spread of abnormal cells, also called malignant tumours. Cancer cells have the ability to continue to grow, invade and destroy surrounding tissue then leave the original site and travel via the lymph or blood systems to other parts of the body where they may establish further cancerous tumours.

Cancer detection rate – Number of women detected with cancer during a screening or surveillance episode, reported per 1,000 women screened or per 1,000 screens. When reported by the total number of women screened rather than the total number of screens undergone by these women, the results are not comparable across studies. This is because studies have different lengths of screening intervals and lengths of follow-up.

Case-control study – An epidemiological study involving the observation of *cases* (persons with the disease, such as breast cancer) and a suitable *control* (comparison, reference) group of persons without the disease. The relationship of an attribute to the disease is examined by comparing the past history of the people in the two groups with regard to how frequently the attribute is present.

Cohort study – An epidemiological study in which subsets of a defined population can be identified who are, have been, or in the future may be exposed or not exposed, in different degrees, to a factor or factors hypothesised to influence the probability of occurrence of a given disease or other outcome. Studies usually involve the observation of either a large population, or for a prolonged period, i.e. years, or both.

Confidence interval (CI) – The computed interval with a given probability, eg 95 per cent, that the true value of a variable such as a mean, proportion or rate is contained within the interval. The 95 per cent CI is the range of values in which it is 95 per cent certain that the true value lies for the whole population.

Confounder – A third variable that indirectly distorts the relationship between two other variables because it is independently associated with each of the variables.

Confounding – A situation in which the measure of the effect of an exposure on risk is distorted because of the association of exposure with other factor(s) that influence the outcome under study.

Control – In clinical trials comparing two or more interventions, a control is a person in the comparison group who receives a placebo, no intervention, usual care or another form of care. In case-control studies a control is the person in the comparison group without the disease or outcome of interest.

Critical appraisal – The process of assessing and interpreting evidence by systematically considering its validity, results and relevance.

Cross-sectional study – A study that examines the relationship between diseases, or other health related characteristics, and other variables of interest as they exist in a defined population at one particular time.

Diagnosis – The process of identifying a disease by its characteristic signs, symptoms and findings on investigation.

Diagnostic test efficacy – The impact and usefulness of a diagnostic test expressed in terms of its technical properties.

Efficacy – The extent to which an intervention produces a beneficial result under ideal conditions.

Epidemiology – The study of the distribution and determinants of health-related states or events in specified populations.

Equipoise – The principle of Equipoise involves the ethical treatment of human subjects in experimental conditions. A subject should only be submitted to a randomized, controlled design if there is substantial uncertainty about which of the treatments would benefit the subject most.

Evidence-based – Based on valid empirical information.

Evidence table – A summary display of selected characteristics, eg methodological design or results of studies of a particular intervention or health problem.

External peer reviewer – A person with relevant content, methodological or user expertise who critically examines reviews in her/his area of expertise.

External validity (synonyms: applicability, generalisability, relevance, transferability) – The degree to which the results of an observation hold true in other settings.

False-negative result – A negative test result obtained in a person who *does* have the condition being tested for.

False-positive result – A positive test result in a person who *does not* have the condition being tested for.

Fine needle aspiration (FNA) - A diagnostic process involving the insertion of a needle into a mass and the extraction of cellular material into a syringe. The needle is moved in a to-and-fro fashion, obtaining enough material for microscopic diagnosis. This procedure is generally accurate and frequently prevents the patient from having an open, surgical biopsy, which is more painful and costly.

Follow-up – The ascertainment of outcomes of an intervention at one or more stated times after the intervention has ended.

Generalisability (synonyms: applicability, external validity, relevance, transferability) – The degree to which the results of a study or systematic review can be extrapolated to other circumstances, in particular to routine healthcare situations.

Gold standard – The method, procedure or measurement that is widely accepted as being the best available, against which new interventions should be compared. It is particularly important in studies of the accuracy of diagnostic tests.

Grey literature – That which is produced at all levels of government, academia, business and industry, in print and electronic formats, but which is not controlled by commercial publishers.

Heterogeneity – Used in a general sense to describe the variation in, or diversity of, participants, interventions, and measurement of outcomes across a set of studies, or the variation in internal validity of those studies. Used specifically as statistical heterogeneity to describe the degree of variation in the effect estimates from a set of studies. Also used to indicate the presence of variability among studies beyond the amount expected due solely to the play of chance.

Heterogenous – Used to describe a set of studies or participants with sizeable heterogeneity. The opposite of homogeneous.

High risk groups – Usually refers to groups that have been identified as having a higher than average incidence of the disease in question.

Histology – The microscopic study of the minute structure and composition of tissues.

Homogenous – Used in a general sense to mean that the participants, interventions and measurement of outcomes are similar across a set of studies. Used specifically to describe the effect estimates from a set of studies where they do not vary more than would be expected by chance.

Incidence – the number of new events or cases, such as of disease, occurring during a certain period, in a specified population.

Inpatient – A patient who is formally admitted to a healthcare facility.

Intermediary outcomes – *See surrogate endpoints*.

Internal validity – The extent to which the design and conduct of a study are likely to have prevented bias. Variation in quality can explain variation in the results of studies included in a systematic review. More rigorously designed, i.e. better quality trials are more likely to yield results that are closer to the truth.

Inter-rater reliability (synonym: inter-observer reliability) – The degree of stability exhibited when a measurement is repeated under identical conditions by different raters. Reliability refers to the degree to which the results obtained by a measurement procedure can be replicated. Lack of inter-rater reliability may arise from divergences between observers or instability of the attribute being measured.

Intervention – The process of intervening on people, groups, entities or objects in an experimental study. In controlled trials, the word is sometimes used to describe the regimens in all comparison groups, including placebo and no-treatment arms.

Intervention group – A group of participants in a study receiving a particular healthcare intervention.

Lead-time bias – Lead-time refers to the period of time between the detection of a medical condition by surveillance and when it would ordinarily be diagnosed by symptoms. Bias can occur as surveillance prolongs the measurable disease time but not actual survival time, in the absence of effective treatment. This needs to be taken into account in studies of surveillance which use early detection of tumours or survival as an outcome.

Length bias – Length bias occurs as slow-growing lesions are diagnosed by surveillance in excess of those diagnosed by symptoms. The effect is that a greater number of slow-growing cancers, i.e. those with better prognoses, are diagnosed by surveillance with an apparent effect of prolonged survival time. This needs to be taken into account in studies of surveillance which use early detection of tumours or survival as an outcome.

Loss to follow-up – *See attrition*.

Matching – The process of making a study group and a comparison group comparable with respect to extraneous factors.

Mean – Calculated by adding all the individual values in the group and dividing by the number of values in the group.

Median – Any value that divides the probability distribution of a random variable in half. For a finite population or sample the median is the middle value of an odd number of values arranged in ascending order, or any value between the two middle values of an even number of values.

Medical Subject Headings (MeSH) – Terms used by the United States National Library of Medicine to index articles in Index Medicus and MEDLINE. The MeSH system has a tree structure in which broad subject terms branch into a series of progressively narrower subject terms.

Meta-analysis – The process of using statistical methods to combine the results of different studies. The systematic and organised evaluation of a problem using information from a number of independent studies of the problem.

Misclassification – The erroneous classification of an individual, a value, or an attribute into a category other than that to which it should be assigned.

Morbidity – Illness or harm.

Mortality – Death.

Natural history – The course of a disease from onset to resolution.

Negative predictive value (NPV) – The probability a person does not have the disease when the test is negative.

Odds ratio (OR) – A measure of the degree or strength of an association. In a case-control or cross-sectional study, it is measured as the ratio of the odds of exposure or disease among the cases to the odds among the controls.

Outcome – A component of a participant's clinical and functional status, after an intervention has been applied, that is used to assess the effectiveness of an intervention. *See also primary outcome, secondary outcome.*

Outpatient – A person who goes to a healthcare facility for a consultation and who leaves the facility within three hours of the start of the consultation. An outpatient is not formally admitted to the facility.

Population screening programmes – Population screening programmes involve screening entire populations or a large and easily identifiable group within a population. The target population group for screening may be defined geographically or by other characteristics such as gender, age or ethnicity. The New Zealand breast screening programme is an example of a population screening programme.

Positive predictive value (PPV) – The probability that a person actually has a disease, when the screening test is positive.

Power – The ability of a study to demonstrate an association if one exists.

Prevalence – Point prevalence is the number of events in a given population at a designated time while period prevalence is the number of events in a given population during a specified period.

Primary care – First contact, continuous, comprehensive and coordinated care provided to individuals and populations undifferentiated by age, gender, disease or organ system.

Prospective study - In evaluations of the effects of healthcare interventions, a study in which people are identified according to current risk status or exposure, and followed forwards through time to observe outcomes. Randomised controlled trials are always prospective studies. Cohort studies are commonly either prospective or retrospective, whereas case-control studies are usually retrospective. In epidemiology, prospective study is sometimes misused as a synonym for cohort study. *See also retrospective study.*

Protocol – The plan or set of steps to be followed in a study.

P-value – The probability, ranging from zero to one, which the results observed in a study, or results more extreme, could have occurred by chance if in reality the null hypothesis was true.

Random sample – A sample that is arrived at by selecting sample units such that each possible unit has a fixed and determinate probability of selection.

Randomised controlled trial (RCT) – An epidemiological experiment in which subjects in a population are randomly allocated into groups to receive or not receive an experimental preventative or therapeutic procedure, manoeuvre, or intervention. Randomised controlled trials are generally regarded as the most scientifically rigorous method of hypothesis testing available in epidemiology.

Reference standard – An independently applied test that is compared to a screening or diagnostic test being evaluated in order to verify the latter's accuracy. A reference standard, therefore, provides an accurate or 'truth' diagnosis for the verification of positive and negative diagnoses. It is sometimes described as providing 'final truth determination'.

Relative risk (RR) – The ratio of the risk of disease or death in those exposed to the risk compared to the risk among those unexposed. It is a measure of the strength or degree of association applicable to cohort studies and RCTs.

Reliability – The degree to which results obtained by a measurement procedure can be replicated. Lack of reliability can arise from divergences between observers or measurement instruments, measurement error or instability in the attribute being measured.

Retrospective study – A study in which the outcomes have occurred to the participants before the study commenced. Case-control studies are usually retrospective, cohort studies sometimes are, randomised controlled trials never are. *See also prospective study.*

Risk factor – An exposure or aspect of personal behaviour or lifestyle, which on the basis of epidemiologic evidence is associated with a health related condition.

Screening – Screening is the examination of asymptomatic people in order to classify them as likely or unlikely to have the disease that is the object of screening. The aim of screening is to detect disease before it is clinically apparent, and for this to improve the outcome for people with the disease.

Secondary care – Surgical or medical services that are generally provided in a hospital setting. In many cases, access to these services is by referral from a primary-care health professional such as a general practitioner.

Selection bias – Any error in selecting a study population that results in the people who are selected to participate in a study not being representative of the reference population or, in analytic studies, the comparison groups are not comparable.

Sensitivity (Se) – Sensitivity is the proportion of truly diseased persons in a screened population who are identified as diseased by a screening test. Sensitivity is a measure of the probability of correctly diagnosing a case, or the probability that any given case will be identified by the test.

Specificity (Sp) – Specificity is the proportion of truly non-diseased persons who are so identified by a screening test. It is a measure of the probability of correctly identifying a non-diseased person with a screening test.

Statistical power – *See power*.

Statistically significant – A result that is unlikely to have happened by chance. The usual threshold for this judgement is that the results, or more extreme results, would occur by chance with a probability of less than 0.05 if the null hypothesis was true. Statistical tests produce a p-value used to assess this.

Stratification – The process by which groups are separated into mutually exclusive sub-groups of the population that share a characteristic: e.g. age group, sex, or socioeconomic status. It is possible to compare these different strata to try and see if the effects of a treatment differ between the sub-groups. See also sub-group analysis.

Sub-group analysis – An analysis in which the intervention effect is evaluated in a defined subset of the participants in a trial, or in complementary subsets, such as by sex or in age categories. Trial sizes are generally too small for sub-group analyses to have adequate statistical power. Comparison of sub-groups should be by test of interaction rather than by comparison of p-values.

Surrogate endpoints (synonyms: intermediary outcomes, surrogate outcomes) – Outcome measures that are not of direct practical importance but are believed to reflect outcomes that are important; for example, blood pressure is not directly important to patients but it is often used as an outcome in clinical trials because it is a risk factor for stroke and heart attacks. Surrogate endpoints are often physiological or biochemical markers that can be relatively quickly and easily measured, and that are taken as being predictive of important clinical outcomes. They are often used when observation of clinical outcomes requires long follow-up.

Surrogate outcomes – See surrogate endpoints.

Surveillance – The monitoring of individuals known to have a disease or to be at increased risk of disease.

Surveillance, Epidemiology and End Results (SEER) registry – A set of geographically defined, population-based, central cancer registries in the United States, operated by local non-profit organisations under contract to the National Cancer Institute (NCI). Registry data are submitted electronically without personal identifiers to the NCI on a biannual basis, and the NCI makes these data available to the public for scientific research.

Symptomatic – Symptomatic people are those who have one or more symptoms, eg skin changes) that may be due to a disease, eg breast cancer).

Systematic review – Literature review reporting a systematic method to search for, identify and appraise a number of independent studies.

True negative – A test correctly identifies a person without the disease.

True positive – A test correctly identifies a person with the disease.

Tumour – An abnormal growth of tissue.

Chapter 1: Introduction

BACKGROUND

Need for the proposed systematic review

This systematic review was requested by Dr Madeleine Wall, Clinical Leader BreastScreen Aotearoa, National Screening Unit, Public Health Directorate, Ministry of Health, New Zealand.

The cancer control strategy and the New Zealand Health Strategy are the two overarching documents that support the development of New Zealand national guidelines for women at high risk of breast cancer. Any initiatives that support the early identification of women with breast cancer and the introduction of strategies to reduce the psychological, emotional and physical effects of cancer are highly valued.

The provision of surveillance for women at high risk of breast cancer is a high profile media topic and an area of interest for the current Minister of Health. Numerous health reports have been written and the number of ministerial inquiries is increasing as the level of awareness around breast cancer increases. The lack of appropriate guidelines for risk identification, surveillance and management of these women has been identified as a problem in attempting to implement policy.

There are few international guidelines exclusively addressing surveillance for women at high risk of breast cancer. The National Comprehensive Cancer Network in the USA recently published guidelines for hereditary breast and/or ovarian cancer (HBOC) which advised monthly breast self-examination (BSE) from 18 years, semi-annual clinical breast exam (CBE) from 25 years and annual XRM and MRI also from 25 years, or individualised based on the earliest age of onset in the family (Daly et al. 2006) An overview of the current clinical guidelines in Germany reported an intensified surveillance programme for women with a known breast cancer gene mutation or a lifetime risk of breast cancer of over 20 per cent (Kuschel et al. 2006). This advised frequent BSE, six-monthly CBE and US and annual MRI from 25 years, or five years earlier than the youngest affected family member, and annual XRM from 30 years. In the UK, a guideline was commissioned by the National Institute for Clinical Excellence for the classification and care of women at risk of familial breast cancer in primary, secondary and tertiary care (McIntosh et al. 2004). The recommendations were specific to the degree of breast cancer risk. Surveillance with XRM was not advised for women less than 30 years of age and, on the basis of the existing evidence, MRI and US were not recommended in routine surveillance. These guidelines were updated following the emergence of further evidence on surveillance with MRI in women at high risk of breast cancer (National Collaborating Centre for Primary Care 2006). A cost utility model was used in this work. It was recommended that annual MRI surveillance be provided for women aged 30 to 49 years with a known breast cancer gene mutation, or from 20 years if a TP53 carrier; for women aged 30 to 39 years at a 10-year risk greater than 20 per cent, and from 40 to 49 years in women at a 10-year risk greater than 20 per cent, or greater than 12 per cent where XRM is difficult due to dense breast tissue.

The requirement for national guidelines for women at high risk of breast cancer has been identified as a priority by the BreastScreen Aotearoa Advisory Group and the National Screening Unit (NSU) Consumers' Reference Group. The NSU supports the development of guidelines, which it recommended be undertaken by a representative group of interested stakeholders and not exclusively the NSU.

In order to underpin any guideline, an evidence-based review of the literature is required to identify the population at risk, as well as methods of identifying individual risk and methods of surveillance of those high-risk groups. In particular, an assessment of the relevance of international literature to the New Zealand population of women at high risk of breast cancer is required. This will inform the development of national guidelines or identify gaps in our knowledge, which necessitate further research before guideline development occurs.

This report contains the evidence-based review of literature examining surveillance of women at high risk of breast cancer.

Burden of disease from breast cancer in New Zealand

Breast cancer is the most common cause of cancer registration and cancer death among women in New Zealand. In 2002 there were 2,364 registrations for malignant neoplasms of the breast among women in New Zealand (New Zealand Health Information Service 2006a). During 2002-2003 there were 2,476 discharges from publicly funded hospitals for malignant neoplasms of the breast, with a mean stay of 9.9 days, and 304 discharges for carcinoma *in situ* of the breast, with a mean stay of 2.6 days among women (New Zealand Health Information Service 2006b). There were 625 deaths from malignant neoplasms of the breast among women in 2002 (New Zealand Health Information Service 2006a).

The number of registrations for malignant neoplasms of the breast generally increases with age, peaking in the early 50-year age group (New Zealand Health Information Service 2006a). This is likely to be partly an effect of screening, by biennial XRM, which has been provided by BreastScreen Aotearoa to all women between the ages of 50 and 64 years in New Zealand since 1998. The age range was extended to 45 to 69 years in 2004. In the group of women at high risk of developing breast cancer the mean age of onset of disease is significantly younger and the incidence of disease much higher than women at average risk. Therefore, additional surveillance needs to be considered for this group.

Women at high risk of breast cancer

Breast cancer is a multi-factorial disease. Known risk factors can be categorised as (McIntosh et al. 2004):

- population risks that all women are exposed to, such as increasing age;
- risks of sub-populations, based on family history;
- risks for each individual woman.

The protocol for this review did not define women at high risk of breast cancer. However the literature available focused on women with a family history of breast cancer, including women with and without known genetic mutations that predispose to breast cancer. All age groups were included and some studies included women with individual risk factors, such as a personal history of breast cancer.

To assess women for a family history of breast cancer it is recommended to obtain a three-generation family history (Daly 2004). The major features of an inherited predisposition are: early-age of onset of breast cancer in the index case and in relatives; multiple affected family members; bilateral cancers in paired organs; multiple primary tumours in an individual; specific cancer constellations, eg breast and ovary; and an autosomal pattern of inheritance (Frank and Critchfield 2001).

It has been estimated that up to 27 per cent of women may have an inherited predisposition to breast cancer (McIntosh et al. 2004). However, only some of the genes responsible have been identified. Known genetic mutations, which confer a very substantial increased risk of developing breast cancer, are thought to be carried in 3-5 per cent of women (Claus et al. 1994). BRCA1 and BRCA2 are two such genetic mutations, and are implicated in both breast and ovarian cancer. Carriers of these mutations reportedly have a lifetime cumulative risk of 50-85 per cent of developing breast cancer (Rijnsburger 2005). The prevalence of these genetic mutations varies between ethnic groups and is especially high in women of Ashkenazi Jewish descent. A statistical computer programme called BRCAPRO has recently been validated as an accurate tool for determining the probability of carrying BRCA1 or BRCA2 genetic mutations (Pichert et al. 2003). Some rarer mutations associated with breast cancer are TP53 and PTEN. Several uncommon genetic syndromes also confer an increased risk of breast cancer, including Li Fraumeni syndrome, Muir Torre syndrome, Peutz Jeghers syndrome, Cowden's disease and ataxia telangectasia.

Women with a clear family history of breast cancer, in whom an underlying genetic mutation has not yet been identified, are also at increased risk of neoplasia. The inability to detect a mutation may be due to there being no living family member affected with breast cancer to test, or because the family's mutation is, as yet, undiscovered. Therefore, a negative genetic test result does not rule out increased risk of breast cancer unless it is negative for a specific mutation which has been identified to run in the

family. Several empirical and statistical models have been designed to estimate the magnitude of the risk of breast cancer in individuals with a family history for breast cancer but no known mutation. Some of these focus on aspects of family history alone and others also incorporate individual risk factors. A review of these models (Claus 2001) suggests that two of the former type, developed by Claus et al. (1994) and Berry et al. (1997), are the most effective at estimating risk in women for whom there is no known mutation. These calculations primarily consider the age of the women in question, the number of first-degree and second-degree relatives affected, and the age of onset of breast cancer in any affected first-degree and second-degree relatives. A threshold is then set, usually a lifetime risk greater than 20 per cent (Kuhl et al. 2005a), over which women are considered to be at significantly increased risk compared with the general age-matched population.

This review includes women at high risk with both known and unidentified genetic mutations. For women with no known mutation, the risk stratification model and threshold used in each study is identified where possible.

It is unknown how many women in New Zealand comprise this sub-population at high risk of breast cancer. It has been estimated that for a total population of 1 million, with an age and sex structure comparable to that of England and Wales, there would be 20-40 families whose family history of breast cancer would indicate that members had a high risk of developing breast cancer (McIntosh et al. 2004). Estimates of population carrier frequency have been reported as 0.006 per cent for both BRCA1 and BRCA2 mutations (Pharoah et al. 1998). This would translate to approximately 230 women being carriers of either BRCA1 or BRCA2 mutations in the New Zealand 2001 census population of 1,914,273 women (Statistics New Zealand 2002). However, these data cannot be extrapolated to the New Zealand population because it differs demographically from the populations from which these estimates were generated.

For women at high risk of breast cancer, there are three options for managing their risk. These are:

- Chemoprevention – Tamoxifen;
- Surgical – bilateral mastectomy or bilateral salpingo-oophorectomy (BSO);
- Surveillance.

The focus of this review is on surveillance.

Surveillance

Surveillance, as opposed to screening, refers to monitoring individuals known to have a disease or to be at increased risk of disease. The NHC has developed criteria to inform the assessment of screening programmes in New Zealand, which are in **Table 1** (National Health Committee 2003). Similar issues need to be considered for any proposed surveillance programme. However, the benefit–to-risk ratio of surveillance as opposed to population screening is more favourable. This is because a greater proportion of a surveillance population is likely to benefit from monitoring due to the high prevalence of disease in this population. This population is also more willing to accept the risks associated with surveillance than would healthy, average-risk individuals (New Zealand Guidelines Group 2004). A discussion of some issues that need to be addressed in considering surveillance for women at high risk of breast cancer follows.

Table 1. Criteria for assessing screening programmes

1.	The condition is a suitable candidate for screening
2.	There is a suitable test
3.	There is an effective and assessable treatment or intervention for the condition identified through early detection
4.	There is high-quality evidence, ideally from randomised controlled trials, that a screening programme is effective in reducing mortality and morbidity
5.	The potential benefit from the screening programme should outweigh the potential physical and psychological harm caused by the test, diagnostic procedures and treatment
6.	The healthcare system will be capable of supporting all the necessary elements of the screening pathway, including diagnosis, follow-up and evaluation
7.	There is consideration of social and ethical issues
8.	There is consideration of cost-benefit issues

Surveillance for women at high risk of breast cancer

Condition is a suitable candidate for surveillance

The population of women at high risk of breast cancer may be relatively small but the likelihood of disease in these women is extremely high. Due to the early onset of disease, it results in considerable loss of quality-adjusted life years. The burden of this disease affects both the individuals at high risk and the greater community, and this renders it a suitable candidate for surveillance.

The primary objective of screening is to detect disease at an early stage, before it causes symptoms. The individual can then be treated to ameliorate or cure the disease (Markham et al. 1997). This is also the primary objective of surveillance and depends on the natural history of the disease allowing early detection, as well as treatment at an early stage improving the prognosis.

The natural history of breast cancer has two important phases, the *in situ* and the invasive phase. *In situ* carcinoma is confined within the epithelial compartment and can be successfully cured by surgical excision. Invasive carcinoma has breached the compartmental barrier of the epithelium and infiltrates the connective tissue of the breast. Although it is possible to eradicate invasive carcinoma from its primary site of growth, it has the potential to establish secondary deposits at distant sites, i.e. metastases. The treatment of metastatic breast cancer is palliative rather than curative (Fentiman and D'Arrigo 2004).

The American Joint Committee on Cancer (AJCC) and the International Union Against Cancer (IUCC) developed staging systems for breast cancer based on the tumour node metastasis (TNM) system which dates back to 1942. In 1987 the systems were aligned and the sixth edition of the AJCC cancer staging manual was released in 2002 (Singletary and Connolly 2006). The staging system refers to the size of the primary tumour (T), the presence and extent of regional lymph node involvement (N) and the presence of distant metastases (M). The AJCC has then grouped the TNM classifications in to Stages 0-4 (Greene et al. 2002). An abridged version of this is presented in **Table 2**.

Table 2. AJCC's stage grouping of the TNM classification for breast cancer (adapted from the sixth edition of the AJCC staging cancer manual, 2002)

AJCC Stage Grouping	Abridged Description
Stage 0	Tumour is *in situ* and no nodes involved or metastases
Stage 1	Tumour is invasive but ≤ 2 cm in greatest dimension and no nodes involved or metastases
Stage 2	Tumour is invasive and ≤ 5 cm in greatest dimension, including no evidence of a primary tumour, with or without ipsilateral axillary lymph nodes and no metastases OR Tumour is invasive and > 5 cm greatest dimension with no nodes involved or metastases
Stage 3	Tumour is of any size, including no evidence of a primary tumour, with movable or fixed ipsilateral axillary lymph nodes and no metastases OR Tumour is of any size with direct extension to the chest wall or skin with or without nodes and with no metastases OR Any tumour with spread to regional nodes other than the ipsilateral axillary nodes
Stage 4	Any tumour with any regional lymph node spread and distant metastases.

(Refer to URL: http://www.cancer.gov/cancertopics/pdq/treatment/breast/HealthProfessional/page3)

The breast cancer stage at diagnosis has been shown to directly relate to survival. This is shown for 10-year survival in Figure 1 (Bland et al. 1998). It is possible to detect *in situ* breast cancer, i.e. ductal carcinoma *in situ* (DCIS) and lobular carcinoma *in situ* (LCIS) with surveillance. Without this surveillance, these early-stage lesions would not be detectable as they are not palpable and would not present symptomatically. Therefore surveillance has the potential to improve breast cancer survival through early detection.

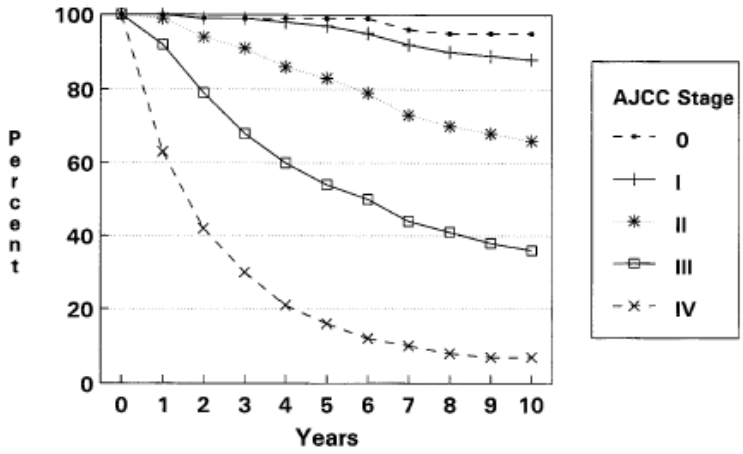

Figure 1 Relative survival for breast carcinoma patients according to AJCC stage group[1] (*Cancer* 1998 September 15; 83(6): 1262-73).

Suitable test

In considering whether a test is suitable for screening, the NHC advised consideration of the following test characteristics (National Health Committee 2003):

- Safe: harm is minimised.
- Simple: easy to perform and interpret.
- Reliable: the test is repeatable, and gives consistent results.
- Valid: the test is capable of measuring what it set out to measure.
- Highly sensitive: high probability that the test will give a positive result when the person being screened has the condition. Sensitivity should be sufficient to lead to a substantial impact on the disease from a population perspective.

[1] This material is reproduced with permission of Wiley-Liss, Inc., a subsidary of John Wiley & Sons Inc.

- Highly specific: high probability that the test will give a negative result when the person being screened is disease-free. This is important to minimise harm from false-positive screening results.

These test characteristics can also be applied for surveillance. However, the threshold for accepting a test as suitable may be altered by the higher benefit-to-risk ratio in a surveillance population.

Measuring diagnostic test accuracy using sensitivity (Se) and specificity (Sp) provides important information when considering its use. Traditional definitions of Se and Sp are usually based on evaluations of tests at one point in time, with a reference standard test used to determine the true disease. This ideal is not always achievable in a clinical setting. In verifying a diagnosis of breast cancer the best reference standard is pathological confirmation following biopsy or surgical excision. Ethically, this cannot be performed unless there is a suspicion of disease arising from surveillance. The only way to confirm the absence of disease in those that are negative is through follow-up over time. This means that most studies considering this issue are prone to verification bias which may underestimate false negative surveillance tests.

Predictive value measures are also frequently used as an additional method of evaluating diagnostic test validity, and can be particularly useful in clinical settings. Predictive values measure whether or not an individual actually has the disease, given the results of the test. The positive predictive value (PPV) is the probability that the person has a disease, given a positive test result, and the negative predictive value (NPV) is the probability that the person is free from disease, given a negative test result. It should be noted that the predictive measures will vary with the underlying prevalence of disease in the population. The PPV of a test will be much higher in a population of women at high prevalence of breast cancer compared to the general population. It is therefore necessary to evaluate the test's performance in study populations with the same risk of disease as the population of interest. This is straightforward for women identified as BRCA1 and BRCA2 mutation carriers, but less clear for those at high risk without proven mutations. In the latter group, the aim will be to identify the risk stratification model and threshold used in each study.

The mainstay of screening and surveillance for breast cancer has been clinical breast examination (CBE) and mammography (XRM). Breast self-examination (BSE) was at times advised, but has been found by randomised controlled trial (RCT) not to reduce breast cancer mortality and to increase unnecessary biopsies (Thomas et al. 2002). Latterly, other imaging modalities have been considered, particularly for women at high risk of breast cancer. Characteristics of these tests are described briefly below, along with the issues surrounding each test's use for surveillance in women at high risk of breast cancer.

Clinical breast examination

Clinical breast examination (CBE) is a systematic examination by a clinician of the four quadrants of each breast and the axillary areas. The performance of CBE alone is not suitable for surveillance as, although it is easy to perform, it is not capable of detecting *in situ* tumours or tumours that are under approximately 10mm in size (Hughes et al. 1999). This means that it may detect tumours before they present symptomatically but not as early as they are detectable by radiological imaging. Spratt et al. (1995; 1996) investigated doubling time in the growth of tumours and, based on this, it was estimated that the time between a tumour being detectable by XRM and by CBE would be approximately 3.4 years (Hughes et al. 1999). However, CBE is usually carried out in conjunction with radiological imaging and in some cases does detect tumours that were not identified by other means.

Mammography

Mammography (XRM) is an imaging modality that utilises low dose ionizing radiation to examine the breasts. The results of eight randomised controlled trials have established the ability of mammographic screening to improve outcome in breast cancer (Smith and Andreopoulou 2004). A decrease in mortality of up to 30 per cent has been shown through mammographic screening of women over the age of 50-69 years (Tabar et al. 2000). The evidence for women aged less than 50 years is less clear. A randomised controlled trial of mammographic screening of women from age 40 was carried out in the UK. An interim analysis using surrogate outcome measures to predict mortality was published in 2005 (Moss et al. 2005). This analysis suggested that a reduction in breast cancer mortality may be observed in the trial, but firm conclusions must await the analysis of observed mortality from breast cancer. Women at high risk require surveillance from an early age, often from 30 years or five years before the youngest affected relative if they were diagnosed under the age of 30. There is no RCT evidence for

mammographic surveillance of women at high risk, and from such a young age. It is unlikely that such evidence could be obtained. It would no longer be considered ethical to conduct an RCT as this would mean one arm of the study receiving no surveillance. Women at high risk are usually offered mammographic surveillance based on the evidence of its use as a screening tool in older women. However, it is known that XRM is less accurate in younger women due to the higher density of their breasts and also perhaps due to the phenotype of their tumours which have a smoother, more benign appearance on XRM (Hartman et al. 2004; Tilanus-Linthorst et al. 2002).

Women at high risk of breast cancer also require more frequent XRM as they are prone to more rapidly developing, biologically aggressive tumours. A disadvantage of XRM in women at high risk is its use of radiation. If surveillance commences at a young age and is repeated frequently they will be exposed to a high cumulative dose of radiation. This is increased by the need for recall XRM when breast density makes the images difficult to interpret. BRCA1 and BRCA2 mutation carriers are of particular concern. These mutations have been implicated in cell cycle regulation and DNA repair (Robson 2002), which means that carriers may be more susceptible to the mutagenic effects of low-dose radiation. These disadvantages suggest that it may be preferable to utilise other modalities of surveillance in women at high risk of breast cancer either in addition to, or instead of, XRM.

BIRADS is the Breast Imaging Reporting and Data System that was developed in 1993 by the American College of Radiology to standardise XRM reporting. It is used to define a cut-off between a 'normal' mammogram and an 'abnormal' mammogram requiring further investigation. It is important to establish the systems of classification and the thresholds used in studies of surveillance to assess consistency and determine whether the studies are comparable. The classification system is outlined in **Table 3** (Eberl et al. 2006). BIRADS classification has also been used with US and MRI, the other potential surveillance modalities which will be discussed.

Table 3. BIRADS system of classification for XRM images

BIRADS Category	Assessment
0	Assessment incomplete
1	Negative
2	Benign finding
3	Probably benign finding
4	Suspicious abnormality
5	Highly suspicious of malignancy
6	Known biopsy-proven malignancy

Ultrasonography

Ultrasound (US) imaging, also called sonography, obtains images through the use of high-frequency sound waves. The reflected sound wave echoes are recorded and displayed as real-time visual images. US has traditionally been used as a diagnostic tool for breast imaging rather than for screening or surveillance. An advantage of US, particularly for women at high risk, is that it does not use ionising radiation. It is also the simplest way to guide biopsies and is therefore a unique problem-solving tool (Rizzatto 2001). The disadvantages of US are that storing images for review is not as simple as other imaging modalities and its accuracy is extremely operator dependent, causing poor inter-observer reliability.

MRI

MRI uses the signal produced by hydrogen ions or protons placed in a powerful magnetic field stimulated by radio waves to produce images. Initially it was thought that MRI would not be a useful tool for imaging breast disease as the signal from breast cancer was very similar to that of normal fibroglandular tissue. However, in 1989 two publications reported that the use of intravenous contrast agent caused breast cancers to become rapidly enhanced and conspicuous against the normal background tissue (Heywang et al. 1989; Kaiser and Zeitler 1989). Subsequently, MRI has been used for locating occult primary breast tumours, pre-surgical and post-surgical assessment of breast tumours, and assessment of treatment response to systemic therapy. It has also been trialed in the surveillance of women at high risk of breast cancer (Bartella and Morris 2006).

The advantage of MRI is that, like US, it does not use ionising radiation. The only contraindications for its use are metal in the body (prosthetic hip, heart pacemaker, artificial heart valve, implanted port, infusion catheter, or any metal plates, pins, screws or surgical staples). A large body mass index (BMI) and claustrophobia may also cause difficulties. Disadvantages of MRI are that it is expensive and not readily available in all facilities. This is especially because MRI breast imaging requires specialised equipment, such as breast coils and facilities, to allow MRI-guided biopsy. These are all required in order to perform adequate surveillance, as is a radiologist with significant experience in breast MRI. As with any modality of surveillance, the accuracy depends heavily on the experience of the interpreting radiologist. The total number of breast MRI studies read each year in the USA is estimated to be 0.02 per cent of the screening XRMs which are read. This may well account for the lower measures of accuracy (PPVs) of MRI in some studies, and should improve over time as experience is gained (Robson 2004).

Effective treatment or intervention for the condition identified through early detection

Treatment for breast cancer depends on the stage and histology at diagnosis and usually involves surgery, with or without systemic therapy. Surgical management has shifted dramatically since the 1980s with an increase in breast conserving procedures (Bland et al. 1998). With early detection the surgery required is generally less radical and the overall outcome can be improved. As discussed, a decrease in mortality has been demonstrated through screen detection and treatment of breast cancer in women of all risk groups over 50 years of age.

Aim of review

The aim of this review is to identify the evidence for the surveillance of women at a high risk of breast cancer with the aforementioned modalities. As discussed in the case of XRM, there is unlikely to be RCT evidence due to the ethical considerations of withholding potentially beneficial surveillance from individuals at such high risk of disease. It was stated recently on this topic that "rigorous scientific evaluation is both ethical and essential to establish that a test does more good than harm, whether for the general population or for those with a greater risk of breast cancer" (Irwig et al. 2006). However, it is unlikely that such research would be considered acceptable by the population eligible to participate.

REVIEW SCOPE

This systematic review focuses on the surveillance of women at high risk of breast cancer. The definition of high risk refers to the subpopulation of women with a family history of breast cancer, including both those with and without identified genetic mutations. It was requested that the scope of the systematic review be limited to considering the following questions:

- Does mammography improve the accuracy of detecting breast cancer among women with high risk of breast cancer compared with usual care?
- Does ultrasound improve the accuracy of detecting breast cancer among women with high risk of breast cancer compared with usual care?
- Does MRI improve the accuracy of detecting breast cancer among women with high risk of breast cancer compared with usual care?
- Does mammography improve health outcomes among women with high risk of breast cancer compared with usual care?
- Does ultrasound improve health outcomes among women with high risk of breast cancer compared with usual care?
- Does MRI improve health outcomes among women with high risk of breast cancer compared with usual care?

The health outcomes of interest were breast cancer mortality, cancer detection rate (invasive disease and DCIS), tumour size, tumour stage, lymph node status and interval cancers.

The search was limited to full reports published in English and published between 1996 and July, 2006. Full details of inclusion and exclusion criteria are provided in the next chapter.

If this review suggests that a surveillance strategy would be accurate and effective in improving health outcomes for women at high risk of breast cancer, further issues that are not included in the scope of

this review certainly would require consideration. These include: any possible harm caused by the test, diagnostic procedures and treatment; that the healthcare system is capable of supporting such a surveillance programme; plus consideration of social, ethical and cost-benefit issues.

STRUCTURE OF REPORT

This report includes nine chapters. In **Chapter 2** the methods, including the search strategy, selection criteria, and study selection and appraisal methods, are detailed. **Chapter 3** examines the accuracy and efficacy of surveillance with XRM compared to CBE in women at high risk of breast cancer. The accuracy and effectiveness of surveillance with US is compared to CBE in **Chapter 4** and compared to XRM and XRM combined with US in **Chapter 5**. The accuracy and effectiveness of surveillance with MRI is compared to CBE in **Chapter 6** and compared to XRM and XRM combined with MRI in **Chapter 7**. **Chapter 8** considers the accuracy of surveillance with a combination of XRM, US and MRI. Each chapter presents a summary of relevant findings of secondary research, i.e. systematic reviews, where available, and the primary research is considered. Tables are provided within each chapter that present highly summarised and aggregated data from the studies selected and appraised. The corresponding, more detailed, evidence tables for the studies that were included in the review are found at the end of each chapter and present each appraised study's methods, results, limitations, and authors' conclusions**. Chapter 9** summarises results, briefly discusses methodological limitations in the area, and presents key conclusions.

Chapter 2: Methodology

SELECTION CRITERIA

Selection criteria for this systematic review are listed in **Table 4**.

Table 4. Inclusion/exclusion criteria for studies of surveillance of women at high risk of breast cancer

Characteristic	Criteria
Inclusion criteria	
Publication type	Clinical studies using human subjects
Patients	Asymptomatic women at high risk of breast cancer
Sample size	At least 20 human patients were tested by one of the interventions outlined below
Intervention/test	Mammography Ultrasound MRI Combinations of the above
Comparator	Usual care – including clinical breast examination but not mammography, ultrasound or MRI When the intervention is a combination test, comparison with a single test
Outcome	Breast cancer related mortality, cancer detection rate, tumour size, tumour stage, node status, interval cancers, measures of test performance (sensitivity, specificity, and positive and negative predictive values)
Exclusion criteria	
Publication type	Non-systematic reviews, letters, editorials, expert opinion articles, conference proceedings, comments and articles published in abstract form
Publication superseded	Publication superseded by subsequent publication with longer follow-up data and overlap in the patient population
Test	Studies of ultrasonography with water baths or frequency probes with a resolution less than 7.5mHz
Language	Non-English articles
Time period	Studies published before 1996

SEARCH STRATEGY

A systematic method of literature searching and selection was employed in the preparation of this review as follows:

Searches were limited to English language material published from 1996 onwards. The searches were completed on 20 July 2006. Correspondence and news items were exclusions in the search strategy if available in the individual database.

Principal sources of information

The following databases were searched using the search strategy outlined in **Appendix 1**:

Bibliographic databases

- Cochrane Central Register of Controlled Trials
- Current Contents
- Embase
- Medline
- Pubmed (last 60 days)

Review databases

- ACP Journal Club
- Cochrane Database of Systematic Reviews
- Database of Abstracts of Reviews of Effectiveness (DARE)
- Health Technology Assessment database
- NHS Economic Evaluation database

Hand searching of journals, contacting of manufacturers or contacting of authors for unpublished research was not undertaken in this review. A complete list of the sources searched for this review is given in **Appendix 2.**

Search terms used

- Index terms from Medline (MeSH terms): breast neoplasms, mammography, ultrasonography, magnetic resonance imaging, mass screening, genes-BRCA1, genes-BRCA2, family health, family, predictive value of tests, sensitivity and specificity, genetic predisposition to disease, genetics.fs[as a floated subheading].
- Index terms from Embase (where different from the MeSH terms): breast cancer, cancer screening, breast examination, echography, nuclear magnetic resonance imaging, genetic predisposition, genetic susceptibility, familial cancer, BRCA1 protein, BRCA2 protein, cancer genetics, genetics, diagnostic accuracy, diagnostic error, diagnostic value, intermethod comparison.
- The above index terms were used as keywords in databases where they were not available and in those databases without controlled vocabulary.
- Additional keywords (not standard index terms) were used in all databases: (high or increase$) adj2 risk), (high or increase$) adj2 rate), (high or increase$) adj2 incidence), first degree relative, family adj2 history, familial, screen$, surveillance, breast adj3 examination, positive predictive value, ppv, negative predictive value, npv, likelihood ratio$, false negative$, false positive$, diagnostic accuracy, interval cancer$.

Validated international filters for identifying studies with high-quality designs were applied to the results of the subject searches.

STUDY SELECTION

A two-stage process was used to select studies for appraisal. Initially, the titles and abstracts, where available, identified from the search strategy, were scanned and excluded as appropriate. The full-text articles were retrieved for the remaining studies and these were appraised if they fulfilled the study selection criteria outlined above.

The search strategy identified 2,780 studies. One hundred and fifty six full-text articles were obtained after excluding studies from the search titles and abstracts. A further 118 of these full-text articles did not fulfil the inclusion criteria and are presented in **Appendix 5**. Therefore, 38 articles were fully appraised and are included in this report, and presented in **Appendix 6**. Other cited publications, e.g. those providing background material, are presented in the **References**.

APPRAISAL OF STUDIES

The evaluation initially classified studies according to National Health and Medical Research Council (National Health and Medical Research Council 2000) levels of evidence criteria, to rank them in terms of quality according to a pre-determined 'evidence hierarchy' (see **Appendix 4**). These evidence levels are only a broad indicator of the quality of the research. The levels describe groups of research which are broadly associated with particular methodological limitations. However, these levels are only a general guide to quality because each study may be designed and/or conducted with particular strengths and weaknesses. High-level evidence is provided by a well-conducted randomised-controlled trial (RCT). In areas where high levels of evidence did not exist, an alternative method of appraisal was used, which is described in **Table 5**. NHMRC checklists of quality issues to consider in appraising research studies were also used relevant to study design.

Table 5. Grading system for the appraisal of included studies

Validity Criteria	Description	Grading System
Appropriate comparison	Did the study evaluate a direct comparison of the index test strategy versus the comparator test strategy?	C1 direct comparison CX other comparison
Applicable population	Did the study evaluate the index test in a population that is representative of the subject characteristics (age and sex) and clinical setting (disease prevalence, disease severity, referral filter and sequence of tests) for the clinical indication of interest?	P1 applicable P2 limited P3 different population
Quality of Study	Was the study designed to avoid bias? High quality = no potential for bias based on predefined criteria Medium quality = some potential for bias in areas other than those pre-specified as key criteria Poor quality = potential for bias based on key pre-specified criteria	Q1 high quality Q2 medium quality Q3 poor quality or insufficient information

Summaries of appraisal results are shown in tabular form as **Evidence Tables** and include:

- reference (authors, publication date) and country where study was principally conducted design
- evidence level (applying NHMRC criteria)
- study setting
- patient characteristics including number of patients for intervention and comparator groups;
- description of intervention and comparator
- patient inclusion and exclusion criteria
- analyses comparing intervention and comparator groups at baseline
- eligible outcome measures used and timing of follow-up intervals
- results of analyses comparing intervention and comparator groups on eligible outcomes, including statistically tested comparisons and reporting relevant statistical data
- authors' conclusions
- comments on the study's limitation relevant to its internal validity.

Conclusions are drawn based on the study design and the specific problems associated with individual studies.

Systematic reviews and meta-analyses are described and critiqued in terms of their search strategy, inclusion/exclusion criteria, data synthesis and interpretation

KEY OUTCOME MEASURES FOR PRIMARY STUDIES

There are two sets of key outcome measures for primary studies. These relate to the measure of test performance and health outcome and are listed below. The majority of studies are likely to focus on intermediate outcomes because of the time inherent in assessing mortality. Surveillance-related improvements in intermediate outcomes cannot be extrapolated to a decrease in mortality. However, intermediate outcomes do have the advantage of not being affected by improving treatment for breast cancer over time (Irwig et al. 2006).

Test performance

The accuracy of the test will be measured in terms of sensitivity and specificity, along with negative and positive predictive values. Methodology for calculating these is presented in **Appendix 3.**

Health outcome

Health outcome measures are breast cancer related mortality, cancer detection rate, tumour size, tumour stage, node status and interval cancers. The measure of interval cancers is considered crucial to assessing the impact of a surveillance strategy. A reduction in interval cancers represents the potential benefit of early detection rather than over-detection (Irwig et al. 2006).

Difficulties were encountered in this review with utilising cancer detection rate as an outcome. Some studies reported these as cancers per 1,000 surveillance rounds and others have used cancers per 1,000 women under surveillance. As the studies had different intervals between rounds of surveillance and ran for different lengths of time, the rates per 1,000 women under surveillance are not comparable between studies. In addition, some studies combined prevalent and incident surveillance results even though detection rates are generally lower for incident surveillance rounds than for prevalent surveillance rounds. All the available results are presented. However, it is emphasised that care must be taken in any comparisons considered between studies.

LIMITATIONS OF THE REVIEW

This study has used a structured approach to review the literature. However, this approach had some inherent limitations. Namely, systematic reviews are limited by the quality of the studies included in the review and the review's methodology.

This review has been limited by the restriction to studies published in English. Restriction by language may result in study bias, but the direction of this bias cannot be determined. In addition, the review has been limited to the published academic literature and has not appraised unpublished work. Restriction to the published literature is likely to lead to bias since the unpublished literature tends to consist of studies not identifying a significant result.

Papers published pre-1996 were not considered due to the relatively recent understanding of the familial risk of breast cancer and introduction of technologies under consideration, especially MRI.

The studies were initially selected by examining the abstracts of these articles. Therefore, it is possible that some studies were inappropriately excluded prior to examination of the full-text article. However, where detail was lacking or ambiguous, papers were retrieved as full text to minimise this possibility.

All of the studies included in this review were conducted outside New Zealand. Therefore, their generalisability to the New Zealand population and context may be limited and needs to be considered.

This review was confined to an examination of the effectiveness of the interventions and did not consider the acceptability or any ethical, economic or legal considerations associated with these interventions. Interventions were not assessed in terms of their impact on general quality of life.

Although two researchers appraised the articles included in this review, they worked on delineated subsections of the project and did not cross-validate the data extraction and appraisal process.

The review scope was developed with the assistance of Ministry of Health staff. It had the goal of providing information on the accuracy and health outcome of methods of surveillance for women at high risk of breast cancer.

This review was conducted over a limited timeframe (July 2006 to November 2006).

This review has greatly benefited from the advice provided by the consultant peer reviewer. However, it has not been exposed to wider peer review.

For a detailed description of interventions and evaluation methods, and results used in the studies appraised, the reader is referred to the original papers cited.

Chapter 3: Accuracy and efficacy of mammography

SECONDARY RESEARCH: STUDY DESIGNS AND QUALITY

The search strategy identified only one relevant review of the accuracy and efficacy of mammographic surveillance in women at high risk of breast cancer. This was carried out by consultants in a tertiary-care genetics service in the UK (Lucassen et al. 2001). The methods and conclusions are described in **Table 6**.

The inclusion and exclusion criteria set by Lucassen et al. were not explicit, and are not necessarily concordant with the criteria applied in this review. Therefore the results must be interpreted with caution. The articles included were published between January 1995 and December 2000. Reviews and non-English papers were excluded. Seven studies were included in total. No details were given of the studies' design, except that no randomised controlled trials were identified.

The sample sizes ranged from less than 100 to more than 25,000 women. Considerable heterogeneity was demonstrated between these studies in terms of the level of breast cancer risk in the women included, the age at which surveillance commenced, and the surveillance protocols. Four studies undertook annual XRM, two were 'variable' and one was biennial. Five protocols incorporated regular CBE; two advised BSE and two did not involve any breast examination. Various comparison groups were utilised. Women matched for age but with no family history were used in three studies; women over 50 years in the UK national health service breast screening programme (NHSBSP) were used in three studies; one compared high, medium and low risk women, determined by degree of family history; and the last used symptomatic women with a family history as the comparator. The comparison of surveillance modalities, XRM versus CBE alone, generally received little consideration.

The outcomes of interest were the cancer detection rates and the stage of tumours detected in the four studies which contained pathological data. Three studies demonstrated that the cancer detection rate for women under 50 years of age at a high risk of breast cancer was similar to women over 50 years of age in the NHSBSP. Two studies showed tumours to be detected at an earlier stage in women at high risk who were under surveillance as opposed to those not under surveillance. However, earlier detection does not improve survival without effective treatment, and the effects of lead-time bias and length bias from early detection were not discussed in this review. Lead-time refers to the period of time between the detection of a medical condition by surveillance and when it would ordinarily be diagnosed by symptoms. Bias occurs as surveillance prolongs the measurable disease time but not actual survival time in the absence of effective treatment. Length bias occurs as slow-growing lesions are diagnosed by surveillance in excess of those diagnosed by symptoms. The effect is that a greater number of slow-growing cancers, i.e. those with better prognoses, are diagnosed by surveillance, with an apparent effect of prolonged survival time. These factors need to be taken into account in studies of surveillance which use early detection of tumours or survival as an outcome.

The authors concluded that the evidence for mammographic surveillance of women at increased risk for breast cancer was weak. They also examined some of the limitations and associated harms of mammography. The rationale of women at high risk being of equivalent risk to women over 50 years who are included in the NHSBSP was offered as justification for surveillance in this high-risk population. The limited pathological data were interpreted as suggesting a survival advantage in women with a family history of breast cancer who received surveillance. However, there was no evidence to show that mortality from breast cancer in this population of women was reduced by early detection.

There was very limited information provided in this review, particularly regarding the appraisal and limitations of the studies. Therefore, the studies reviewed have also been included in the primary literature section of this review.

Table 6. Secondary research appraised relevant to accuracy and efficacy of mammography surveillance on outcomes from breast cancer

Source	Search method	Criteria for inclusion/exclusion	Results	Comments
(Lucassen et al., 2001) Wessex Regional Genetics Service, Southampton, and CRC Primary Care Education Research Group, Department of Primary Health Care, University of Oxford Institute of Health Sciences (UK)	Search: January 1995-December 2000. Databases searched: Pubmed clinical query site, Medline, Cochrane database of Systematic Reviews and Best Evidence. Key words: "detection breast cancer in women with a family history" and diagnosis, "management women with a family history of breast cancer" (Pubmed). Mammography or breast screening and family history (Medline).	Clear PICO question (describing the Population, Intervention, Comparison and Outcome of interest) is not specified Examined articles on effectiveness of mammography in women at a high risk of breast cancer. Populations of studies were selected by varying definitions of high risk. Included studies with different frequency of mammography and different ages of commencement of surveillance. Some in conjunction with self and/or clinical breast examination. Inclusion criteria: No inclusion criteria given Exclusion criteria: Non-English papers and reviews.	No RCTs 7 studies included but unclear on study type Cancer detection rate: 4 of the 7 studies found that the cancer detection rate for women at a high risk aged <50 years was similar to women >50 in a screening programme. Tumour stage: 4 papers commented on tumour stage. Two of the 4 studies reported that tumours detected by surveillance were of an earlier stage, and one also reported that they were more likely to be node negative.	Did not specify clear PICO questions. Not clear on method or extent of appraisal of papers included Authors' conclusions: The authors concluded that the published studies provide evidence that the detection rate of cancer in women under 50 with a family history of breast cancer is equivalent to that in women over 50 in the general population who are screened. The limited pathology data also suggest that it is reasonable to expect a survival advantage in women with a family history of breast cancer who receive mammographic surveillance. However, they acknowledge that there was no evidence, at that time, to show that mortality from breast cancer in this group of women at high risk will be reduced by early detection. They also recognise that, although this evidence supports mammographic surveillance in women below 50 with a family history, there is no consensus on what level of family history is strong enough to qualify a woman for such surveillance.

PRIMARY RESEARCH

The search identified 24 eligible primary research studies. Below is an overview of study designs and aspects of quality represented by these studies. Full details of the papers appraised, including methods, key results, limitations and conclusions, are provided in evidence **Table 11**. Studies are presented in chronological order of publication within the tables.

Study design and quality assessment

The RCT is the most robust design to compare diagnostic tests' usefulness for surveillance, and could compare surveillance with XRM to surveillance with CBE alone or to no surveillance. However, as in the review by Lucassen et al. (2001), no such evidence was identified. Two factors were referred to in the introduction that prevent such research from being undertaken. The first is that there is RCT evidence from the screening population of women over the age of 50 that XRM is effective in detecting breast cancer and reducing mortality by around 30 per cent. (Tabar et al. 2000). The second factor is that, despite there being no RCT evidence (Duffy et al. 2006), CBE is generally accepted to be less effective for detecting breast cancer than XRM. This is because, as discussed, lesions are only palpable once they are over a certain size. Therefore, there would be concern about equipoise in conducting such a study. As a result, lower level designs have to be utilised and alternative comparisons drawn to determine the efficacy and accuracy of surveillance for women at high risk of breast cancer with XRM.

The 24 eligible studies were all graded as level III-2. They consisted of 17 prospective cohort studies, three retrospective cohort studies, two that contain both retrospective and prospective data, one cross-sectional study and one matched case-control study. They mostly examined the accuracy and efficacy of XRM, with or without CBE, in a population of women at high risk of breast cancer. Some studies also utilised other modalities of surveillance and these will be discussed in subsequent chapters. Outcome comparisons were then drawn, as outlined in the review by Lucassen et al. (2001), between the study cohort and either the general population of women over 50 years that receive XRM screening or a population of women who did not receive surveillance and presented symptomatically with breast cancer. The former comparison was based on the assumption that if similar accuracy and efficacy was found between the two groups, then surveillance in women at high risk should be acceptable as it is in the over 50-year age group. However, this assumption does not consider the harms of XRM that are possibly greater in a younger population of high-risk women. The second comparison aimed to demonstrate that tumours detected by surveillance have more favourable characteristics, such as size, stage and lymph node status, than those presenting symptomatically. The assumption was that detection at an earlier stage in tumour development would result in a better response to treatment and a better prognosis. However, these are intermediate outcomes which can be affected by biases, and benefits of surveillance demonstrated by these outcomes do not necessarily translate to a reduction in mortality. Only two studies addressed mortality in a surveillance population and a population without surveillance (Elmore et al. 2005; Maurice et al. 2006).

The studies were all of moderate quality in design and conduct and were all subject to several limitations. They were all likely to be affected by verification bias because the reference standard for diagnosis was different in the case of a positive surveillance result versus a negative result. In all the studies, positive surveillance results were followed by biopsy or surgical excision and histopathological confirmation. However, verification of negative surveillance results was only possible through clinical follow-up over the interval between rounds of surveillance. This follow-up would detect false negatives that arose as interval tumours but would not detect false negatives that did not present symptomatically. In the majority of the studies there was no blinding of the radiologist to the women's risk status. This knowledge may have affected their degree of suspicion and therefore the thoroughness in which they carried out CBE or interpreted the mammograms. The studies also tended to be lacking in information on the characteristics of their populations, other than basic demographics and risk status. They often did not comment on the characteristics of women who declined to participate in surveillance or who were lost to follow-up. Therefore, there may be selection bias that is unaccounted for. In addition, studies mostly did not account for factors that may increase a woman's risk of breast cancer, such as OCP or HRT use, or on the use of risk reduction strategies such as BSO or Tamoxifen. The percentage of women in the cohort who had undergone BSO could particularly alter the risk profile of the group as

this procedure has been shown to reduce the risk of breast cancer by up to 50 per cent (Kuschel et al. 2006).

Study setting

There was considerable heterogeneity in the settings from which these studies recruited participants. Eleven studies were undertaken in single centres and 13 were multi-centre. The centres of recruitment were usually breast screening or genetic screening clinics. One study recruited patients from a genetic registry for breast cancer. When appraising a study, it is important to consider the setting as this will determine the prevalence and spectrum of disease in the participant population (Deeks 2001). However, in these studies, this was also determined by the risk stratification that participants underwent.

Risk stratification

The methods of risk stratification varied greatly between the studies. The majority used family history factors alone and many referred to the model developed by Claus et al. (1994), which has risk tables to stratify women's' future risk of breast cancer Other studies developed their own systems of stratification and included factors such as personal history of breast cancer, previous biopsy, HRT use and atypical hyperplasia. This heterogeneity makes it difficult to draw comparisons across these studies. Due to the large number of studies and the variety of methods employed, the remainder of the information on these studies, including risk stratification strategies, is presented individually.

Chart et al. (1997)

Study sample

This prospective cohort study recruited 1,044 women at high risk of breast cancer from two breast clinics in Toronto, Canada. There were no age restrictions and the mean age of commencing surveillance in the two clinics was 42.7 years (SD, 10.9 years) and 39.5 years (SD, 10.8 years). Risk stratification was unique to the study and, in addition to family history factors, included previous benign or *in situ* breast disease, reproductive history, exposure to radiation and alcohol intake. Women were considered to be 'high' risk if their risk factors gave them a RR of more than four, 'moderate' risk if the RR was two to four and 'slightly increased' risk if their RR was less than two. Twelve per cent of women were lost to follow-up, but they were distributed among all the risk categories.

Interventions and comparators

The intervention was determined by the level of risk of breast cancer. Women at 'high' or 'moderate' risk commenced surveillance at 40 years, or 10 years before the earliest age at which cancer was detected in the family. Women at 'slightly increased' risk only commenced surveillance after 40 years of age. Women at 'high' risk received CBE every six months and annual XRM. Those at 'moderate' and 'slightly increased' risk received CBE and XRM annually and were advised to perform monthly BSE. The XRM views taken were not specified. The system of classification of images used was not documented, and it was not reported whether the radiologists interpreting the XRM images were blinded to the results of CBE. The average follow-up was 21.9 months. Comparisons were drawn between the outcomes in the different risk categories and between CBE and XRM.

Outcomes

Cancer detection rate

Nineteen tumours were detected during the study, 13 on the prevalent round and six on incident rounds. Ten were in 'high risk' women, two were in 'moderate' risk women and seven were in 'slightly increased' risk women. This gave an overall cancer detection rate of 18 per 1,000 women under surveillance. This was 12.4 per 1,000 women under surveillance for the prevalent round and 5.7 per 1,000 women under surveillance for the incident rounds. The mode of detection was only reported for the incident tumours. In the 'high risk' women, two tumours were detected by CBE and one by XRM. In the 'moderate' risk group, no tumours were detected by surveillance. In the 'slightly increased' risk group, XRM detected three tumours. Therefore the cancer detection rate for CBE was 1.9 per 1,000 women under surveillance (2/1,044), compared to 3.8 per 1,000 women under surveillance (4/1,044) for XRM.

Tumour characteristics

There were three incident tumours that were invasive and three that were DCIS. They were all sized from 10-15mm and were all lymph node negative. Of the prevalent tumours, six were *in situ* and seven were invasive. The tumours detected by CBE were both invasive, whereas three of the four tumours detected by XRM were *in situ.*

Interval tumours

BSE detected four interval tumours in the incident phase. There was also one tumour that presented in a woman who had not yet commenced surveillance. All of the interval tumours were invasive, sized from 10-15mm and were lymph node negative.

In summary, this study suggests that XRM is more effective than CBE at detecting breast cancer in women at high risk and is better at detecting *in situ* lesions. However, the statistical significance of these findings is not documented. The small number of tumours detected in this study means that it may be underpowered to demonstrate differences between these modalities of surveillance or between risk groups. There were a high number of interval tumours in the women at 'moderate' and 'high' risk and this suggests that this surveillance was not adequate for these risk groups. This may be because the surveillance was not sensitive enough in this population or because women at high risk have faster growing tumours. More regular surveillance may be able to partially compensate for this. However, surveillance with other modalities, alone or in combination with XRM, may perhaps be required for these risk groups in order to reduce the number of interval tumours.

Lalloo et al. (1998)

Study sample

This prospective cohort study recruited 1,259 women from one family history clinic in Manchester, U.K. They were all aged less than 50 years of age and the mean age at entry to the study was 39.1 years with a range of 28-49. Risk stratification was performed according to family history, using the Claus tables. Women were required to have a lifetime risk of breast cancer of one in six or greater, equivalent to a fourfold increase in risk for these women. Attendance rates of this population were high, at 95.2 per cent and 98.9 per cent for the prevalent and incident screens respectively.

Interventions and comparators

Surveillance consisted of annual XRM and CBE for all women commencing at 35 years, or at five years younger than the earliest diagnosis of breast cancer in the family. The prevalent surveillance round was two-view XRM (oblique and cranio-caudal) whereas the incident surveillance rounds were single-view (oblique). The system of classification of images used was not documented. All XRM images were interpreted by experienced radiologists, but it was not reported whether they were blinded to the results of CBE. The average follow-up was 30 months with a range 1-54 months. Comparisons were made between the CBE and XRM, and the number of cancers expected to arise in the general population with data from the regional cancer registry.

Outcomes

Cancer detection rate

Fourteen tumours were detected, seven at the prevalent surveillance round and seven at the incident surveillance round. This gave an overall cancer detection rate of 11 per 1,000 women under surveillance, with 5.6 per 1,000 women under surveillance for both the incident and prevalent screen. However, if calculated by the number of surveillance screens rather than the number of women under surveillance, the detection rate was 5.6 per 1,000 prevalent surveillance screens and 4.8 per 1,000 incident surveillance screens. Eight of the cancers were not palpable and were detected by mammography alone. Therefore the cancer detection rate for CBE was 4.7 per 1,000 women under surveillance (6/1,259). It is not possible to calculate the cancer detection rate for XRM alone as it is not documented how many of the tumours were palpable but not detected by XRM. The cancer detection rate is shown to be similar to women over 50 years in the National Health Service Breast Screening Service (NHSBSP), which is 5.7 per 1,000 prevalent screens and 3.8 per 1,000 incident screens. The number of invasive cancers detected was five times the number expected to arise in the general population without surveillance and 1.42 (95% CI, 0.72 to 2.48) times the number expected to arise in

the study population without surveillance. These latter comparisons are based on modelling and rely on the accuracy of these models.

Tumour characteristics

The detected tumours were invasive in nine cases and *in situ* (DCIS or LCIS) in five cases. They ranged in size from 7-45mm and four were lymph node positive, with two having no results of lymph node sampling. The tumour characteristics were not stratified by modality of surveillance.

Interval tumours

There were two interval cancers during the study. These were both invasive, less than 15mm in size and one was node positive.

In summary, this study shows that surveillance with CBE and XRM is effective at detecting early breast cancer in women at high risk and less than 50 years of age. The cancer detection rate is equivalent to that of accepted screening programmes for women over 50 years of age. It also appears to reinforce the need for XRM in addition to CBE, although there is a lack of information on tumour detection by modality to verify this. This study was also limited by the relatively small number of tumours in the cohort.

Kollias et al. (1998)

Study sample

This prospective cohort study recruited 1,371 women from a family history breast screening cancer clinic in Nottingham, UK. They were all less than 50 years in age and the mean age at the start of screening was 41 years (range 18-49 years). Risk stratification was performed according to family history using the Claus tables, but also incorporated atypical hyperplasia as a risk factor. They were required to have a lifetime risk of breast cancer of at least one in nine. There was no comment as to whether women with a past history of breast cancer were included.

Interventions and comparators

Surveillance consisted of biennial XRM and annual CBE. Patients were also instructed in BSE. Surveillance commenced 10 years younger than the earliest diagnosis of breast cancer in the family. XRM was two-view at the prevalent surveillance round (cranio-caudal and medio-lateral) and one-view (oblique) at the incident surveillance rounds. The mean follow-up was 22 months (range 0-96 months). The system of classification of images used was not documented. All XRM images were interpreted by radiologists experienced in breast imaging, but it is not reported whether they were blinded to the results of CBE. A comparison was made with the cancer detection rate in women over 50 years screened by the NHSBSP. Tumour characteristics were compared with women not under surveillance, less than 50 years of age, who presented with symptomatic breast cancer and had a family history that would have qualified them for surveillance in this study.

Outcomes

Cancer detection rate

Nineteen tumours were detected, 11 at the prevalent surveillance round and eight at incident surveillance rounds. This gave an overall cancer detection rate of 14 per 1,000 women under surveillance (8 per 1,000 women under prevalent surveillance and 5.8 per 1,000 women for the incident surveillance). This translated to a cancer detection rate of 8 per 1,000 surveillance screens and 3.3 per 1,000 surveillance screens for the prevalent and incident surveillance rounds respectively. At the prevalent surveillance round two tumours were not detected on XRM, but all incident tumours were detected by XRM. This gave an overall cancer detection rate of 12 per 1,000 women under surveillance (17/1,371) for XRM. It is not documented how many of the tumours were palpable. The cancer detection rate was similar to that of the NHSBSP detection rate of 6.5 per 1,000 prevalent screening visits and 3.8 per 1,000 incident screening visits.

Tumour characteristics

Eight of the prevalent tumours were invasive and three were *in situ*. Six of the incident tumours were invasive and two were *in situ*. The size and lymph node involvement of the invasive tumours is only

reported for all tumours together, including interval cancers. There were 15 tumours sized 0-20mm and eight were sized over 20mm. Fifteen tumours were lymph node negative and eight were positive. A significant difference (p= 0.01) was seen between the proportion of cases of DCIS seen in the group under surveillance (6/29, 21%) and in symptomatic women not under surveillance (2/54, 3.7%). However, no differences were demonstrated in tumour size, histological grade or lymph node status. When the Nottingham Prognostic Index (NPI) was applied, a higher proportion of the group under surveillance (12/29, 41%) were categorised as having a good prognosis than the symptomatic group (16/54, 30%), but this was not statistically significant (p=0.37).

Interval tumours

There were 10 interval cancers and one cancer that arose in a woman who had not yet commenced surveillance. Nine of the interval tumours were invasive and one was *in situ*.

In summary, this study shows that surveillance with CBE and XRM is effective in detecting early breast cancer in women at high risk of breast cancer and less than 50 years of age. The cancer detection rate is equivalent to that of accepted screening programmes for women over 50 years of age. The rate of interval cancer is higher than that in the NHSBSP, suggesting that annual rather than biennial surveillance may be warranted in this group of women. More DCIS tumours appear to be detected in the group under surveillance as compared to the group not under surveillance, although other tumour characteristics did not show a statistically significant difference between these groups. This may in part be due to over-detection in the group under surveillance. There is not enough information documented to compare the detection rate by modality of screening.

Federico et al. (1999)

Study sample

This prospective cohort study recruited 151 women from family pedigrees of breast cancer compiled by the University of Modena, Italy. The age for inclusion was 30 to 65 years of age, but the mean age of the cohort at entry to screening was not reported. Risk stratification was performed by criteria specific to this study, involving family history (first and second degree) of breast or ovarian cancer. The women were then categorised as HBC (hereditary breast cancer), HBOC (hereditary breast and ovarian cancer), SHBC (suspected hereditary breast cancer), FBC (familial breast cancer) and a control group with no criteria of increased risk. There was no mention of whether women with a past history of breast cancer were included or excluded. An inclusion criterion was that all women must reside within the province of Modena. Forty-five women who were selected to participate did not receive surveillance due to a lack of compliance or prior involvement with another institution. The characteristics of these women were not given.

Interventions and comparators

Surveillance consisted of biannual (twice a year) CBE and XRM, according to the women's level of risk. HBC and SHBC women received XRM biennially (every two years) from 30 to 36 years and then annually from 37 to 65 years. FBC women received XRM biennially under 50 years and then annually after 50 years. The routine practice of BSE was not recommended. The views taken at XRM were not reported. The mean follow- up was 24 months. The system of classification of images used was not documented. It was not reported whether radiologists interpreting the XRM images were blinded to CBE results. Comparisons were made between screening programmes carried out in Northern and Central Italy in women over 50 years.

Outcomes

Cancer detection rate

Six tumours were detected in total. Three were in the HBC and HOBC women, two in the SHBC women and one in the FBC women. Three were detected on the prevalent surveillance round and three on the incident surveillance rounds. This gave an overall cancer detection rate of 40 per 1,000 women under surveillance (20 per 1,000 women under surveillance for both the prevalent and the incident surveillance). The exact number of surveillance screens the women underwent is not documented, but the overall cancer detection rate by surveillance screens was 19.1 per 1,000 surveillance screens. The six tumours were all detected by XRM. It is not documented how many were palpable. The cancer

detection rate was reported to be higher than that seen in the breast cancer screening programmes currently operating in Italy for the general population of women over 50 years, which is 7.7 to 8 per 1,000 screens.

Tumour characteristics

One prevalent tumour was *in situ* and two prevalent and three incident tumours were invasive. The tumours ranged in size from 0.9-35mm. Two of the six tumours had lymph node involvement. The tumour characteristics were not stratified by modality of surveillance.

Interval tumours

One tumour was described as an interval tumour but was detected on CBE, so could be considered to be detected by surveillance as the protocol involved CBE.

Mortality

Of the women diagnosed with cancer, one had died by the time of publication, but there had not been adequate follow-up to assess mortality (range 2-62 months).

In summary, this study shows that surveillance with CBE and XRM is effective at detecting breast cancer early in women at high risk of breast cancer. The cancer detection rate is higher than that of accepted screening programmes for women over 50 years of age in Italy. There was insufficient information on detection by modality of surveillance to be able to draw comparisons between CBE and XRM.

Moller et al. (1999)

Study sample

This research collated data from seven prospective cohort studies in the European Union. The reporting of this paper was poor as it did not present the total number of women screened. There were no reported age restrictions and the mean age was only given for diagnosis (48.6 years, range 28-71 years). Risk stratification was based on the Claus tables and women were eligible for surveillance if their risk was at least twice that of the general population. Women were only included if they were asymptomatic and had no past history of breast cancer.

Interventions and comparators

Surveillance varied between centres but included XRM, usually annually, from age 35-50 years. This started at an earlier age if there was very early onset disease in the family. For women over 50 years, surveillance intervals in some centres were longer, ranging from 18 months to two years. This was combined with CBE, but the regularity of this is not mentioned. SBE was also advocated. The average length of follow-up is not reported. The system of classification of images used was not documented. It was not documented whether radiologists interpreting XRM were blinded to the results of CBE. Comparisons were made between XRM and CBE and between the survival of women screened and sporadic cases of breast cancer in an unscreened population.

Outcomes

Cancer detection rate

There were 121 tumours detected by surveillance. Forty were detected at the prevalent surveillance round and 81 at incident surveillance rounds. The overall cancer detection rate could not be calculated as the number of women under surveillance or number of surveillance screens done was not documented. In the prevalent surveillance, eight tumours were not detected by XRM and were detected by CBE. In the incident surveillance, twenty tumours were not detected by XRM and were detected by CBE. There is no comment on the number of tumours which were detected by XRM but not CBE.

Tumour characteristics

In the prevalent surveillance round there were eight (20%) *in situ* tumours, 32 (80%) invasive tumours and nine (23%) were lymph node positive. In the incident surveillance rounds there were 22 (27%) *in*

situ tumours, 59 (73%) invasive tumours and 13 (16%) were lymph node positive. The tumour characteristics were not stratified by mode of detection.

Interval tumours

There were 29 interval tumours and also 11 tumours that arose in the time between women being referred for surveillance and actually receiving it. Out of the interval tumours, two (7%) were *in situ*, 27 (93%) were invasive and 11 (38%) had positive lymph nodes.

Survival

Five-year overall survival was calculated as 0.89 (SE 0.05) and five-year event-free survival was 0.86 (SE 0.06). Stage-specific survival was similar to that reported for sporadic breast cancer, but the overall survival was better. This is said to indicate that prognosis is related to stage at diagnosis and that the effect of the surveillance intervention was mediated through early diagnosis. However, the survival figures from the study may be affected by lead-time bias and length bias, and they were not adjusted for this.

In summary, surveillance with CBE and XRM in women at high risk of breast cancer is effective in detecting early breast cancer and may offer a survival advantage over women who are not under surveillance. It suggests that CBE is necessary in addition to XRM, although not enough information is documented to calculate and compare cancer detection rates by modality. A survival advantage is suggested in women under surveillance as opposed to those who were not under surveillance and had sporadic tumours. Due to the high number of interval cancers, it also suggests that more intensive surveillance or surveillance with other modalities may be warranted in this population of women.

Macmillan et al. (2000)

Study sample

This multi-centre study collected data prospectively within nine breast units and retrospectively within 13 breast units in the UK. There were 8,783 women recruited altogether. All women were aged less than 50 years. The average age at entering surveillance was not given but the median age at diagnosis was 41 years at the prevalent surveillance round and 44 years at the incident surveillance rounds. Risk stratification was performed with family history criteria stipulating that women must have at least one first-degree relative affected by breast cancer before the age of 50 years to qualify for surveillance. Women were also only included if they were asymptomatic. There is no mention of whether women were included or excluded if they had a past history of breast cancer.

Interventions and comparators

Surveillance varied between the different breast units. XRM was offered annually by 11 units, biennially by 10 units and every 18 months by one unit. CBE was offered annually by 20 units and one unit also performed US. The surveillance was commenced five years before the age of the youngest affected relative in 12 units, and 10 years before this age in five units. One unit started surveillance at 30 years, one at 35 years, one at 40 years and two units had no fixed criteria for surveillance. There were a total of 9,075 women years of follow-up (1.03 per woman, range not given). The system of classification of images used was not documented. It is not reported whether the radiologists interpreting the XRM images were blinded to the results of CBE. Comparisons were made between the cancer detection rate in the study and in the NHSBSP study of women aged 50-64 years, and the Gothenburg trial of screening women aged 40-49 years. There was also comparison between CBE and XRM detection, and the prognostic index in the study and that of the NHSBSP and of women presenting symptomatically with breast cancer. Lastly, annual and biennial surveillance were compared.

Outcomes

Cancer detection rate

Eighty-three cancers were detected in this cohort. There were 42 prevalent and 41 incident tumours. The corresponding cancer detection rates were 9.4 per 1,000 women under surveillance overall, and 4.8 and 4.7 per 1,000 women under surveillance for the prevalent and incident surveillance respectively. The cancer detection rates were 4.78 and 4.52 per 1,000 prevalent and incident surveillance screens

respectively. Data on the mode of detection was only available for 67 tumours. Of the prevalent tumours, 47 per cent were palpable and 85 per cent were visible on XRM. Of the incident tumours, 62 per cent were palpable and 100 per cent were visible on XRM. The raw data for this comparison are not documented. The cancer detection rates were comparable to the NHSBSP of 6 per 1,000 prevalent screens and 4.6 per 1,000 incident screens, and much higher than in the Gothenburg trial, which had 1.5 per 1,000 prevalent screens and 1.7 per 1,000 incident screens. A significant difference was not able to be demonstrated between annual and biennial surveillance, although the cancer detection rate was higher in the former, at 5.71 per 1,000 surveillance screens compared with 3.64 per 1,000 surveillance screens (RR=0.64, 95% CI = 0.34 to 1.19, p=0.15).

Tumour characteristics

Complete data on tumour characteristics were only available for 75 of the 83 tumours. Eleven (26%) of the prevalent tumours were *in situ* and 31 (74%) were invasive. Their mean size was 19.9mm and 12 were node positive. Six (15%) of the incident tumours were *in situ* and 35 (85%) were invasive. Their mean size was 13.9mm. The tumour characteristics were not stratified by modality of surveillance. A higher proportion of women in the study group had a more favorable NPI than women of a similar age who presented symptomatically with breast cancer, but women in the NHSBSP had the highest proportion of favorable NPI, though raw data were not presented. The tumour characteristics, in terms of proportion of tumours that were *in situ,* were also better ($p < 0.001$) in the study group (17/75, 22%) than the symptomatic group (26/440, 6%), but not significantly different (p=0.48) from the NHSBSP (48/264, 18%). There was no statistically significant difference in the proportion of node negative tumours between the study group and the NHSBSP or symptomatic group (p = 0.17 and 0.9 respectively).

Interval tumours

There were 20 interval tumours, with a mean size of 19.4mm. Three (15%) were *in situ*, 17 (85%) were invasive and seven (35%) were node positive.

In summary, this study shows that surveillance with CBE and XRM is effective in detecting early breast cancer in women at high risk of breast cancer. The cancer detection rate is equivalent to that of accepted screening programmes for women over 50 years of age and higher than that of women at average risk under 50 years of age. There appears to be better tumour characteristics and prognostic index in women under surveillance than symptomatic women who had not undergone surveillance. The importance of XRM in addition to CBE is also reinforced, although the raw data were not documented for this to be verified. This study is limited by its partly retrospective nature, by the heterogeneity in surveillance protocols between different units and the lack of information from some units.

Kerlikowske et al. (2000)

Study sample

This cross-sectional study recruited 389 533 women from seven screening registries in six U.S. centres, 50 834 of which had a family history of breast cancer. The age criterion for inclusion was 30-69 years. The mean age was not presented. Risk stratification was by family history and women were included in this group if they had at least one first-degree relative with breast cancer. Inclusion criteria also stipulated the time of referral for surveillance as 1985 to 1997, and only the first mammogram of each woman was included. Women were excluded if they had a past history of breast cancer, a palpable mass on CBE or if they lived outside the catchment area for the tumour registry and results databases being used.

Interventions and comparators

The surveillance examined in this study was the prevalent mammogram only. However, a high percentage of women reported previous use of mammography with the screening registry, i.e. 81.7 per cent of women with a family history and 80.2 per cent of women without. Twelve months' follow-up was examined after the surveillance examination to determine true and false negatives and positives. Two standard mammographic views were taken per breast. BIRADS was used to classify the images and images classed as 0 (incompletely assessed), 3, 4 or 5 were considered abnormal. It was not reported whether radiologists interpreting the XRM images were blinded to the results of CBE. Comparisons were made between women with and without a family history of breast cancer and

between different age groups. XRM was also compared with CBE as all the tumours detected in this study had not been palpable.

Outcomes

Cancer detection rate

There were 1,650 tumours detected and 309 of these were in women with a family history of breast cancer. This gives a cancer detection rate in these high risk women of 6 per 1,000 women under surveillance or 6 per 1,000 surveillance screens. In the women with a family history, 70 tumours were *in situ* and 239 were invasive. This is the same as the proportion in women without a family history. The cancer detection rate was found to be 1.5 times higher in women with a family history of breast cancer than those without a family history ($p<0.001$). The rate increased with age in both groups. When stratified by age, the cancer detection rate of women at high risk of breast cancer was equivalent to women a decade older without such a history. This detection rate was in women with no palpable tumours, so it demonstrates the importance of utilising XRM in addition to CBE.

Sensitivity

The measures of accuracy were all presented stratified by age group and the presence of a family history of breast cancer. Some overall figures are documented in the text but no raw data were given to allow verification of these calculations.

The overall sensitivity for women at high risk is not presented. The sensitivity of XRM increased significantly with age. This ranged from 63.2 per cent (95% CI, 41.5 to 84.8) for ages 30-39 years to 83.8 per cent (95% CI, 76.8 to 90.9%) for ages 60-69 years in women with a family history of breast cancer ($p=0.006$, Chi squared test for trend). There was no significant difference in sensitivity found between women at high risk of breast cancer and women without such a history ($p=0.1$).

Specificity

The overall specificity for women with a family history is not documented. The specificity of XRM did not alter significantly with age in women with a family history. The specificity in women with a family history was lowest in the 40-49 year old group at 86.7 per cent (95% CI, 86.3 to 87.2) and highest in the 30-39 year old group at 89.4 per cent (95% CI, 88.6 to 90.2). The specificity was lower among women with a family history of breast cancer than in those without such a history ($p<0.001$).

PPV

The overall PPV for women with a family history is not documented. The PPV increased significantly with age in women with and without a family history ($p=0.001$, Chi squared test for trend for both groups) and was higher in women with a family history than in those without such a history (3.7% versus 2.9%, $p= 0.001$, chi squared test for trend).

Tumour characteristics

There was no documentation of tumour characteristics or node status.

In summary, this study showed that surveillance with XRM is effective in detecting early breast cancer in women with a family history and detected tumours not identified by CBE. The study was limited by its cross-sectional design.

Nixon et al. (2000)

Study sample

This study examined a prospective cohort of data from a randomised controlled trial. There were 3,226 women with a family history recruited that were within the Swedish Two County Trial (Tabar et al., 1992), of mammographic screening. No age restrictions were specified and the mean age was not given. The results were stratified by two age groups, 40-49 years and 50-74 years. Risk stratification was not described, other than saying these women had a 'family history'. It is not clear if women with a personal history of breast cancer were included or excluded.

Interventions and comparators

The original RCT compared mammographic screening with no screening. Family history data were only recorded for the women who received screening. Screening occurred biennially in women aged 40-49 years and at 33-month intervals in women aged 50-74 years. There appeared to be seven or eight years of data, but the mean follow-up time was not given. The system of classification of images used was not documented. It was not reported whether the radiologists interpreting the XRM images were blinded to the results of CBE. Comparisons were drawn between women with and without a family history of breast cancer and between the two age groups. CBE was not used in this study.

Outcomes

Cancer detection rate

Forty-five tumours were detected overall in the women with a family history of breast cancer. This gives a cancer detection rate of 14 per 1,000 women screened. Overall, there was a higher cancer detection rate in women with a family history than there was for the whole cohort receiving screening, i.e. 9 per 1,000 women screened.

Tumour characteristics

The detected tumour characteristics refer to invasive cancers only. Thirty-two of the tumours were less than 20mm in size (71%) and 10 were over or equal to 20mm (22%). The size of three tumours was not recorded. Nine were lymph node positive (20%) and 29 were lymph node negative (64%). The tumour characteristics were documented as not being significantly different between the two cohorts, although no *p* values are given.

Interval tumours

There were 15 interval tumours in women with a family history (4.6 per 1,000 women screened). This was a higher interval cancer rate than in women without such a history, i.e. 2.7 per 1,000 women screened. However, the percentage of screen-detected to interval cancers was similar in both cohorts. Of the 15 interval tumours, in women at high risk, nine were less than 20mm (60%) and six were 20mm or over (40%). Eleven were node positive (73%) and four were node negative (27%).

In summary, this study showed that screening in women with a family of history of breast cancer has as high, if not higher, cancer detection rate as in women without this history. This will be due to the higher prevalence of breast cancer in women with a family history. There is a high interval cancer rate in women with a family history. This suggests that more intensive screening or additional modalities of screening may be required for surveillance in this population to improve the sensitivity and reduce the number of interval tumours. This study has many limitations due to its design. There appears to be data missing. There is also little information on risk stratification, which diminishes the external validity of the study.

Myles et al. (2001)

Study sample

This prospective study recruited 2,998 women with a moderate family history of breast cancer from a family history clinic in Manchester, UK. There were no age criteria given and the mean age was not given, but the range was 19-71 years. Risk stratification was performed according to the Claus tables. The women included had sufficient family history to indicate a moderate risk but not sufficient to warrant gene mutation analysis. There were no other inclusion or exclusion criteria and no mention whether women with a past history of breast cancer were included or excluded.

Interventions and comparators

Surveillance consisted of annual XRM, although this was reduced to biennial in the latter years of the study due to pressures on resources. CBE was also used and some women received US. The mean follow-up time was not given. The system of classification of images used was not documented. It was not reported whether the radiologists interpreting the XRM images were blinded to the results of CBE. Comparisons were made between the cancer detection rate and the incidence of tumours in this cohort and the figures predicted by the Claus tables.

Outcomes

Cancer detection rate

Forty-one cancers were detected. This gives an overall cancer detection rate of 13.6 per 1,000 women under surveillance. There were 15 prevalent tumours (5 per 1,000 prevalent surveillance screens) and 26 incident tumours (4.9 per 1,000 incident surveillance screens). The incidence predicted by the Claus tables was 3.73 per 1,000 person years and the cancer detection rate is just over this incidence. The cancer detection rate is not stratified by the mode of detection.

Sensitivity

The sensitivity of this surveillance strategy is modelled to be 70%, i.e. estimated that 70% of tumours would be detected by surveillance and 30% would arise between surveillance rounds as interval tumours.

Tumour characteristics

No characteristics of the tumours were presented.

Interval tumours

There were nine interval tumours

Mortality

Using the estimated sensitivity and also data from the Swedish Two County Trial (Tabar et al., 1992), suggesting a 59 per cent lower mortality from screen detected tumours compared to non-screen detected, a mortality reduction of 41 per cent is estimated from this surveillance strategy.

In summary, this study suggests that surveillance with XRM is effective in women at high risk of breast cancer on the basis that it detects as many tumours as would be predicted to arise in this population, according to the Claus tables. It also estimates a reduction in mortality based on the above modelling. There were still a considerable number of interval cancers, which indicates that perhaps more intensive surveillance or additional modalities are required for this cohort. The study is limited by the vague description of the surveillance, which is not explicit about the surveillance interval and does not describe the role of CBE or US. This was also a relatively low-risk cohort, and this reduces its comparability with other studies.

Brekelmans et al. (2001)

Study sample

This study combined analysis of a retrospective and prospective cohort. A total of 1,198 women at high risk of breast cancer were recruited from a large cancer clinic in the Netherlands. Data were collated retrospectively for women under surveillance prior to 1995 and prospectively for women after this date. There were no age restrictions specified and the mean age at the start of surveillance was 38 years with a range of 21-70 years. Risk stratification was performed by genetic testing and the use of the Claus tables. Women were included if they had over a 15 per cent lifetime risk of breast cancer. The women were divided into groups of mutation carriers (60-85% lifetime risk), 'high' risk (30-50% lifetime risk) and 'moderate' risk (15-30% lifetime risk). There was no documentation on whether women with a past history of breast cancer were included or excluded and no other inclusion or exclusion criteria.

Interventions and comparators

Surveillance consisted of biannual CBE (annual in some moderate risk women). Annual XRM and monthly BSE was recommended. The youngest age for XRM was 25 years, although this was commenced earlier in women from families with a young age of onset. After 1995, MRI was offered to some women with dense breasts and US was used when clinically indicated. The mean follow-up period was three years with a range of 0-22 years. There were a total of 3,607 person years of follow-up. The system of classification of images used was not documented. It was not reported whether the radiologists interpreting the XRM images were blinded to the results of CBE. Comparisons were made between the study population and the expected number of cancers in a population aged 40-50 years, according to the national cancer registry. The modalities of screening, CBE and XRM, were also compared.

Outcomes

Cancer detection rate

Twenty-six tumours were detected in total, with three prevalent and 23 incident. The overall cancer detection rate was 21.7 per 1,000 women screened or 7.2 per 1,000 person years. The cancer detection rates were presented by the three risk groups, but they exclude *in situ* tumours and include the interval tumours. These were recalculated as 3.3 per 1,000 person years, 7.9 per 1,000 person years and 18.6 per 1,000 person years in the 'moderate', 'high' and mutation carrier risk groups respectively. Nine tumours were detected by XRM alone, one was detected by CBE alone and 12 were detected by both CBE and XRM. There were also three detected by MRI alone and one detected by MRI and CBE. Cancer detection was also found to increase with age. Therefore the cancer detection rate for XRM and CBE together was six per 1,000 person years (22/3,607), for XRM alone was 5.8 per 1,000 person years (21/3,607) and for CBE alone was 3.9 per 1,000 women years (14/3,607). The ratio of observed (in the study) to expected (age matched population in national cancer registry) breast cancers was seven overall. This was 23.7 in the carrier group, seven in the 'high' risk group and just under three in the 'moderate' risk group.

Sensitivity

The overall sensitivity of the surveillance protocol was 74 per cent (95% CI 57-88). However, there were some tumours detected by the selective use of MRI. The sensitivity by mode of detection was 40 per cent (95% CI 24-58) for CBE alone, 60 per cent (95% CI 42-76) for XRM alone and 66 per cent (95% CI 48-81) for CBE and XRM. Comparison of the proportion detected by CBE and XRM with CBE alone did not reveal a statistically significant difference ($\chi 2$ with Yates correction 3.67, P=0.055). The overall sensitivity was shown to be lowest in the mutation carrier group (56%) although the difference between the risk groups were not statistically significant (p=0.21). There is also an increase in sensitivity with age but this result is also not significant (p=0.61). The overall sensitivity (including the three tumours detected by MRI) is said to be comparable to that of the Dutch Breast Screening Programme.

Tumour characteristics

Four of the tumours detected by surveillance were *in situ* and the rest were invasive. All of the interval tumours were invasive. The other tumour characteristics were presented with the tumours detected by surveillance and interval tumours combined. Ten tumours were sized over 10mm, eight were 10-15mm and 11 were over 15mm, with a range of 16-40mm. Eleven were node positive (two prevalent, six incident and three interval). There were more node positive tumours in the proven mutation carriers and the younger age group, but these results were not statistically significant. There were no significant differences in tumour characteristics between the risk groups (*p* values all >0.05). Tumour characteristics were not stratified by mode of detection.

Interval tumours

There were nine interval tumours (2.5 per 1,000 person years), four in the mutation carrier group, five in the 'high' risk group and none in the 'moderate' risk group.

Mortality

Overall, three participants died, two of metastatic breast cancer and one of another cause.

In summary, this study demonstrates a higher prevalence of breast cancer in women at high risk. It suggests an equivalent sensitivity of this surveillance strategy to the existing Dutch Breast Screening Programme, although figures for the programme are not documented. It also suggests a benefit of surveillance with XRM in addition to CBE, although the difference was not statistically significant. The fact that all interval tumours occurred in the mutation carrier and 'high' risk groups suggests that these women may require more intensive surveillance or additional modalities of surveillance. The retrospective nature of data collated prior to 1995 means that it is more prone to bias. The selective use of MRI and US may also have confounded the results and the way in which the results were presented in this paper was unclear.

Gui et al. (2001)

Study sample

This prospective cohort study recruited 2,578 women from a breast diagnostic unit in London, UK. There were 1,500 women at standard risk of breast cancer and 1,078 women at moderate to high risk of breast cancer. There were no age restrictions and the median age at commencement of surveillance was 48 years in the 'standard' group and 44 years in the 'moderate to high' risk group. Risk stratification was performed according to the Claus tables. The women at 'standard' risk had less than a 16 percent lifetime risk of breast cancer. The women at 'moderate to high' risk had at least a 16% lifetime risk of breast cancer. In this latter group, 22 per cent had over a 25 per cent lifetime risk of breast cancer. Other inclusion criteria were: that the women were already patients at the unit and were known to be cancer-free at the start of the study. These were therefore all incident surveillance rounds.

Interventions and comparators

Surveillance consisted of annual XRM and CBE from the age of 35 years, or commencing five years younger than the earliest diagnosis within the family if that was earlier than 40 years. However, women less than 25 years old did not udergo surveillance. The mean follow-up period was 3.9 years for the 'standard' risk group and four years for the 'moderate to high' risk group. The system of classification of images used was not documented. It was not reported whether the radiologists interpreting the XRM images were blinded to the results of CBE. Comparisons were made between the two risk groups and cancer detection rates from the NHSBSP. The modalities of surveillance, CBE and XRM, were also compared.

Outcomes

Cancer detection rate

Thirty-one tumours were detected overall, 19 in the 'moderate to high' risk group and 12 in the 'standard' risk group. The cancer detection rate for the 'moderate to high' risk group was reported as 4.4 per 1,000 women years. The detection rate reported for the 'standard' risk group was 2 per 1,000 women years. In the 'moderate to high' risk group, six tumours were detected by CBE and XRM, 10 were detected by CBE alone and three by XRM alone. These results are difficult to interpret as they include the interval tumours. It was stated that out of the 26 tumours detected by CBE in the standard and moderate to high risk group, 17 were found by patients (interval tumours) and nine by clinicians. This means only nine overall were really detected by CBE, a cancer detection rate of 1.4 per 1,000 women years. As the interval tumours were said to be detected by CBE, not BSE, they have been included in the cancer detection rates. This is misleading. It is not possible to recalculate the detection rates as it is not clear how many of these interval tumours were in each risk group. It is also not possible to recalculate the detection rate for XRM alone as the interval tumours must have received diagnostic XRM once detected by BSE and it is not stated how many were XRM positive and how many were XRM negative. The cancer detection rate in the 'moderate to high' risk group of 4.4 per 1,000 women years is compared to that of the NHSBSP rate of 3.8 per 1,000 and used as justification for this screening. However, the overall cancer detection rate in this study would be much lower if the interval cancers were removed. This cannot be calculated for the risk groups individually, but it would lower the overall rate from 3 per 1,000 women years to 1.4 per 1,000 women years.

Sensitivity

The sensitivities reported in this study are calculated with false negatives (interval tumours detected by BSE) included in the numerator. These results give a 100 per cent sensitivity overall, 55 per cent for XRM alone and 84 per cent for CBE alone. For the women at 'moderate to high' risk the sensitivities were 47 per cent for XRM alone and 84 per cent for CBE alone. In women at 'standard' risk the sensitivities were 66 per cent for XRM alone and 85 per cent for CBE alone. These results are very misleading and suggest an advantage of CBE over XRM, as all the false negatives are included in the CBE category although they were detected by BSE. When recalculated without the false negatives in the numerator, the sensitivity for CBE and XRM for all women ('standard' and 'moderate to high' risk) is only 45 per cent. The recalculated sensitivity of CBE alone for all women is 29 per cent. Unfortunately, it is not possible to recalculate the results for XRM alone as some of the false negatives had diagnostic XRM once detected by BSE and it is not stated which are in the XRM positive group or

the XRM negative group. It is also not possible to recalculate the results stratified by risk group as the false negatives are not attributed to the risk groups they arose in.

Tumour characteristics

In the 'moderate to high' risk group there were 17 invasive and two *in situ* tumours. Tumour size ranged from 10-30 mm. The lymph node status was only known for nine tumours and six were positive. The tumour characteristics were not stratified by modality of surveillance.

Interval tumours

Seventeen interval tumours were detected by BSE that were included in the results as CBE, and therefore the study overestimated the efficacy and accuracy of CBE. These were among the 'moderate/high' risk group and the standard 'risk' group, and thereby prevented the calculation of results of surveillance detected tumours for women at high risk of breast cancer alone.

Mortality

One woman died in each risk group of metastasized disease.

In summary, the results of this study are unclear. The authors emphasised the importance of CBE over XRM but once the interval tumours are removed, XRM appears to be more effective at detecting tumours. This calculation cannot be done for women at high risk alone as there is not enough data presented. The combination of both CBE and XRM is more effective at detecting early breast cancer in women at high risk, than either modality alone. In addition the high number of interval tumours in this cohort suggested that more intensive surveillance or other modalities of surveillance were required.

Hou et al. (2002)

Study sample

This prospective cohort study recruited 935 women who were relatives of breast cancer patients in a Taiwanese hospital. All participants were aged over 35 years of age and the mean age at surveillance was 48.6 years with a range of 35-75 years. There was no specific risk stratification process, but all participants had at least one first-degree (mother, sister or daughter) or second-degree (grandmother) relative with breast cancer. Exclusion criteria were: pregnant or lactating women and a past history of breast cancer or known metastatic disease.

Interventions and comparators

Surveillance consisted of annual CBE, XRM and US. The US results will be discussed in a following chapter. The BIRADS system was used for both US and XRM with scores of 4 and 5 leading to biopsy. It is not reported whether the radiologist interpreting the XRM images was blinded to the results of CBE. The median follow-up time was 41.8 months with a range of 12-82 months). Comparisons were drawn between the different modalities of surveillance.

Outcomes

Cancer detection rate

Twenty-one cancers were detected by the overall surveillance strategy, including US. This gives a cancer detection rate of 22 per 1,000 women under surveillance. Seven of the tumours were detected by CBE (7 per 1,000 women under surveillance) and 11 by XRM (12 per 1,000 women under surveillance).

Sensitivity

The sensitivities presented in the paper are only calculated with the cancers detected by surveillance as the denominator, and they did not include the interval cancer. The documented figures are 33.3 percent for CBE and 52.4 per cent for XRM. No confidence intervals are documented. If calculated with the interval tumour as a false negative, the respective results are 32 per cent (95% CI, 13.9 to 54.9%) and 50 per cent (95% CI, 28.2% to 71.8%).

Specificity

It is unclear how the specificities have been calculated. The documented figures are 83.5 per cent for CBE and 82.2 per cent for XRM. No confidence intervals are documented.

Tumour characteristics

Sixteen tumours were invasive, two were DCIS, two were mucinous carcinomas and one was a medullary carcinoma. The mean tumour size was 12mm and seven were lymph node positive. These characteristics were not stratified by mode of detection.

Interval Tumours

There was one interval tumour.

Survival

The five-year overall survival was 90.4 per cent and the disease-free survival rate was 80.9 per cent. This calculation was not adjusted for lead-time bias or length bias.

In summary, this study suggested that XRM was more accurate and effective than CBE at detecting tumours in women at high risk of breast cancer in Taiwan. However, no measures of statistical significance are documented in this study. It is noted that these findings may be specific to Asian women and may not be reproducible in a Western population.

Scheuer et al. (2002)

Study sample

This prospective cohort study recruited 165 women from 1,865 patients who had received genetic testing at a cancer centre in the USA. There were 251 women who consented to participate, but those who chose bilateral mastectomy after discovering their mutation status did not participate in the surveillance. There was no age restriction and the mean age of genetic testing was 47.7 years with a range of 24-79 years). Risk stratification was not performed as all these women were proven BRCA1 or BRCA2 mutation carriers. A high percentage of these women had a prior history of breast cancer, but the exact number is not specified (59% of the group of 251 that consented to participate). Five per cent of the group of 251 that consented to participate had had a risk-reducing BSO. Eight women were lost to follow-up.

Interventions and comparators

Surveillance consisted of CBE two to four times a year and annual XRM, both commencing at the age of 25 years. Monthly SBE was advised from the age of 18 years. Some women also received US and MRI, but there is little information on the use of these modalities. The mean follow-up period was 24.8 months with a range of 1.6-66 months. The system of classification of images used was not documented. It was not reported whether radiologists interpreting the XRM images were blinded to the results of CBE. Comparisons were made between the use of CBE and XRM for surveillance.

Outcomes

Cancer detection rate

Seven tumours were detected by surveillance. This gives a cancer detection rate of 31 per 1,000 women under surveillance. Six of these tumours were in women with a prior history of breast cancer. Five tumours were detected by XRM (30 per 1,000 women under surveillance). The authors did not state whether these were palpable or not. One tumour was detected by CBE performed in the XRM screening interval. One tumour was detected by MRI alone.

Sensitivity

The sensitivity of XRM and CBE combined was 50 per cent (95% CI 21-79) while the sensitivity for XRM alone was 42 per cent (95% CI 15-72). The difference in sensitivity between these two groups was not statistically significant (p=0.68). It was not possible to calculate the sensitivity for CBE alone as the number of tumours detected by XRM which were palpable on CBE is not reported.

Tumour characteristics

Half of the tumours detected by XRM and MRI were invasive and half were *in situ*. The invasive tumours were all less than 20mm in size and one was lymph node positive. The tumour identified by CBE was 25mm in size and both invasive and lymph node positive. The tumour characteristics were not stratified by mode of detection.

Interval tumours

Five interval tumours were detected by SBE. The interval tumours were less than or equal to 18mm in size and one was lymph node positive.

In summary, this study shows that there is no statistically significant difference between the sensitivity of a combination of surveillance by XRM and CBE, and XRM alone. From the documented data it is not possible to compare CBE alone with XRM or XRM plus CBE. The high interval cancer rate again suggests that more intensive surveillance or additional modalities of surveillance are required in this high-risk population. The external validity of this study to all women at high risk is reduced by the participants being at such high risk, i.e. mutation carriers, a high proportion of who had a personal history of breast cancer, and the results of surveillance may have been confounded by the use of MRI and US.

Trecate et al. (2003)

Study sample

This prospective cohort study recruited 23 women at high risk of breast cancer from the National Cancer Institute in Milan, Italy. There was no age restriction and no average age of the cohort was given. The age range was 30-61 years. Risk stratification was specific to this study. The women included were either BRCA1 or BRCA2 mutation carriers, had a one-in-two probability of being a carrier or greater than a 50 per cent risk of carrying a susceptibility gene for familial breast cancer based on family history. Women with a personal history of breast cancer were included (six women).

Interventions and comparators

Surveillance depended on the age group of the women. All ages had CBE every six months. Mammography was annual and commenced at 25 years with bilateral one view, and then increased to bilateral double view from 30 years and above. Annual US was performed alone from 20-25 years, then with XRM from 25-35 years, then six months after XRM from 35-40 years and above 40 years only if requested by the radiologist. The US and MRI results will be reported in subsequent chapters. The method of classifying the images was not documented. It was not reported whether the radiologists interpreting the images were blinded to the results from other modalities of surveillance. Follow-up was not documented. The study was conducted over a seven-month period but the dates were not given. It is unclear if this work may have been related to the study by Podo et al. (2002). Comparisons were made between the different modalities of screening.

Outcomes

Cancer detection rate

Four breast cancers were detected by the overall surveillance strategy, including US and MRI. This gives a detection rate of 170 per 1,000 women screened. Three were detectable by CBE (130 per 1,000 women screened) but none of the tumours were detected by XRM.

Measures of accuracy

No measures of accuracy were calculated in this study.

Tumour characteristics

All four tumours were invasive. Only two tumours had the size recorded and these were 10mm and 30mm. No record of the lymph node status was documented. There was no stratification of tumour characteristics by modality of surveillance.

Interval tumours

No interval tumours were documented.

In summary, this study suggests that CBE is more effective than XRM for the surveillance of women at very high risk of breast cancer, i.e. mostly mutation carriers with a high proportion having a personal history of breast cancer. The results are extremely limited by the very small sample size, small number of tumours detected and the lack of detail documented in the publication. The study focuses on very high-risk women and may not be generalisable to all women at high risk of breast cancer.

Kriege et al. (2004)

Study sample

This prospective cohort study recruited 1,909 women from six familial cancer clinics in the Netherlands. The mean age was 40 years with a range of 19-72 years. Women younger than 25 were only included if they had a family history of breast cancer being diagnosed before 30 years of age. Risk stratification was performed according to the Claus tables. Women were included if they had a cumulative lifetime risk of breast cancer of 15 per cent or higher. Within the group, 358 women were proven mutation carriers (276 BRCA1 and 77 BRCA2), one woman had both mutations, two women had PTEN mutations and two had TP53 mutations. Women were divided in to three groups, mutation carriers (50% to 85% lifetime risk), a high-risk group (30% to 49% lifetime risk) and a moderate-risk group (15% to 29% lifetime risk). Women were excluded if they had symptoms suggestive of breast cancer or a personal history of breast cancer.

Interventions and comparators

Surveillance consisted of six-monthly CBE and annual XRM and MRI. The mean follow-up was 2.7 years with a range of 0.1-3.9 years. There was a total of 5,249 women years. The results of MRI will be reported in a subsequent chapter. The interpretation of MRI and XRM images was performed blind to the results of the other, but it does not report if radiologists were blinded to the results of CBE. Comparisons were made between the risk groups and the modalities of surveillance. Additional comparisons were made of the tumour characteristics between the study group and two control groups. The first control group was from all women who had breast cancer diagnosed in 1998 in the Netherlands using data from National Cancer Registry. The second control group comprised unselected patients who had received a diagnosis of breast cancer between 1996 and 2002 and who were participating in a prospective study of the prevalence of gene mutations. Both control groups were matched for age with the patients in the study group in five-year categories. BIRADS was used to classify the images and the results were presented according to different BIRADS cut-off points.

Outcomes

Cancer detection rate

Fifty-one tumours arose overall during this study. Forty-five of these were detected by the entire surveillance strategy, including MRI. The cancer detection rates were documented, including the interval cancers. The recalculated overall rate, without interval cancers, was 23 per 1,000 women under surveillance (8.6 per 1,000 women years). The cancer detection rates by risk group were 53 per 1,000 women under surveillance (21.9 per 1,000 women years) for mutation carriers, 14 per 1,000 women under surveillance (5 per 1,000 women years) for the high risk group and 22 per 1,000 women under surveillance (7.8 per 1,000 women years) for the moderate risk group. Three tumours were detected by CBE (1.5 per 1,000 women under surveillance). With a BIRADS cut-off of 4 and above, there were 11 tumours detected by XRM (6 per 1,000 women under surveillance) and with a BIRADS cut-off of 3 and above, there were 18 tumours detected by XRM (9 per 1,000 women under surveillance).

Sensitivity

The measures of accuracy were calculated from 45 tumours. This figure was derived from the 51 tumours that arose, minus one Hodgkins lymphoma, and minus five more tumours with missing data or follow-up. There were four interval cancers within these 45 tumours, but it is not documented by which modalities the interval tumours were included under and therefore it is not possible to calculate the following results without the interval tumours.

The sensitivity of CBE was 6.7 per cent (95% CI, 1.4 to 18.3%). The sensitivity of XRM at a BIRADS cut-off of 4 and above was 24.4 per cent (95% CI, 12.9 to 39.5) and at a BIRADS cut-off of 3 and above was 40 per cent (95% CI, 25.7 to 55.7%).

Specificity

The specificity of CBE was 99.9 per cent (95% CI, 99.8 to 99.9%). The specificity of XRM at a BIRADS cut-off of four and above was 99.7 per cent (95% CI, 99.4 to 99.8%) and at a BIRADS cut-off of 3 and above was 95 per cent (95% CI, 94.3 to 95.6%).

PPV

The PPV of CBE was 50 per cent (95% CI, 11.3 to 88.2%). The PPV of XRM at a BIRADS cut-off of 4 and above was 47.8 per cent (95% CI, 26.8 to 69.4%) and at a BIRADS cut-off of 3 and above was 8 per cent (95% CI, 4.8 to 12.3%).

AUC

The AUC for XRM was calculated as 0.686. The AUC was not calculated for CBE.

Tumour characteristics

There were six DCIS lesions and 44 invasive tumours overall. Of the invasive tumours, 43.2 per cent (19/44) were less than 10mm in size, 31.8 per cent (14/44) were 10-20mm in size and 25 per cent (11/44) were greater than 20mm in size. The number of tumours less than 10mm in size was significantly higher in the study cohort than the National Cancer Registry control group ($p<0.001$) and the genetic study control group ($p=0.04$). Lymph nodes were negative in 66.7 per cent (28/42) of the study cohort. This was also significantly higher in the study cohort than the number of node negative tumours in the National Cancer Registry control group ($p<0.001$) and the genetic study control group ($p=0.001$). There were no statistically significant differences in proportion of DCIS tumours, size of tumours and lymph node status between the different risk groups in the study. The tumour characteristics were not stratified by modality of surveillance.

Interval tumours

There were four interval tumours within the 45 tumours used for the analysis. Their size ranged from 4-45 mm and three out of the four were lymph node negative.

In summary, this study suggests that surveillance with XRM is more sensitive than CBE for detecting tumours in this cohort of women at high risk of breast cancer. The study also demonstrates that the tumours detected in the women under surveillance had statistically significantly more favourable tumour characteristics, i.e. tumour size and lymph node status, than those detected in the two control groups of women not receiving surveillance. The interval tumours cannot be commented on as they are representative of the surveillance strategy as a whole (CBE, XRM and MRI) and cannot be related to separate modalities. The results and conclusions of MRI testing in this study are discussed in a subsequent chapter.

Warner et al. (2004)

Study sample

This prospective cohort study recruited 236 female BRCA1 and BRCA2 mutation carriers from familial cancer centres in southern Ontario and Montreal in Canada. There were no age restrictions and the mean age at first surveillance was 46.6 years with a range of 25-65 years. Risk stratification was performed by all participants being BRCA1 or BRCA2 mutation carriers. This was, therefore, a very high risk group, 31 per cent of whom were of Ashkenazi Jewish descent. In addition, 30 per centhad a personal history of breast cancer. Exclusion criteria were: a past history of unilateral breast cancer if the contralateral breast was not intact, pregnant or lactating women, history of bilateral breast cancer currently undergoing chemotherapy or known to have metastatic disease and women weighing over 91kg (technical reasons). Thirty-one women left the study before completing three rounds, 16 underwent bilateral mastectomy, three were too large for MRI machine, three stopped due to pregnancy, four developed metastatic cancers, four were lost to follow-up and one did not wish to continue participating.

Interventions and comparators

Surveillance consisted of biannual CBE and annual XRM, US and MRI, all performed on the same day. The MRI and US results will be discussed in a subsequent chapter. Surveillance commenced at least one year after the woman's last mammogram. CBE was coded as normal, suggestive of benign disease, indeterminate, or suspicious of malignancy. Indeterminate CBE exams were repeated after three months. MRI was performed with 1.5T magnet (Signa, General Electrical Medical Systems). US used a 7.5MHz transducer and the first seven patients did not receive US. All participants underwent the first screen, but only 58 per cent had the second and 36 per cent the third. BIRADS was used to classify the images and scores of 4 or 5 were biopsied. Each imaging study was read and scored independently by a different radiologist who specialized in breast imaging, and radiologists were blinded to the results of CBE. All patients were followed up for a minimum of one year after their last surveillance examination. Comparisons were drawn between different modalities of surveillance.

Outcomes

Cancer detection rate

Twenty-two cancers were detected overall in 21 women. Seven of these women had a past history of breast cancer. This gives an overall cancer detection rate, including US and MRI, of 93 per 1,000 women under surveillance. Two were detected by CBE (8 per 1,000 women under surveillance) and eight by XRM (34 per 1,000 women under surveillance). Two tumours were detected by XRM alone.

Sensitivity

All the measures of accuracy in the paper are presented individually for each year of surveillance. These results have been combined to give overall results for the three rounds of surveillance. There were not enough raw data to calculate measures of accuracy for CBE except for the sensitivity.

The sensitivity of XRM was 36 per cent (95% CI, 17.1 to 59.3%) and of CBE was 9 per cent (95% CI, 1 to 29%)

Specificity

The specificity of XRM was 99 per cent (95% CI, 98.7 to 99.9%).

PPV

The PPV of XRM was 88 per cent (95% CI, 51.7 to 99.7%).

NPV

The NPV of XRM was 97 per cent (95% CI, 94.8 to 98.3%).

AUC

The AUCs for XRM was 0.77. The AUC for CBE is also given at 0.48 and the combination strategy of CBE and XRM was 0.77. There were no confidence intervals documented for the AUCs.

Tumour characteristics

Sixteen tumours were invasive and six were DCIS. The mean size of the invasive tumours was 11mm at the first surveillance round and 13mm at the second round. Fifteen cases had lymph node sampling and two were node positive. The tumour characteristics are not documented as stratified by modality of surveillance.

Interval tumours

One interval tumour was detected seven months after a third screen. Retrospectively, this was visible on XRM at the last surveillance visit.

Mortality

All 22 patients with tumours were still alive and disease-free at the time of publication of the article.

In summary, this study suggests a superior efficacy and accuracy of XRM to CBE in detecting early breast cancer in women at high risk of breast cancer. The results of this study are limited to the very high risk population of women who are proven BRCA1 or BRCA2 mutation carriers, including those

with a personal history of breast cancer. It is therefore not generalisable to all women with an increased risk of breast cancer due to a family history. Further studies with larger numbers and longer follow-up, and including women of other risk groups are required.

Murday et al. (2004)

Study sample

This retrospective cohort study recruited 192 women from cancer clinics in the South Thames region, UK. All the women were less than 50 years of age at their first appointment. The mean age of surveillance is not provided, but the mean age at diagnosis was 39.9 years with a range of 29-48 years. Risk stratification was performed according to national guidelines in the UK. For surveillance, women required at least one first-degree relative with breast cancer diagnosed before 40 years of age, or two relatives diagnosed at less than 60 years of age, one of which had to be a first-degree relative. The risk status of the women was then recalculated using the Claus tables, a cyrillic computer program and modified by Bayes theorem. This was to take into account the number of unaffected relatives in the family, without which the risk is overestimated. Other inclusion criteria were: having attended the cancer clinics in this region prior to 1996, to have received surveillance in the same hospital in which they had genetic counselling, and to have no past history of breast cancer.

Interventions and comparators

Surveillance consisted of annual XRM and CBE from 35 years. Women were also encouraged to perform monthly SBE. XRM was two-view. Twenty-three women had US instead of XRM. Nineteen of these women were less than 35 years and four chose US over XRM. Some of the US screens were surveillant but some were also diagnostic. There were 280 person years of follow-up in total. BIRADS was used to classify the images, but the cut-off for an abnormal image was not documented. It was not reported whether radiologists interpreting the XRM images were blinded to the results of CBE. Comparisons were made between different levels of risk and also between XRM and CBE. Information was also collected on a control group of sisters of participants. However, this comparison was not successful as there was a poor response from this group, they were difficult to age match and there was a lack of information on any surveillance that they might be receiving.

Outcomes

Cancer detection rate

Six cancers were detected by surveillance. Three of these were detected at the prevalent round and three during the incident rounds. This gives a cancer detection rate of 31 per 1,000 women under surveillance or 15.6 per 1,000 women under surveillance for both the prevalent and incident rounds. All six tumours were visible on XRM and half of these were also palpable at CBE. The cancer detection rate by modality was therefore 31 per 1,000 women under surveillance for XRM and 15.6 per 1,000 women under surveillance for CBE. The study results were confusing as the interval tumours were considered to be part of the surveillance protocol and therefore detected by surveillance, but in this report they have been separated out. The comparison of different risk groups found that the majority of cancers were detected in the high-risk and mutation carrier groups. All cancers were detected in the groups that had greater than a 20 per cent calculated chance of having a high-risk gene in their family.

Sensitivity

The sensitivity of XRM and CBE was reported as 78 per cent (95% CI 40-97). However, this calculation included one of the false negatives in the numerator. The justification for this was that it was retrospectively visible on the last surveillance screen. If the sensitivity is recalculated with this tumour in the denominator only, then it is 67 per cent (95% CI 30-93). The sensitivity of CBE alone is 33 per cent (95% CI 7-70) and the sensitivity of XRM alone is 67 per cent (95% CI 30-93). Therefore there were no tumours detected by CBE that were not visible on XRM also. There was no statistically significant difference in sensitivity between XRM and CBE with CBE alone. If measures of accuracy are looked at by risk groups, the numbers are very small. There was one tumour in a mutation carrier and this was not detected by CBE or XRM. There were seven tumours in the 'high' risk group and the sensitivities for CBE and XRM within this group are 43 per cent (95% CI 10-82) and 86 per cent (95% CI 42-100) respectively. There was only one tumour in the 'medium' risk group and this was not detected by either modality. There were no tumours in the 'low' risk group.

Specificity

The specificity for CBE and XRM combined was reported as 84 per cent. There were not enough raw data to be able to reproduce these figures.

PPV

The PPV for CBE and XRM combined was reported as 9 per cent. There were not enough raw data given to be able to reproduce these figures.

Tumour characteristics

Of the tumours detected by surveillance, four were invasive, one was *in situ* and one had both invasive and *in situ* components. Tumour size was not recorded for four tumours and was 30mm and 90mm in the other two tumours. Two tumours had lymph node spread. The three palpable tumours were invasive. Two of the three non-palpable tumours were *in situ* and one was invasive.

Interval tumours

Three interval tumours were detected by BSE, two of which were not visible by XRM. Two of the interval tumours were invasive, one was *in situ* and none of them were node positive.

In summary, this study suggested that surveillance was most warranted in women with over a 20 per cent chance of having a high-risk gene in their family. More cases were detected by XRM than by CBE, but the comparison was limited by the small number of cases. There was still a high rate of interval tumours in this study, which suggested that more intensive surveillance or additional modalities may be required in the high-risk group. In this study, only one third of women actually received the recommended interval between their prevalent and first-incident round of surveillance. This is a further limitation of this study, although this reflects the real-life situation. The number of interval cancers may have been lower if the surveillance intervals had been adhered to.

Banks et al. (2004)

Study sample

This prospective cohort study recruited 122,355 women from the Million Women study. All women were attending the NHSBSP at 10 breast screening units in the UK. Of this cohort, 10,959 women had a family history of breast cancer. The women were all aged between 50 and 64 years of age. The mean age of the cohort was not given. Risk stratification was not described other than that they had a first-degree relative (mother or sister) affected by breast cancer. All women had no history of cancer other than non-melanoma skin cancer. Women were excluded if they had a positive screen but no diagnosis of breast cancer, and were recalled for screening before three years, as they were considered to no longer be on the routine surveillance.

Interventions and comparators

Surveillance consisted of three-yearly XRM. Clinical breast examination does not appear to have been performed in this cohort. There was no mention of the mammographic views used. The NHSBSP used single-view (mediolateral oblique) XRM initially, but some centres used more than one view and from 1995 two-view XRM was made mandatory for prevalence screens. Screens were classified as abnormal if the women were recalled for further investigation. Comparison of the accuracy of XRM was made between women with and without a family history of breast cancer in the NHSBSP. This was not the main focus of the study and other factors examined were HRT use, previous breast operations, previous use of OCP, regular strenuous exercise, smoking, alcohol consumption, menopausal status, age, parity, tubal ligation and body mass index. The women were followed up for 12 months. If diagnosis, confirmed by histology, occurred within three months of screening it was considered to have been screen diagnosed. If diagnosis occurred over three months from screening it was considered not to be screen detected.

Outcomes

Sensitivity

The overall sensitivity for XRM in this study was 86.6 per cent. For women with a family history of breast cancer the sensitivity of XRM was 83.8 per cent (95% CI, 74.6 to 90.0%). For women without a family history of breast cancer the sensitivity of XRM was 89.4 per cent (95% CI, 86.3 to 91.9%). The *p* value for comparing the sensitivity in these two groups was 0.1. There were no raw data to verify these calculations.

Specificity

The overall specificity for XRM in this study was 96.8 per cent. For women with a family history of breast cancer, the specificity was 97.2 per cent (95% CI 96.9 to 97.5%). For women without a family history of breast cancer, the specificity was 97.3 per cent (95% CI 97.2 to 97.4%). The *p* value for comparing the specificity in these two groups was 0.5. There were no raw data to verify these calculations.

In summary, this study shows that there is no significant difference in the accuracy of XRM between women with and without a family history of breast cancer between the ages of 50 and 64 years in the NHSBSP. Factors that were shown to significantly affect the accuracy of XRM were HRT use, previous breast surgery and body mass index. This study is not representative of all women with a family history of breast cancer as it does not include women less than 50 years of age so is therefore less relevant to the present review. It is in these younger women that breast density is thought to affect the accuracy of XRM, and this is of relevance to women at high risk of breast cancer as they are the women who require screening at a younger age. The results of this study would also be affected by the follow-up period being only 12 months and the screening interval being three years. Interval cancers would arise and lower the sensitivity had the cohort been followed for the entire screening interval.

Halapy et al. (2004)

Study sample

This retrospective cohort study recruited 143,574 women from the Ontario Breast Screening Programme (OBSP) in Canada, of which 21,749 had a family history of breast cancer. The women were all over 50 years of age. The mean age was not given. Risk stratification was performed according to methods specific to this study and relied upon self-reported information from the women. A family history was defined as having at least one affected first-degree relative. More details are presented in **Table 11**. There were 14,325 women with a 'moderate' family history and 7,424 women with a 'strong' family history. Other inclusion criteria were: having a screen between January 1996 and December 1997, residing in Ontario, no past history of breast cancer or augmentation mammoplasty and being free from acute breast symptoms. Exclusion criteria were: having received only CBE surveillance and any screens other than the first one performed in this time period.

Interventions and comparators

Surveillance consisted of biennial XRM and CBE. Some women at high risk were recalled annually, but the number of these women was not reported. XRM was two-view. The follow-up period is not specified in this study, but two years of data appear to have been included. Screening was classified as abnormal if the women were referred on for further investigations. It was not reported whether radiologists interpreting the XRM images were blinded to the results of CBE. Comparisons were drawn between the different risk categories, age groups and modalities of screening.

Outcomes

Cancer detection rate

Sixty-five tumours were detected in women with a 'strong' family history, 133 in those with a 'moderate' family history and 788 in those with no family history. The overall cancer detection rate for women with a 'moderate' and a 'strong' family history was 9.1 per 1,000 women screened (9.1 per 1,000 screens). The respective cancer detection rates (and 95% CI) were 10.6 (6.7 to 14.5), 9.7 (7 to 12.4) and 8.3 (7.5 to 9.1) for prevalent screens, and 7.7 (5.2 to 10.2), 9.0 (7.1 to 10.9) and 5.3 (4.8 to 5.8) for incident screens. The cancer detection rate is also presented for the prevalent and incident

screens by mode of detection, CBE or XRM (see **Table 11**). This showed that XRM was more effective at detecting tumours than CBE in both prevalent and incident screens and in most risk groups. Only rates are presented and no raw figures are given. There are no measures of significance presented for CBE versus XRM and due to the lack of raw data these cannot be calculated. Cancer detection by modality is also stratified by risk group and incident or prevalent screen. In women at increased risk, cancer detection for CBE ranges from 3.2 to 6.1 per 1,000 screens. For XRM it ranges from 7.3 to 9.9 per 1,000 screens. Comparisons across risk groups showed that cancer detection rates increased with level of risk and with increasing age.

PPV

The only measure of accuracy examined in this study was the PPV. The results were stratified by degree of family history, modality of screening and whether it was a prevalent or incident screen. The PPV was shown to generally be higher in those with a 'moderate' and 'strong' family history of breast cancer, than in those without such a history. This is not surprising given the increased prevalence in the higher risk groups. The PPV was also higher for XRM than CBE across all family history groups, and was shown to increase with increasing age of all family history groups. For XRM in the prevalent screen, the PPV (and 95% CI) were 8.7 per cent (7.8 to 9.5%), 9.9 per cent (7.2 to 12.6%) and 12.0 per cent (7.7 to 16.4%) for the no family history, 'moderate' family history and 'strong' family history groups respectively. For CBE in the prevalent screen, the PPV(and 95% CI) were 6.0 per cent (5.2 to 6.85), 6.4 per cent (3.9 to 8.8%) and 7.3 per cent (3.9 to 10.8%) for the no family history, 'moderate' family history and 'strong' family history groups respectively. There were not enough raw data to calculate *p* values for the comparison of PPV between CBE and XRM and none were documented for this comparison.

Tumour characteristics

In women with a 'strong' family history, 15 (23%) tumours were *in situ* and 50 (77%) were invasive. Of the invasive tumours, 88 per cent were less than 20mm in size and 68 per cent were node negative. In women with a moderate family history, six (4.5%) tumours were *in situ* and 127 (95.5%) were invasive. Of the invasive tumours, 87 per cent were less than 20mm in size and 81 per cent were node negative. The only significant difference in tumour characteristic was that women with a 'moderate' family history of breast cancer had fewer *in situ* tumours than those with no family history. The tumour characteristics were not stratified by modality of screening.

In summary, this study suggested that XRM was more effective than CBE in detecting early breast cancer in women at high risk of breast cancer over the age of 50 years. However, there were no measures of statistical significance given and not enough raw data documented to calculate them. This study also examined HRT as a confounder and discovered that current HRT use removed the association of family history and cancer detection rate. No comment was made on interval tumours in this study or screening intervals. The external validity of this study to all women at a high risk of breast cancer is restricted due to the exclusion of women less than 50 years of age.

Halapy et al. (2005)

Study sample

This retrospective cohort study recruited 115,460 women from the Ontario Breast Screening Programme (OBSP) in Canada, of whom 16,813 had a family history of breast cancer. The women were a subgroup from the previous study by Halapy et al. (2004). The age range of this subgroup was specified as greater than 50 years and less than 69 years. The mean age of this cohort is not given. Risk stratification was performed according to methods specific to this study and relied on self-reported information from the women. A family history was defined as having at least one affected first-degree relative. More details are presented in **Table 11.** There were 11,025 women with a 'moderate' family history and 5,788 women with a 'strong' family history. Other inclusion criteria were: having a screen between January 1996 and December 1997, residing in Ontario, no past history of breast cancer or augmentation mammoplasty and being free from acute breast symptoms. The exclusion criteria were: having received only CBE surveillance and any screens other than the first one performed in this time period.

Interventions and comparators

Surveillance consisted of biennial XRM and CBE. Some women at high risk were recalled annually, but the number of these women was not reported. XRM was two-view. The follow-up period was 12 months post-screening. Screens were classified as abnormal if women were referred on for further investigations. It was not reported whether radiologists interpreting the XRM images were blinded to the results of CBE. Comparisons were drawn between the different risk categories, age groups and modalities of screening.

Outcomes

Cancer detection rate

The findings were similar to the 2004 study, namely that the cancer detection rate increases with risk category and with age and that XRM has a better detection rate than CBE. Overall in this subgroup there were 16,813 women with either a 'moderate' or 'strong' family history'. All of them received XRM and 16,712 had undergone CBE. In total there were 154 cancers detected. This gives a cancer detection rate of 9.1 per 1,000 women screened. There were 49 tumours (3 per 1,000 women screened) detected by CBE and 105 (6 per 1,000 women screened) detected by XRM.

Sensitivity

Sensitivity was stratified by family history, age and screening modality. Sensitivity was found to be lower in women with a family history when compared to those without such a history, but this result was not statistically significant. Sensitivity was consistently higher for XRM than CBE across all family history groups. For XRM, the sensitivity was 89.4 (95% CI, 82.9 to 96.0%) and 76.3 (95% CI, 62.8 to 89.8%) for those with a 'moderate' family history and 'strong' family history respectively. For CBE, the sensitivity was 40.0 (95% CI, 29.6 to 50.4%) and 40.5 (95% CI, 24.7 to 56.4%) for the 'moderate' family history and 'strong' family history groups respectively.

Specificity

Specificity was also stratified by family history, age and screening modality. Specificity differed very little according to family history and was much the same between the two modalities For XRM, the specificity was 93.9 (95% CI, 93.4 to 94.3) and 94.6 (95% CI, 94.0 to 95.2%) for the 'moderate' family history and 'strong' family history groups respectively. For CBE, the specificity was 94.8 (95% CI, 94.4 to 95.2%) and 94.0 (95% CI, 93.4 to 94.6%) for the 'moderate' family history and 'strong' family history groups respectively.

Interval tumours

The interval cancers were defined as invasive tumours arising within 12 months of the screening date. There were 61 interval tumours altogether. Forty-eight tumours were in the women with no family history, seven were in the 'moderate' risk group and six were in the 'strong' family history group. This gives interval cancer rates (and 95% CI) of 4.9 (3.5 to 6.3), 6.4 (1.7 to 11.1) and 10.5 (2.1 to 18.8) per 10,000 person years respectively. However, the difference between the risk groups was not statistically significant. The interval cancer rates were higher for surveillance by CBE (46.8 per 10,000 person years) than surveillance by XRM (8.2 per 10,000 person years) in the 'moderate' family history group ($p<0.001$). The interval cancer rates were similarly higher for surveillance by CBE (38.6 per 10,000 person years) than surveillance by XRM (15.7 per 10,000 person years) in the 'strong' family history group ($p = 0.02$). This reflects the more effective cancer detection by XRM, as any cancers not detected by one modality were considered interval cancers in addition to those that arose during the 12 months follow-up.

In summary, this study reiterates the findings of the previous study (Halapy et al. 2004) for women in a restricted age group (50-69 years of age) and provides more raw data to support these findings. However, no measures of statistical significance were calculated for the differences in accuracy between CBE and XRM. The interval cancer rates reinforce the superior efficacy of XRM to CBE in breast cancer screening. The screening interval in the OBSP was biennial for most women in the programme and the interval cancer rate, especially in women at 'moderate' and 'strong' risk, was high even after 12 months follow-up. This suggests that annual screening may be more effective in these women. The external validity of this study to all women at a high risk of breast cancer is restricted due to the exclusion of women less than 50 years of age.

Kuhl et al. (2005b)

Study sample

This prospective cohort study recruited 529 women from high breast cancer risk clinics in a single hospital in Germany. There was no age restriction and the mean age of the whole cohort was 41.7 years (range 27-59 years). Risk stratification was performed according to the Consortium on Familial Breast and Ovarian Cancer of the German Cancer Aid. All participants had a greater than 20 per cent lifetime risk of breast cancer. In women who did not have a personal history of breast cancer, the Claus tables were also used to stratify risk. Women with a personal history of breast cancer were included provided they had not had bilateral mastectomy, had not had chemotherapy within the last 12 months and had no metastases;139 women had a personal history. Being asymptomatic was another inclusion criterion.

Interventions and comparators

Surveillance consisted of biannual CBE and US and annual XRM and MRI. If abnormalities found on CBE or US at the round without XRM or MRI, these additional modalities were used to further investigate this. The results of US and MRI screening will be discussed in subsequent chapters. Surveillance commenced at 30 years, or five years before the youngest family member affected with the disease. In the first two years, women under 30, or 30-39 years with dense breasts, did not receive XRM, but this was subsequently abandoned and all women received XRM. These data were not included in the calculation of accuracy measures. MRI of both entire breasts was performed on a 1.5T system (NT/INTERA; Philips, Best, the Netherlands). US was performed with 7.5-13MHz probes. Each imaging study was read and scored independently by a different radiologist who had substantial experience with the respective imaging technique. The radiologists were informed about the clinical findings from CBE and the risk status of the patient but were blinded to the results of the respective other imaging modalities. BIRADS was used to classify the images and scores of 4 or 5 went for biopsy. The mean follow-up time was 5.3 years with a range of 2-7 years). The number of total annual surveillance rounds for which data on all three imaging modalities was available was 1,452, and this was used in the calculation of accuracy measures. Comparisons are made between the three risk groups and the different modalities of surveillance.

Outcomes

Cancer detection rate

A total of 43 tumours arose in 41 patients during the study period. Forty of these were detected by imaging. That gives a cancer detection rate for the overall surveillance strategy, including US and MRI, of 76 per 1,000 women under surveillance. Eleven, i.e. 25 per cent of these patients had a prior history of breast cancer. CBE identified only one tumour (2 per 1,000 women under surveillance) which was also detected on imaging. Fourteen tumours were detected by XRM (26 per 1,000 women under surveillance) and only one was diagnosed by XRM that was not diagnosed by MRI.

Sensitivity

The sensitivity for XRM was 32.6 per cent (95% CI, 19 to 48.5%) and for CBE was 2.3 per cent (95% CI, 0.1 to 12.3%).

When stratified by risk groups, XRM became less sensitive as the lifetime risk of breast cancer increased. The sensitivity was 50 per cent in those with a 20 per cent lifetime risk, 25 per cent in those with a 21-40 per cent lifetime risk and 25 per cent for the mutation carrier group. The average age of the women decreased as the risk of breast cancer increased, and this may have contributed to the decreasing sensitivity. However, the difference in age is only small and is unlikely to account for the whole effect. The mean ages and age ranges in the three groups were 43.8 years (35-59 years), 40.3 years (31-57 years) and 38.9 years (27-51 years) in the lifetime risk of 20 per cent, 21-40 per cent and the mutation carriers respectively. A more aggressive nature of tumours or a different histopathology i.e. prominent pushing margins, are other factors that may have contributed to the decrease in sensitivity in the highest risk women.

Specificity

The overall specificity for XRM was 96.8 per cent (95% CI, 95.7 to 97.7%). There were insufficient data to calculate the specificity of CBE.

Stratification by risk group or by a past history of breast cancer did not appear to affect the specificity

PPV

The overall PPV for XRM was 23.7 per cent (95% CI, 14 to 37%). There were insufficient data to calculate the PPV of CBE.

Stratification by risk group or by a past history of breast cancer did not appear to affect the PPV either.

Tumour characteristics

Thirty-four tumours were invasive and nine were DCIS. The tumour characteristics are presented by mode of detection. Of the 14 cancers detected by XRM, 10 were invasive. The invasive cancers had a mean size of 13.2mm and four were node positive. The tumour characteristics were not stratified by CBE.

Interval tumours

The interval tumour rate is given as 2 per cent in this cohort. It is unclear if this is a percentage of the women under surveillance or of the tumours that arose. It was also documented that there was one interval cancer that arose between surveillance rounds. However, it was reported that 40 of the 43 cancers were detectable by imaging, which would suggest three interval cancers. These figures were reported in an unclear manner.

In summary, this study suggests that the addition of XRM to CBE does improve the sensitivity of surveillance of women at high risk of breast cancer. The data on interval tumours is somewhat unclear in its documentation. This study included women at high risk who had a personal history of breast cancer, but the majority of the results were not significantly different if stratified by personal history.

Elmore et al. (2005)

Study sample

This matched case-control study recruited 3,752 women from six health plans in five states in the USA. The cases were 1,351 women who had died from breast cancer, comprising a random sample of all those that were eligible. The controls were 2,501 live women matched to the cases for age and risk level. The age restrictions were 40-65 years of age. This was split in to two age categories, 40-49 years and 50-65 years. The mean age was not given. Risk stratification was performed according to family history or a breast biopsy noted in the medical records before the index date, i.e. the date of first suspicion of breast abnormalities in the cases and the same date used in matched case controls. There were 411 cases and 599 controls out of all the participants who had an increased risk of breast cancer. More inclusion details are presented in **Table 11.** Exclusions for cases were: medical chart information not being available for review or not being reviewed, and if no eligible control was found.

Interventions and comparators

The occurrence of surveillance in the cases and controls was examined. Surveillance consisted of XRM and CBE, but it was unclear what surveillance intervals were used or when surveillance had commenced. The results were examined by risk category, age groups and by modality of surveillance (XRM, CBE or both).

Outcomes

The odds ratios (OR) for breast cancer mortality were all generally less than one, indicating that those who received surveillance were less likely to die of breast cancer. However, all but one of these ORs had 95 per cent confidence intervals that included the value of 1 and were therefore not significant. The only exception was the OR for surveillance by CBE in women at an increased risk aged 50-65 years, which is 0.61 (95% CI, 0.39 to 0.97). The OR for women at increased risk aged 40-65 years under surveillance with either CBE or XRM was 0.78 (0.50 to 1.23). The OR for women at increased risk aged 40-65 years under surveillance with CBE alone was 0.80 (0.59 to 1.08). The OR for women at increased risk aged 40-65 years under surveillance with XRM alone was 1.05 (0.8 to 1.39).

In summary, there was no statistically significant association observed between surveillance by CBE, XRM or both these modalities, and breast cancer mortality except for CBE in one particular age and risk group. However, the design of this study limits the reliability of these findings. There was potential for confounding and bias, particularly misclassification and selection bias. The sample of women at high risk was also small and may limit the power of these results to show a significant difference. The external validity is limited by the risk stratification which relied on previous biopsy as well as family history and on the lack of information on the actual surveillance women received.

Gui et al. (2006)

Study sample

This prospective cohort study recruited 1,132 consecutive women attending a breast diagnostic unit in London, UK. There were no age restrictions and the mean age of entry to the study was 54 years, 49 years and 47 years in the 'standard', 'moderate' and 'high' risk groups respectively. Risk stratification into these three groups was performed by criteria based on guidance provided by the National Institute for Clinical Excellence (NICE). These criteria related purely to family history. 'Standard' risk was defined as less than 17 per cent, 'moderate' risk as between 17 per cent and 30 per cent, and 'high' risk as greater than 30 per cent lifetime risk of breast cancer. There were 192,803 and 137 women in these groups respectively. There was no mention of whether they included women with a past history of breast cancer. Only 406 women were completely asymptomatic and had no clinical signs.

Interventions and comparators

Surveillance for women at 'moderate' or 'high' risk consisted of annual CBE and XRM from the age of 35 years. US was sometimes used diagnostically if there was uncertainty. After 50 years of age, women at 'moderate' risk underwent surveillance every 18 months while women at 'high' risk continued to undergo surveillance annually until 69 years. The 'standard' risk women were discharged to the NHSBSP unless there was any clinical indication for follow-up. Follow-up was at least one 12-month or 18-month surveillance interval, depending on the surveillance being received. The system of classification of images used was not documented. It was not reported whether radiologists interpreting the XRM images were blinded to the results of CBE. Comparisons were made between the different risk categories, the modes of detection and the cancer detection rates in the study and in the NHSBSP.

Outcomes

Cancer detection rate

Seven cancers were detected during the active study period. This gives a cancer detection rate of 6.2 per 1,000 women under surveillance. There were two tumours in the 'standard' risk group (10.4 per 1,000 women under surveillance), three in the 'moderate' risk group (3.7 per 1,000 women under surveillance) and two in the 'high' risk group (14.6 per 1,000 women under surveillance). Four tumours were detectable by CBE and all were detectable by XRM (5.3 and 6.2 per 1,000 women under surveillance respectively). The cancer detection rate overall and in each risk group was equivalent or greater than that of the NHSBSP (3.8 per 1,000 women screened).

Sensitivity

The sensitivity of XRM was 86 per cent (95% CI, 42-100%) and the sensitivity for CBE was 14 per cent (95% CI 0-58).

Specificity

The specificity of XRM was 99 per cent (95% CI 98-99).

PPV

The PPV of XRM was 38 per cent (95% CI 15-65).

NPV

The NPV of XRM was 99 per cent (95% CI, 99.3 to 99.9%)

There is not enough raw data to calculate the specificity, PPV or NPV for CBE alone.

Tumour characteristics

The tumour characteristics were presented for the surveillance-detected and interval cancers combined, plus eight surveillance-detected tumours that occurred outside the active study period. There were 13 invasive and four *in situ* tumours. The median invasive tumour size was 15mm with a range of 7-8mm and the median DCIS size was 4mm with a range of 2-30mm). Two of the 13 invasive tumours had lymph node spread. Tumour characteristics were not stratified by modality of surveillance.

Interval tumours

There were two interval tumours, one in the 'standard' risk group and one in the 'moderate' risk group.

Mortality

There were four deaths overall in the cohort, but none was related to breast cancer.

In summary, this study shows that surveillance with XRM and CBE in women at high risk of breast cancer is effective in detecting early breast cancers, based on an equivalent cancer detection rate to that of accepted screening programmes for women over 50 years of age. It also reinforces the need for XRM in addition to CBE. The study design is confused by the use of US scanning, which is not thoroughly discussed and may have confounded the results. It was unclear if the study was adequately powered to determine equivalence.

Comment is made that the cancer detection rate is still high among women in the moderate risk category, with the mean age of diagnosis being 55.6 yreas, and that this suggests that they should continue intensive surveillance into their 50s. However, little comment is made about the cancer detection rate in the standard risk women. The risk in this group is even greater, with the mean age of diagnosis being 58.8 years, and there is still one interval cancer in this group. However, this may be a product of selection bias in this group. It is stated in the methodology that the standard risk women in this study were those that had not yet been discharged back to the NHSBSP and this was usually because of a clinical indication for continued follow-up. Therefore, it can probably be assumed that these women were not representative of women at standard risk in general.

Maurice et al. (2006)

Study sample

This prospective cohort study recruited 3,016 women at high risk of breast cancer from a family history clinic in Manchester, UK, including 32 known mutation carriers. The women were all aged less than 50 years. The mean age of surveillance was not given but the mean age at diagnosis was 45 years. Risk stratification was performed according to the Claus tables. Women were included if their lifetime risk of breast cancer was greater than one in six. It was not mentioned if women were included if they had a past history of breast cancer, but they had to be asymptomatic.

Interventions and comparators

Surveillance was by XRM and CBE at 12-18 monthly intervals commencing at presentation to the clinic, but not usually younger than 35 years. If there were relatives affected at an early age then surveillance would commence five years before the earliest breast cancer diagnosis, but not before 30 years. XRM was two-view from 1999 onwards and one-view for surveillance prior to this. All women were offered instruction in BSE. All women were followed up for two years after the end of the active study period. The average follow-up was 3.6 years. The system of classification of images used was not documented. It was not reported whether the radiologists interpreting the XRM images was blinded to the results of CBE. Comparisons were made with cancer detection rates in the NHSBSP and also with the tumour characteristics of women, less than 50 years, presenting with symptomatic breast cancer.

Outcomes

Cancer detection rate

There were 45 screen detected cancers, 19 prevalent and 26 incident. Seventeen of these tumours were in known mutation carriers. This gives an overall cancer detection rate of 15 per 1,000 women under surveillance. It was documented that the cancer detection rate at the prevalent screen was 5.97 per

1,000, but this does not agree with the figures given (19/3,016). The cancer detection rate at the incident screen was given as 4.84 per 1,000, but this cannot be verified as it was not documented how many screens were performed over the study. These cancer detection rates were compared with women in the NHSBSP (5.5 per 1,000 and 4.6 per 1,000 for the prevalent and incident screens respectively in the NHSBSP). The cancer detection rate was not stratified by modality of surveillance.

Tumour characteristics

Of the prevalent tumours, nine were invasive (47%) and 10 were *in situ* (53%). Seventy-eight per cent of the prevalent tumours were sized less than 20mm and were sized 20-50mm in the remaining 22 per cent. Eight of the invasive tumours had node sampling and seven (88%) of these were node negative. Significantly more prevalent carcinomas were *in situ* than incident and interval tumours (p=0.013).

The characteristics of the incident and interval tumours are presented together. There were 34 invasive tumours (79%) and nine (21%) *in situ*. Twenty-four tumours (71%) were less than 20mm in size, nine (26%) were 20-50mm and one (3%) was sized over 50mm. Thirty-three had node sampling and of those, 20 (61%) were node negative.

When compared with women not under surveillance, presenting symptomatically with breast cancer, the surveillance group had less invasive tumours and more *in situ* tumours (p = <0.001), more small tumours (p = <0.001), more that were node negative (p = 0.013) and fewer breast cancer deaths (p = 0.013). However there was quite a lot of uncertainty in the data for symptomatic women. The figures used for the surveillance group in this comparison included the interval tumours. Overall, the symptomatic women had 1,000 tumours, 918 (92%) invasive and 82 (8%) *in situ.* The characteristics were only looked at for the invasive tumours. However, 213 (23%) had an unknown size, 397 (43%) had not had the grade assessed and 97 (10%) did not have the nodes sampled. In the surveillance group all had the size known, only one (2%) had not had the grade assessed and only two (4%) had not had the nodes sampled. The inclusion of these data may alter the results of the comparisons. Only mortality had no missing data from both groups.

Interval tumours

There were 17 interval cancers.

Mortality

Two of the women with prevalent tumours died and two of the women with incident or interval tumours died.

Survival

Survival curves were calculated for the surveillance and the symptomatic group and adjusted for lead time. Cox regression analysis indicated a relative hazard of death from breast cancer between the study and the symptomatic group of 0.19 (95% CI 0.07 to 0.52, p<0.001). When adjusted for lead time in the study population, the relative hazard between the two groups was 0.24 (95% CI 0.09 to 0.66, p = 0.005). The difference in disease-free survival was also significant. The relative hazard unadjusted for lead time between the two groups was 0.19 (95% CI 0.08 to 0.43, p<0.001) and the relative hazard adjusted for lead time between the two groups was 0.25 (95% CI 0.11 to 0.57, p<0.001).

In summary, this study shows that surveillance with XRM and CBE in women at high risk of breast cancer is effective in detecting early breast cancer. The cancer detection rate is equivalent to that of accepted screening programmes for women over 50 years of age. There also appears to be significantly better tumour characteristics in the women under surveillance than the symptomatic women who had not undergone surveillance, and even a significant reduction in breast cancer mortality and improved survival and disease-free survival. There are also a lot of missing data which, if present, may alter the results considerably. The only result which did not have a great deal of missing data was mortality.

Summary

Twenty-four studies were identified of relevance to the accuracy and efficacy of XRM surveillance of women at high risk of breast cancer. No RCTs have been carried out in this high-risk population as it is considered unethical not to offer screening by XRM to this high-risk population. The majority (22) of

the studies were prospective or retrospective cohort studies. Participants were under surveillance by XRM and CBE, with or without additional modalities of surveillance.

A total of 130,504 high-risk women participated in the 24 studies. Many of the studies were limited by a small sample size or by the small number of tumours detected during surveillance. There was heterogeneity between study designs, including the surveillance intervals, participants' level of risk of breast cancer, participants' age and the inclusion or exclusion of women with a personal history of breast cancer. This made it difficult to compare results across studies.

The main outcomes of the studies were the cancer detection rates and measures of accuracy. There is difficulty in comparing the detection rates across studies, as some studies have reported these as cancers per 1,000 surveillance screens and others have used cancers per 1,000 women under surveillance. As the studies have had different surveillance intervals, and therefore numbers of surveillance screens, the rates reported by 1,000 women under surveillance are not comparable across studies. Only 13 of the 24 studies documented the cancer detection rate by number of surveillance screens, or had the raw data to calculate it. Some studies presented prevalent and incident surveillance results, but others combined prevalent and incident surveillance results, even though detection rates are generally lower for incident surveillance than for prevalent surveillance. Only 11 studies calculated measures of accuracy for surveillance. The cancer detection rate and measures of accuracy are both intermediate outcome measures.

Three studies considered survival as an outcome. Two of these were not adjusted for lead time (Hou et al. 2002; Moller et al. 1999), but one was adjusted (Maurice et al. 2006) and appears to suggest a survival advantage from surveillance in this population. Eight studies considered mortality as an outcome. However, there was mostly insufficient sample size and length of follow-up to demonstrate any significant decrease in mortality (Brekelmans et al. 2001; Federico et al. 1999; Gui et al. 2001; Gui et al. 2006; Warner et al. 2004). The one matched-case control study (Elmore et al. 2005) failed to demonstrate a difference in surveillance status between women who had died of breast cancer and matched controls who had not. These findings must be interpreted with caution due to the retrospective nature of this study and the likelihood of bias and confounding. Myles et al. (2001) do suggest a reduction in mortality associated with screening in a high-risk population, but this is as a result of modelling with figures from studies in the screened population over 50 years of age. The most recent of the 24 studies (Maurice et al. 2006) does, however, appear to show a significant reduction in breast cancer mortality when comparing a surveillance population and a population that did not receive surveillance.

Three main comparisons are made in these studies in order to demonstrate a benefit of surveillance with XRM. The first is to demonstrate that surveillance with XRM has a higher cancer detection rate and is more accurate than CBE alone. Summaries of these results are presented in **Tables 7** and **8**. Overall, there is seen to be higher cancer detection rates and sensitivity with XRM surveillance compared to CBE. This is logical as tumours are only detectable by CBE once they have reached approximately 10mm in size (Hughes et al. 1999). CBE surveillance performed better than XRM in two studies (Gui et al. 2001; Trecate et al. 2003). The results of Gui et al. (2001) are misleading as 17 tumours included in the CBE category were actually detected by SBE and therefore were interval tumours. Trecate et al. (2003) was a very small study of just 23 women at very high risk of breast cancer, i.e. mutation carriers, a high proportion of whom had a past history of breast cancer and had only a four tumours detected. Therefore, the results are unreliable. The sensitivity of XRM is shown to decrease as the lifetime risk of breast cancer increases, with the lowest sensitivity in mutation carriers (Kuhl et al. 2005b).

The second comparison made was between the cancer detection rates in the studies and in established breast screening programmes for women over 50 years of age. The assumption was that if there were similar rates of detection then surveillance for women at high risk should be acceptable. However, this does not take into account the potential harms, for example radiation exposure, for high-risk women undergoing surveillance from an early age and over a longer time period. Cancer detection rates would be expected to be higher in women with a high risk of breast cancer due to the higher prevalence among this group. However, it is also postulated that cancer detection with XRM is reduced in women at high risk. This is because they require screening from a younger age when their breasts are denser, and also due to the histopathological character of their tumours, which appear more benign (Tilanus-

Linthorst et al. 2002). The eight studies which made this comparison are summarised in **Table 9.** The figures suggest equivalency of cancer detection rates to those of the breast screening programmes.

Table 7. Summary of cancer detection rates from the studies of surveillance by XRM and CBE

Study	Cancer detection rate per 1,000 women under surveillance	Cancer detection rate per 1,000 surveillance screens or woman/years	Cancer detection rates by modality of screening
Chart et al. (1997)	18.0 per 1,000 w/s 5.7 per 1,000 w/s (incident)	N/R	CBE 1.9 per 1,000 w/s XRM 3.8 per 1,000 w/s
Lalloo et al. (1998)	11.0 per 1,000 w/s	5.6 per 1,000 prevalent 4.8 per 1,000 incident	CBE 4.7 per 1,000 w/s XRM N/R
Kollias et al. (1998)	14.0 per 1,000 w/s	8.0 per 1,000 prevalent 3.3 per 1,000 incident	CBE N/R XRM 12.0 per 1,000 w/s
Frederico et al. (1999)	39.0 per 1,000 w/s	N/R	CBE N/R XRM 39.0 per 1,000 w/s
Moller et al. (1999)	N/R	N/R	CBE N/R XRM N/R
Macmillan et al. (2000)	9.4 per 1,000 w/s	4.8 per 1,000 prevalent 4.5 per 1,000 incident	No raw data CBE 40% prevalent 62% Incident XRM 85% prev 100% incident
Kerlikowske et al. (2000)	6.0 per 1,000 w/s	6.0 per 1,000 overall	CBE 0 XRM 6.0 per 1,000 w/s
Nixon et al. (2000)	14.0 per 1,000 w/s	N/R	CBE N/R XRM N/R
Myles et al. (2001)	13.6 per 1,000 w/s	5.0 per 1,000 prevalent 4.9 per 1,000 incident	CBE N/R XRM N/R
Brekelmans et al. (2001)	21.7 per 1,000 w/s	7.2 per 1,000 person/years	CBE N/R XRM N/R
Gui et al. (2001)	17.6 per 1,000 w/s (includes interval tumours)	4.4 per 1,000 overall (includes interval tumours)	CBE 14.8 per 1,000 w/s XRM 8.3 per 1,000 w/s (includes interval tumours)
Hou et al. (2002)	22.0 per 1,000 w/s (includes US)	N/R	CBE 7.0 per 1,000 w/s XRM 12.0 per 1,000 w/s
Scheuer et al. (2002)	42.4 per 1,000 w/s	N/R	CBE N/R XRM 30.0 per 1,000 w/s
Trecate et al. (2003)	170 per 1,000 w/s (includes US and MRI)	N/R	CBE 130.0 per 1,000 w/s XRM 0
Kriege et al. (2004)	23.0 per 1,000 w/s (includes MRI)	8.6 per 1,000 woman/years	CBE 1.6 per 1,000 w/s XRM 6 per 1,000 w/s BIRADS >4 9 per 1,000 w/s BIRADS >3
Warner et al. (2004)	93.0 per 1,000 w/s (includes US and MRI)	N/R	CBE 8.0 per 1,000 w/s XRM 34.0 per 1,000 w/s
Murday et al. (2004)	31.0 per 1,000 w/s	N/R	CBE 15.6 per 1,000 w/s XRM 31.0 per 1,000 w/s
Banks et al. (2004)	N/R	N/R	CBE N/R XRM N/R
Halapy et al. (2004)	N/R	9.1 per 1,000 overall	Not reported overall, only stratified by age and risk
Halapy et al. (2005)	N/R	9.1 per 1,000 overall	CBE 3.0 per 1,000 w/s XRM 6.0 per 1,000 w/s
Kuhl et al. (2005b)	76.0 per 1,000 w/s	N/R	CBE 2.0 per 1,000 w/s XRM N/R
Elmore et al. (2005)	N/R	N/R	CBE N/R XRM N/R
Gui et al. (2006)	6.2 per 1,000 w/s	6.2 per 1,000 overall	CBE 5.3 per 1,000 w/s XRM 6.2 per 1,000 w/s
Maurice et al. (2006)	15.0 per 1,000 w/s	5.97 prevalent 4.84 incident	CBE N/R XRM N/R

w/s = women under surveillance N/R = not reported

The cancer detection rates reported by women under surveillance cannot be compared across studies due to the differing surveillance intervals and differing length of the studies.

Table 8. Sensitivity of surveillance for women at high risk of breast cancer by XRM and CBE

Study	Measure of Accuracy	CBE (95% CI)	XRM (95% CI)
Kerlikowske et al. (2000)	Documents accuracy data but all stratified by age groups – refer to text and evidence table		
Brekelmans et al. (2001)	Sensitivity	40% (24% to 58%)	60% (42% to 76%)
Gui et al. (2001) (includes interval tumours with CBE)	Sensitivity	84% (60% to 97%)	47% (24% to 71%)
Hou et al. (2002)	Sensitivity Specificity	31.8% (14% to 55%) 99.4% (98.7% to 99.8%)	50% (28% to 72%) 99.6% (98.9% to 99.9%)
Scheuer et al. (2002)	Sensitivity	50% (21% to 79%)	42% (15% to 72%)
Kriege et al. (2004)	 Sensitivity Specificity PPV NPV Sensitivity Specificity PPV NPV	 6.7% (1.4% to 18.3%) † 17.8% (8.0% to 32.0%) ‡ 99.9% (99.8% to 99.9%) 50% (11.8% to 88.2%) 98.9% (98.6% to 99.2%)	BIRADS ≥ 4 24.4% (12.9% to 39.5%) 99.6% (99.4% to 99.8%) 47.8% (26.8% to 69.4%) 99.1% (98.8% to 99.4%) BIRADS ≥ 3 40% (25.7% to 55.7%) 94.9% (94.3% to 95.6%) 8% (4.8% to 12.3%) 99.3% (99% to 99.5%)
Warner et al. (2004)	Sensitivity Specificity PPV NPV AUC	9.0% (1% to 29%) N/R N/R N/R 0.48	36.3% (17.1% to 59.3%) 99.8% (98.7% to 99.9%) 88.9% (51.7% to 99.7%0 96.9% (94.8% to 98.3%) 0.77
Murday et al. (2004)	Sensitivity	33% (7% to 70%)	67% (30% to 92%)
Halapy et al. (2005)	 Sensitivity Specificity Sensitivity Specificity	Moderate family history 40% (30% to 50%) 94.8% (94% to 95%) Strong family history 40.5% (25% to 56%) 94% (93% to 95%)	Moderate family history 89.4% (83% to 96%) 93.9% (93% to 94%) Strong family history 76.3% (63 %to 90%0 94.6% (94 to 95%)
Kuhl et al. (2005b)	Sensitivity Specificity PPV NPV	2.3% (0.1 to 12%) N/R N/R N/R	32.5% (19% to 48.5%) 96.8% (95.7% to 97.7%) 23.7% (14 to 37%) 97.9% (97% to 98.6%)
Gui et al. (2006)	Sensitivity	14% (0.3% to 58%)	85.7% (425 to 99.6%)

† cut off of 'suspicious' ‡ cut off of 'probably benign' N/R = not reported

Table 9. Comparisons of surveillance by XRM and CRB with women in established breast screening programmes or unscreened women

Study	Cancer detection rate (per 1,000 screens)	Comparison
Lalloo et al. (1998)	Prevalent 5.6 per 1,000 Incident 4.8 per 1,000	NHSBSP Prevalent 5.7 per 1,000 NHSBSP Incident 3.8 per 1,000
Kollias et al. (1998)	Prevalent 8 per 1,000 Incident 3 per 1,000	NHSBSP Prevalent 6.5 per 1,000 NHSBSP Incident 3.8 per 1,000
Frederico et al. (1999)	Overall 19.1 per 1,000	Italian BSP Overall 7.7 to 8 per 1,000
Macmillan et al. (2000)	Prevalent 4.8 per 1,000 Incident 4.5 per 1,000	NHSBSP Prevalent 6 per 1,000 NHSBSP Incident 4.6 per 1,000 Gothbrg trial Prevalent 1.5 per 1,000 Gothbrg trial Incident 1.7 per 1,000
Brekelmans et al. (2001)	Documented that results comparable to Dutch BSP but no data provided	
Gui et al. (2001)	Overall 4.4 per 1,000 (misleading as includes interval tumours)	NHSBSP overall 3.8 per 1,000
Gui et al. (2006)	6.2 per 1,000 screens	NHSBSP overall 3.8 per 1,000
Maurice et al. (2006)	5.9 per 1,000 prevalent 4.8 per 1,000 incident	NHSBSP 5.5 per 1,000 prevalent NHSBSP 4.6 per 1,000 incident

The third comparison is of the characteristics of tumours arising in women under surveillance compared with the characteristics of sporadic tumours arising in women who have not been under surveillance. Kollias et al. (1998) and Macmillan et al. (2000) demonstrated a statistically significantly higher proportion of DCIS detected in women under surveillance as opposed to women who were not under surveillance. However, there is a possibility that some of the diagnoses of DCIS may represent over-detection rather than early detection. Macmillan et al. (2000) also found a statistically significantly higher proportion of women had tumours with good prognostic indices (NPI) in the surveillance group compared with the group without surveillance. This was not significant in the study by Kollias et al. (1998). No significant difference in lymph node status could be demonstrated (Macmillan 2000). These results are summarised in **Table 10**. The assumption is that if screening can detect tumours at an earlier stage of development that early treatment will translate to a decrease in mortality. This has certainly been demonstrated in women over the age of 50 years in established screening programmes (Tabar et al. 2000). However, this may not hold true in women at high risk whose tumours may progress more rapidly and respond differently to treatment. Further study will be needed to demonstrate that early detection and treatment can decrease mortality.

Table 10. Comparisons of prognostic characteristics between tumours detected in women at high risk of breast cancer under surveillance and women not under surveillance

Study	Prognostic Characteristics	Screened women	Unscreened women	P values
Kollias et al. (1998)	DCIS	21% (6/29)	4% (2/54)	0.01
	NPI (excellent/good)	41% (12/29)	30% (16/54)	0.28
Macmillan et al. (2000)	DCIS	22% (17/75)	6% (26/440)	<0.001
	NPI (excellent/good)	57% 43/75	5% (24/440)	<0.001
	Node negative	61% (36/75)	60% (247/440)	0.19

In conclusion, surveillance with XRM in women at high risk of breast cancer appears to be more effective and accurate in detecting early breast cancer than surveillance with CBE alone. The cancer detection rates appear to be equivalent or higher than those in established breast screening programmes for women over 50 years of age. There is evidence from two studies to suggest that surveillance detects tumours at an earlier stage, but no evidence that early treatment results in a decrease in mortality. Only one study demonstrated a significant difference in mortality between a population under surveillance and a population not receiving surveillance (Maurice et al. 2006). Surveillance for breast cancer may result in over-detection rather than early detection. A good marker of improved early detection rather

than over-detection is a reduction in the number of interval tumours arising. The interval tumours were documented in these studies. Unfortunately, in studies examining more than one modality of surveillance the interval tumour rate is a result of the overall surveillance strategy and comparisons cannot be made. A high level of interval tumours remained in these studies of XRM and CBE surveillance. This suggests that to improve detection of early breast cancer in this population of high-risk women and especially those at highest risk, i.e. mutation carriers, more intensive surveillance strategies or the addition of other modalities of surveillance is required. The following chapters will review the evidence for the accuracy and efficacy of surveillance with additional modalities in women at high risk of breast cancer.

Table 11. Primary research studies appraised investigating the accuracy and efficacy of mammographic surveillance compared to usual care on outcomes from breast cancer

Source	Study design Evidence grading	Comparison interventions and dates of testing	Sample	Outcomes and Verification	Results	Comments
Chart et al. (1997) University of Toronto, Canada	Prospective cohort study III-2 (C1 P2 Q2)	Surveillance protocol dependent on level of risk assessed: High risk – receive CBE every 6 months and annual mammography. Surveillance starting at 40 years, or 10 years before the earliest age at which cancer was detected in the family. Moderate risk - Receive BSE monthly, CBE and mammography annually. Surveillance starting at 40 years or 10 years before the earliest age at which cancer was detected in the family. Slightly increased risk - receive BSE monthly, CBE and annual mammography after the age of 40 years.	Sample *n*=1044 women Recruited from 2 Breast clinics Recruited from 2 cancer clinics. All patients referred specifically for risk evaluation. Mean age at surveillance = 42.7 (SD =10.9) years in one clinic and 39.5 (SD=10.8) in the second clinic. Mean age at diagnosis = 47 years (range 32-82 years). Risk category High risk = 381 (36%) Moderate risk= 204 (20%) Slightly increased risk = 401 (38%) No increased risk = 58 (6%)	Outcomes of relevance: • Cancer detection rates at the initial visit and during surveillance. • Mode of detection • Tumour size • Tumour stage • Node status Not explicit about verification of diagnosis – biopsy/pathology must have been performed in those diagnosed with cancer to obtain tumour type, stage and nodal status. Follow-up in surveillance programme was ongoing verification that cancer not present in those not diagnosed – but is not a good reference as follow-up would need to continue beyond the end of study to fully assess patient status at last surveillance screen.	Cancer detection: 23 tumours were said to be detected altogether. 13 were on the prevalent round, 10 on incident rounds. However, this included 4 interval tumours, so 19 were actually surveillance detected. (Also 1 in a woman who had not yet commenced surveillance. The latter is not included in these figures as by definition it was not screen detected). When stratified by risk the figures include the 4 interval tumours; High risk = 10 Moderate risk = 2 Slightly increased risk = 7 Overall 18 per 1,000 women under surveillance, prevalent 12.4 per 1,000 women under surveillance and incident 5.7 per 1,000 women under surveillance. CBE 1.9 per 1,000 women under surveillance XRM 3.8 per 1,000 women under surveillance	Limitations include: Small number of tumours detected. Potential for selection bias as it Is not explicit about how participants were selected and how many chose not to participate. However, there are characteristics presented (age and risk category) of those who were discharged or lost to follow-up. There was no comment on past history of breast cancer, although it appears that this information was collected, or on the use of preventative strategies such as Tamoxifen or bilateral salpingo oophorectomy. Those interpreting the surveillance images were not blinded to the risk status of the women. Risk assessment may differ from other studies and reduce external validity – the assessment involved factors such as previous benign or *in situ* breast disease, reproductive factors, exposure to radiation and alcohol intake. Not just family history factors. Possible verification bias – only those with positive surveillance results had pathological verification; those with negative results were followed up for disease development. As it is an ongoing surveillance programme, some were not followed up for long after their last negative surveillance screen.

Table 11. Primary research studies appraised investigating the accuracy and efficacy of mammographic surveillance compared to usual care on outcomes from breast cancer (continued)

Source	Study design Evidence grading	Comparison interventions and dates of testing	Sample	Outcomes and Verification	Results	Comments
Chart et al. (1997) University of Toronto, Canada *Continued*		Details of risk categorisation are outlined in full in the original article – include family history plus other factors e.g. exposure to radiation, lifestyle factors and reproductive factors. Comparison made between outcomes in different risk categories, and between modalities used in surveillance. Study carried out between October 1990 and December 1996. No mention of classification system for images (e.g. BIRADS)	98 women were discharged to community follow-up due to low-risk status. 131 (12%) were lost to follow-up – these women were from all age groups. More than half were at slightly increased risk, a quarter at high risk, 10% from the moderate and 10% from the no-increased-risk groups. Risk stratification High risk = 1 or more factors estimated to increase their RR more than fourfold. Moderate risk = 1 or more factors estimated to increase their RR 2-4 fold. Slightly increased risk= 1 or more factors estimated to increase their RR less than twofold.	Average follow-up was 21.9 months (SD 21 months). (150 patients in 1 clinic had an average follow-up of 5.1 months (SD 5.8 months). Interval cancers were not commented on as monthly BSE was included as part of the surveillance and therefore cancers detected by this means were considered detected by the surveillance. However, it was stated that in cases of cancer detected 10 weeks to 6 months after assessment, images were reviewed to see if any abnormality was missed.	Mode of detection (only reported for incident tumours) High risk CBE detected 2 tumours Mammography detected 1 tumour Moderate risk None detected by surveillance Slightly increased risk Mammography detected all 3 tumours in this group Tumour size, stage and node status) Of the incident tumours, there were 3 invasive tumours. All were from 10 to 15mm in size. The other 3 were *in situ*. They were all node negative i.e. stage 0 (*in situ*) or stage I. Of the prevalent tumours, 6 were *in situ*, 3 were stage I and 4 were stage II or greater. Interval tumours There were 4 interval tumours detected by BSE. They were all invasive, sized 10 to 15mm and node negative.	Authors' conclusions: Surveillance for women at increased risk for breast cancer may be useful in detecting the disease at an early stage. However, to achieve these results, the regular performance of all 3 methods of detection (mammography, CBE and BSE) is important. The optimal age to commence surveillance remains unknown. Reviewers' conclusions: The surveillance regime evaluated appears to offer earlier diagnosis of breast cancer with smaller, earlier stage tumours. However, the power of the study to demonstrate statistical significance is limited by the small number of tumours. As many tumours were detected by mammography as there were interval tumours. This suggests that the regimen of mammography and CBE was not sufficient by itself in detecting disease.

Table 11. Primary research studies appraised investigating the accuracy and efficacy of mammographic surveillance compared to usual care on outcomes from breast cancer (continued)

Source	Study design Evidence grading	Comparison interventions and dates of testing	Sample	Outcomes and Verification	Results	Comments
Lalloo et al. (1998) Centre for Cancer Epidemiology and the Manchester Breast screening service, Manchester, U.K.	Prospective cohort study III-2 (C1 P1 Q2)	Intervention: Annual mammographic surveillance and CBE provided for all women. Surveillance commencing at 35 years or from 5 years younger than the earliest diagnosis of breast cancer in the family, until 50 years. Prevalent round was two-view mammography (oblique and craniocaudal) whereas the incident rounds were single -view (oblique) Comparison: Extrapolated the number of cancers expected to develop in this population over this time period had they not received surveillance, using data from Claus model. Also compared with the number of cancers expected in the general population from the regional cancer registry data, and compared with detection in the NHSBSP.	Sample = 1,259 women recruited at one family history clinic out of 2,446 women attending the clinic for the first time. Recruited from a breast screening service where they had been referred by GP or surgeon for risk assessment. Mean age at entry to surveillance was 39.1 (range 28-49) years Inclusion criteria: • <50 years • Lifetime risk breast cancer of 1 in 6 or greater (equivalent to fourfold increase in risk for these women) Risk assessed by Claus model (family history factors)	Outcomes of relevance: • Cancer detection rates (incident and prevalent) • Mode of detection • Tumour size • Tumour stage • Node status • Interval cancers Verification of diagnosis of those with a positive result was made by pathology reports. Verification of those with negative surveillance tests consisted of follow-up. The average follow-up was 30 months (range 1-54 months, median 33 months). Interval cancers were verified by reviewing the mammograms following diagnosis.	Cancer detection: 14 cancers were detected at screening – 7 were prevalent and 7 were incident. (2 were in 50 yr olds and so were excluded from the comparative calculations) The overall cancer detection rate was 11 per 1,000 women under surveillance. This was 5.6 per 1,000 prevalent surveillance screens and 4.8 per 1,000 incident surveillance screens. Mode of detection: 8 of the cancers detected were not palpable and therefore detected by mammography alone. Tumour size, stage and node status: Tumour size ranged from 7-45mm (2 were not reported) 5 of the cancers detected were *in situ* (LCIS or DCIS) – 4 prevalent and 1 incident. Of the 9 invasive tumours, 4 were node positive and 2 had no available results, 3 were prevalent and 6 were incident.	Limitations include: No characteristics of the 4.8 and 1.1% of women that failed to attend. Few details of women, such as past history of breast cancer, Tamoxifen use or previous bilateral salpingo oophorectomy. Those interpreting the imaging results were aware of the women's high-risk status. Verification bias was likely as only those with positive results were verified by pathology. The others were only verified with follow-up. Authors' conclusions: The overall detection rates of invasive and non-invasive cancers were 5.6 per 1,000 for prevalent screens and 4.8 for incident screens. Similar detection rates were reported in the older, average risk population screened by the NHSBSP. The results of the study suggest that the lead time gained from detecting cancer by surveillance in this population is unlikely to be, on average, greater than a year and suggests that surveillance of high-risk young women requires an annual screening policy.

Table 11. Primary research studies appraised investigating the accuracy and efficacy of mammographic surveillance compared to usual care on outcomes from breast cancer (continued)

Source	Study design Evidence grading	Comparison interventions and dates of testing	Sample	Outcomes and Verification	Results	Comments
Lalloo et al. (1998) Centre for Cancer Epidemiology and the Manchester Breast screening service, Manchester, U.K. *Continued*		Cancer detection was compared between different modalities – CBE and mammography. Data collected from September 1992 until April 1997. No mention of system of classification for images or cut off for abnormal result.			Interval cancers: 2 interval cancers were detected These were both invasive tumours, one of which was node positive and one node negative. They were both <15mm in size. Cancer detection rates equivalent to the NHSBSP (5.7 per 1,000 prevalent and 3.8 per 1,000 incident screens) The number of invasive cancers expected to occur if this high-risk population did not receive surveillance was calculated to be 8.45. This gave a ratio of cancers observed to those expected of 1.42 (95% C.I. 0.73-2.48). This calculation is broken down into 3 age groups and presented in full in the paper. (The figure looked only at women under 50 years).	Reviewers' conclusions: The surveillance regimen evaluated seems to offer earlier diagnosis of cancer in high-risk women aged <50 years. It demonstrates equivalent cancer detection rates to the NHSBSP. There was not enough data to compare cancer detection rates by modality of surveillance. The study was limited by the small number of tumours. To fully establish the effectiveness of surveillance in young high-risk women, long-term studies are required and most likely multi-centre studies to recruit sufficient numbers. The use of mammography in addition to CBE does seem to be advantageous – detecting 8 tumours that were not palpable.

Table 11. Primary research studies appraised investigating the accuracy and efficacy of mammographic screening compared to usual care on outcomes from breast cancer (continued)

Source	Study design Evidence grading	Comparison interventions and dates of testing	Sample	Outcomes and Verification	Results	Comments
Kollias et al. (1998) Nottingham, UK.	Prospective cohort study III-2 (CX P2 Q2)	Intervention: Biennial mammography and annual CBE. Patients also instructed on BSE and offered open direct access to the clinic if they found a problem between visits. Surveillance commencing at age 10 years younger than the youngest affected relative Prevalent round was two-view mammography (craniocaudal and mediolateral) whereas the incident rounds were single view (oblique) Comparison: Breast cancer incidence was compared with an age matched population in the UK. Breast cancer detection rates at prevalent and incident screens were compared with those in women of 50 years and over attending the NHSBSP (3-yearly mammography). No documentation of system of classification of imaging or cut off for abnormal result.	Sample = 1371 women Recruited at a family history breast screening cancer clinic Unclear who referred by Mean age at start of surveillance was 41years (range 18-49) Inclusion criteria: • <50 years • Asymptomatic • Lifetime risk breast cancer of at least 1 in 9 (one first-degree relative with breast cancer with an onset <60 years or multiple affected relatives aged <60 years) or greater. No minimum age of entry (although mammography not generally performed until the age of 35 years unless the family history suggested multiple affected relatives with a very early age of onset – in these incidences an individualised plan was drawn up.)	Outcomes of relevance: • Cancer detection rates (incident and prevalent) • Mode of detection • Tumour size • Tumour stage • Node status • Interval cancers • Prognosis (using Nottingham prognostic index) Verification of positive results was by histology Verification of negative results was through follow-up. Mean follow-up 22 (range 0-96) months Interval cancers were verified by looking back at mammograms and this picked up one false negative mammogram.	Cancer detection: 19 cancers were diagnosed by surveillance – 11 at the prevalent screen and 8 at incident screens. Cancer detection rate at the prevalent round was 8 per 1,000 surveillance screens The cancer detection rate at the incident rounds was 3.3 per 1,000 surveillance screens. Overall cancer detection rate was 14 per 1,000 women under surveillance. Mode of detection: At the prevalent round 2 cancers were mammographically occult. All incident tumours were demonstrated on mammography. Tumour size, stage and node status. Of all the tumours detected, 23 were invasive and 6 were DCIS Of the cancers detected by the prevalent round, 8 were invasive and 3 were *in situ*. Of the incident rounds, 6 were invasive and 2 were *in situ*. 15 tumours were 0-20mm, and 8 were >20mm in size 15 cancers were lymph node negative and 8 were positive 2 were Grade I, 9 were Grade II and 12 were Grade III.	Limitations include: Few sample characteristics presented, such as past history of breast cancer, Tamoxifen use, BSO or OCP or HRT use. Does not mention any loss to follow-up. Those interpreting the imaging results were aware of the women's high-risk status. Likely verification bias. Authors' conclusions: This study has shown that the incidence of breast cancer in women aged <50 years offered this surveillance is 5 times greater than that of an age-matched population. The cancer detection rate at the prevalent round was similar to that in the NHSBP; the rates were similar for invasive tumours but more *in situ* cancer was detected in the study surveillance group. A significant difference was seen in the proportion of cases of DCIS between the surveillance group and the symptomatic patients (21% versus 4 %, P=0.01). Histological prognostic factors can serve as surrogate predictors of survival in women with breast cancer. The long-term survival of patients with DCIS or categorised in the good prognostic group is only marginally less than that of an age-matched population and represent a population that is potentially 'cured' after local treatment alone. In this way, a survival advantage over a group not receiving surveillance is anticipated. .

Table 11. Primary research studies appraised investigating the accuracy and efficacy of mammographic surveillance compared to usual care on outcomes from breast cancer (continued)

Source	Study design Evidence grading	Comparison interventions and dates of testing	Sample	Outcomes and Verification	Results	Comments
Kollias et al. (1998) Nottingham, U.K. *Continued*		The prognostic features of the surveillance-detected cancers were compared with cancers in women <50 years who had presented with symptomatic primary operable breast cancers, and who had a family history that would have made them eligible for surveillance had they been referred to the clinic. Data collected from January 1988 to December 1995	Risk assessed by Claus model (family history factors) but also incorporated atypical hyperplasia as a risk factor. Median breast cancer lifetime risk for cohort was 16.5% (range 11-45%) and the median calculated RR was 2.3 (range 1.5-6)		Prognosis: 12 tumours were categorised as good prognosis and 17 were moderate/poor prognosis. Interval cancers: 10 cancers were interval cancers, discovered incidentally or by BSE. 6 developed within 12 months of the last normal mammogram – 2 were mammographically occult (even at diagnosis) and 1 was a false negative (evident on previous film). 3 developed within 12 -24 months of the last normal mammogram – 2 were occult. 1 developed in a woman who had not yet commenced surveillance mammography (aged 31years). The overall interval cancer rate in the 24 months post-mammography was 2.5 per 1,000 women under surveillance. **Comparison** Incidence of breast cancer is 1.6 per 1,000 women-years in the age matched general population in England and Wales. Therefore the RR of breast cancer in the surveillance group was 5 times higher.	The rate of interval cancer was higher than in the NHSBSP and perhaps suggest that annual surveillance may be more effective than biennial. A RCT is suggested to address this. The results suggest that young women at risk of breast cancer due to a family history may benefit from regular breast surveillance due to the early detection of *in situ* lesions. Reviewers' conclusions: The surveillance regimen evaluated seems to offer diagnosis of cancer in high risk women aged <50 years at a more favourable stage compared with women who did not receive surveillance and presented symptomatically. However, comparing prognostic characteristics between the surveillance detected tumours and those in a symptomatic population does not necessarily mean that the surveillance is detecting the same tumours earlier but may be detecting a different set of tumours (more indolent, lower grade ones). Additionally, detecting them earlier does not directly translate to a reduction in mortality and has not yet been proven to achieve this in this high-risk population

Table 11. Primary research studies appraised investigating the accuracy and efficacy of mammographic surveillance compared to usual care on outcomes from breast cancer (continued)

Source	Study design Evidence grading	Comparison interventions and dates of testing	Sample	Outcomes and Verification	Results	Comments
Kollias et al. (1998) Nottingham, U.K. *Continued*					Comparison of breast cancer detection rates: Cancer detection rate in the NHSBSP prevalent screen is 6.5 per 1,000 screening visits. The cancer detection rate at the NHSBSP incident screen is 3.8 per 1,000 screening visits There was not enough data to compare cancer detection by modality. The cancer detection rate for XRM alone was 12 per 1,000 women under surveillance, but cannot calculate for CBE. Comparison of prognostic features: A significant difference was seen in the proportion of cases of DCIS between the surveillance group and the symptomatic patients (21% (6/29) versus 4 % (2/54). P=0.01). No differences were demonstrated for tumour size, histological grade or lymph node status. A higher proportion of the surveillance group was categorised as being in the good prognostic group (41% versus 30%). However, the numbers were too small to reach statistical significance (P=0.37)	Regarding cancer detection rate, the author states that the rates were similar to the NHSBSP for invasive tumours but more *in situ* cancer was detected in the study surveillance group. However, the statistical significance of this finding is not commented on. This study is limited by the small number of tumours detected. A larger study of this nature or a study comparing this surveillance regime to one with annual mammography would be valuable.

Table 11. Primary research studies appraised investigating the accuracy and efficacy of mammographic surveillance compared to usual care on outcomes from breast cancer (continued)

Source	Study design Evidence grading	Comparison interventions and dates of testing	Sample	Outcomes and Verification	Results	Comments
Federico et al. (1999) University of Modena, Italy	Prospective cohort study III-2 (CX P1 Q2)	Surveillance protocol involved : CBE every 6 months, by the same physician where possible. Mammographic surveillance from 30 years age and interval depends on risk category: In HBC and SHBC families it is every 2 years from 30-36 years, then annually from 37-65 years. In FBC families it is performed at 2 yearly intervals under 50 years and then annually. BSE was not recommended as a routine procedure, but women were asked to inspect and palpate their breasts and seek advice if they noticed or felt 'something different'. Surveillance for ovarian cancer was also undertaken. Comparison of prevalence rate is made with the first round of BC screening programmes carried out in nearby cities in Northern and Central Italy.	Sample *n*=151 women Identified from family pedigrees of breast cancer compiled by the University of Modena over 5 years. The mean age of the 151 women is not given. The mean age of tumour development in these women is 58.5 years. The mean age of the cohort was not reported. 196 high-risk women were interviewed (out of 592 eligible for inclusion) and asked permission for enrolment. The 45 who did not participate were either due to a lack of compliance or previous enrolment by another institution. The criteria for risk stratification are described in full in the original article and include family history factors alone. The criteria were used to stratify the families into 4 clusters: HBC= hereditary breast cancer	Outcomes of relevance: • cancer detection rate • mode of detection • tumour size • tumour stage • node status • interval tumours • mortality Verification of tumours was done pathologically and they were classified by WHO criteria. Verification of negative results consisted of follow-up. Mean follow-up was 24 months.	Cancer detection: 6 tumours were detected in total. 3 were in the HBC and HBOC risk group, 2 were in the SHBC risk group and 1 was in the FBC risk group. 3 were detected on the prevalent round and 3 were detected on incident rounds. Cancer detection overall was 40 per 1,000 women under surveillance (20 per 1,000 on the prevalent and incident rounds) The cancer detection rate at the prevalent screen was given as 19.1 per 1,000 mammographies. Mode of detection: 6 tumours were detected mammographically. It is not clear how many of these were palpable. One tumour was detected by CBE, and is described as the 'interval tumour' but it is not clear if this is classified in this way because it was mammographically occult or if it was detected between surveillance rounds. Tumour size, stage and node status:	Limitations include: Selection bias – it is not made clear how the women were selected who were asked to participate. It is not documented how their characteristics compared to those who were not asked, or how the characteristics of those who gave permission differed from those who declined to participate. Few sample characteristics presented, such as past history of breast cancer, Tamoxifen use, BSO or OCP or HRT use. No mention of blinding to the women's risk status. Verification bias – mean follow-up was 24 months. Risk assessment strategy is unique to this study. Mortality is not adequately measured due to the short follow-up post-diagnosis in most of the women. The definition of interval cancer is unclear as it says that it was 'clinically detected' but not when, and CBE was part of the surveillance protocol. Authors' conclusions: Generally, although preliminary, our results suggest that the selection procedure developed in this study identifies true high-risk groups which can represent the target for future strategies of breast cancer early detection or prevention.

Table 11. Primary research studies appraised investigating the accuracy and efficacy of mammographic surveillance compared to usual care on outcomes from breast cancer (continued)

Source	Study design Evidence grading	Comparison interventions and dates of testing	Sample	Outcomes and Verification	Results	Comments
Federico et al. (1999) University of Modena, Italy *Continued*		A control group of those at no increased risk of breast cancer is commented on but no comparisons appear to have been drawn. Comparison is also made of detection rate between CBE and mammography. It is unclear exactly the timing of data collection but it must have occurred some time between 1996 and 1999. No mention of the system of classification for images or the cut-off used for an abnormal examination.	HBOC = hereditary breast/ovarian cancer SHBC = suspected hereditary breast cancer FBC = familial breast cancer (There was SFBC and also those who had no criteria. The latter were considered sporadic and included in a control group) Inclusion criteria for the surveillance program were: • positive family history of breast cancer in either the HBC, HBOC, SHBC or FBC groups. • age range 30-65years • residence in the province of Modena		The tumours ranged in size from 0.9-35mm. One tumour was *in situ* (prevalent) and the others were all invasive (2 prevalent, 3 incident). Three were infiltrating ductal, 1 was infiltrating lobular and 1 was infiltrating tubular carcinoma. 2 of the 6 tumours were node positive. Interval tumours: One interval tumour was detected (as mentioned above, by CBE). It was invasive multicentric and node positive. (the verification process/definition of interval tumours was not explicit). Mortality: One of the women with breast cancer died, 59 months after diagnosis. Of those still alive, most were not long post-diagnosis (range 2-62 months). Comparison: The cancer detection rate at the prevalent screen was given as 19.1 per 1,000 mammographies, which is stated to be higher than that in the breast cancer screening programmes currently operating in nearby cities for women at average risk (7.7-8 per 1,000 mammographies).	Surveillance results in a higher cancer detection rate than in the Italian current breast cancer screening programmes, and mammography appears to detect more tumours than CBE alone. Appropriate follow-up and extensive genetic testing will be necessary to check the validity of the operational criteria and the rationale for stratifying individuals with suspected inherited cancer predisposition into four risk categories. Reviewers' conclusions: The main focus of this study was the development of a system of risk stratification for families at a high risk of breast and ovarian cancer. Therefore, there is less detail of the surveillance programme and outcomes. However, it does suggest that the surveillance strategy was efficacious. The cancer detection rate was higher than in screening programmes already in operation for women >50 years. However, no conclusions could be drawn about the characteristics of the tumours detected as there was no comparative data. It was also not possible to compare cancer detection by modality of surveillance. The mortality statistics had insufficient follow-up to be meaningful.

Table 11. Primary research studies appraised investigating the accuracy and efficacy of mammographic surveillance compared to usual care on outcomes from breast cancer (continued)

Source	Study design Evidence grading	Comparison interventions and dates of testing	Sample	Outcomes and Verification	Results	Comments
Moller et al. (1999) Multi-centre study from 7 centres in the EU Demonstration Programme on Clinical Services for Familial Breast Cancer (Norway, Dundee, Manchester, Leiden, Aberdeen, Edinburgh, London) (includes data from Lalloo et al. 1998)	Collation of data from 7 prospective cohort studies III-2 (C1 P1 Q3)	The surveillance protocols are reported to be described in separate papers. They all included mammography, usually annually, from age 35 to 50 years and starting at a younger age if there had been very early onset disease in the family. For women over 50 years, surveillance intervals in some centres have been longer (18 months to 2 years). This was combined with regular expert CBE and instruction on BSE ('breast awareness'). Comparisons: Comparisons were made between the characteristics of surveillance-detected tumours and interval and other tumours (those detected prior to receiving screening).	Sample *n*= The paper does not report how many women were under surveillance in total. 161 tumours were detected in 152 women. Mean age at diagnosis was 48.6years (range, 28-71years) 57% were diagnosed <50years and 19% were diagnosed <40years of age. Recruited from 7 centres in the EU (the breakdown of numbers per study centre is in the original article) Inclusion criteria: • asymptomatic - no signs of breast cancer past or present. Risk stratification: In all centres the women are eligible for surveillance if their genetic risk is at least twice that of the general population, based on the model by Claus et al. (2001)(family history factors alone).	Relevant outcomes: • cancer detection • mode of detection • tumour stage • node status • interval tumours • 5-year overall survival and event-free survival (free from cancer related death and tumour spread) Verification of positive results is by pathology post-excision. Verification of negative results is by follow-up. The mean follow-up time is not reported. Interval tumours had a confirmed previous negative surveillance examination Another category of 'others' was created and included those tumours which presented clinically before a planned first surveillance examination was actually undertaken.	Cancer detection: 121 tumours were detected (some of these were 2 tumours in 1 women) 40 were detected at the prevalent round, and 81 at subsequent rounds. Mode of detection: Of the prevalent round, 8 (16%) were not detected by mammography (presumably by CBE). Of the incident rounds, 20 (18%) were not detected by mammography (presumably by CBE). Tumour stage and node status: In the prevalent round, 8 tumours were *in situ* and 32 invasive. Nine were node positive. In the incident rounds, 22 were *in situ* and 59 were invasive. Thirteen were node positive. Interval tumours: There were 29 interval tumours during the surveillance. Of the interval tumours, 2 were *in situ*, 27 were invasive and 11 had positive nodes. There were also 11 tumours that arose in the period between women being referred for surveillance and them actually having there first examination.	Limitations include: Selection bias – as this study is a conglomeration of data from 7 centres, there is no information on the characteristics of the women or of the selection processes, other than that they were risk stratified by the Claus model. No mention of Tamoxifen use, BSO or OCP or HRT use. No mention of blinding to the women's risk status. Verification bias – the mean follow-up time is not presented. Survival time may be subject to lead time bias and length bias, and does not necessarily reflect a benefit if compared to women with cancer who were not under surveillance. Authors' conclusions: This study demonstrates conclusively that surveillance programmes for women whose family histories suggest they may be at increased risk can detect the majority of breast tumours, including those arising at an early age. More than 75% of tumours were detected in the course of planned surveillance examinations. It is suggested that reducing the surveillance interval further, perhaps to 6 months, would reduce the number of pathologically advanced interval tumours. Trying to ensure that people received screening when it was due would also help this.

Table 11. Primary research studies appraised investigating the accuracy and efficacy of mammographic surveillance compared to usual care on outcomes from breast cancer (continued)

Source	Study design Evidence grading	Comparison interventions and dates of testing	Sample	Outcomes and Verification	Results	Comments
Moller et al. (1999) *Continued*		Comparisons were also made with the age-specific rates of cancer in the general population, between detection with CBE and with XRM and between survival of women under surveillance compared to sporadic cases in a population without surveillance. Comparisons are also drawn with the Norwegian Cancer Registry data on sporadic tumours in women without surveillance. No dates of testing are given as they vary between centres. No comment on the method of classification of images or the cut-off for an abnormal result.		It is proposed that evaluation of surveillance programmes for those at high risk must be based on 'intention to screen' as delays in implementation of that intention (up to 1 year in some centres), as well as prolongation of surveillance intervals, may have adverse effects on programme performance.	Of these, none were *in situ*, 6 were invasive node negative and 5 were invasive node positive. Survival: 5-year overall survival was 0.89 (SE 0.05) and 5-year event-free survival for the whole group of women with tumours was 0.86 (SE 0.06). This broke down into 1 for women with CIS, 0.88 (SE 0.06) for women with node-negative invasive tumours and 0.67 (SE 0.20) for women with node-positive invasive tumours. Comparisons The interval and 'other' tumours were more often node positive (p=0.006) and less often *in situ* (p=0.01). The age-specific rates of the tumours were 8 times higher than the age-specific rates in the general population in Norway. In Manchester the age-specific rates of the tumours were 2.5 times higher than the age-specific rates in the general population.	A reduced surveillance interval may be particularly appropriate in BRCA1 mutation carriers. Stage specific 5-year survival was found to be similar to that reported for sporadic breast cancer. However, the overall 5-year survival was better. This indicates that prognosis is related to stage at diagnosis and that the effect of the surveillance intervention was mediated through diagnosis at an early stage. Reviewers' conclusions: Overall, the reporting is poor as there is no data on the total number of women in this cohort. It is also difficult to assess bias with the lack of individual data on the studies involved. The survival comparison with woman with sporadic cancer may not necessarily prove an advantage of surveillance, due to lead time bias and length bias. There is also little detail of the sporadic group to decide if this is a reasonable comparison. A study with longer follow-up and comparing mortality of those under surveillance and those not receiving surveillance would be required to prove a benefit of surveillance.

Table 11. Primary research studies appraised investigating the accuracy and efficacy of mammographic surveillance compared to usual care on outcomes from breast cancer (continued)

Source	Study design Evidence grading	Comparison interventions and dates of testing	Sample	Outcomes and Verification	Results	Comments
Moller et al. (1999) *Continued*					This was felt to be due to different risk profiles in the different countries i.e. higher number of mutation carriers in Norway. The comparison of surveillance modalities suggested that CBE was necessary in addition to mammography. Unfortunately, the numbers that were not detected by CBE but were by mammography was not given. Therefore, the use of mammography in addition to CBE cannot be commented on. Stage-specific survival was similar to that reported for sporadic breast cancer, while the overall 5- year survival was better.	The study does not provide enough information to calculate the cancer detection rates. The fact that a large number of tumours were not detected by XRM suggests that CBE is important. The number of interval tumours suggests that more intensive surveillance or surveillance with other modalities is required in this population. The 'intention to screen' analysis is a reasonable proposition. It is important for an effective surveillance programme to pay close attention to the logistics of surveillance and to referring women directly, without any delay. There is also insufficient data to compare mammographic surveillance with CBE alone.

Table 11. Primary research studies appraised investigating the accuracy and efficacy of mammographic surveillance compared to usual care on outcomes from breast cancer (continued)

Source	Study design Evidence grading	Comparison interventions and dates of testing	Sample	Outcomes and Verification	Results	Comments
Macmillan, R D. (2000) Nottingham, UK Results from 22 Breast Units in the UK.	Cohort study –data was collected prospectively within 9 units (4,906 women) and retrospectively within 13 units (3,877 women). III-2 (C1 P1 Q2)	Surveillance protocols varied between different units. Mammography was offered annually by 11 units, biennial by 10 units and every 18 months by one unit. CBE was offered annually by 20 units. One unit performed US annually. Surveillance was initiated 5 years before the age at diagnosis of the youngest affected relative in 12 units, and 10 years before this age in 5 units. One unit started at the age of 30 years, one at 35 years, one at 40 years, and 2 units had no fixed criteria for starting. Comparison is made of the cancer detection rate between the study cohort, the NHSBSP f or women aged 50-64, and the Gothenburg trial of screening between ages 40-49(non-high risk women).	Sample *n*= 8,783 women (range 49 to 1731 per breast unit) Median age at diagnosis was 41 years (range 30-49) for the prevalent round, 44 years (ranges 28-49 and 31-49) for the incident round and the interval tumours respectively. Inclusion criteria • asymptomatic women • <50 years age • family history of at least one first-degree relative affected with breast cancer before the age of 50 years. Data for women with a paternal family history giving the same relative risk as the original criteria were also included. Risk Stratification: This was performed as detailed in the inclusion criteria.	22 units were able to give data for the prevalent round, 13 units were able to give data for incident rounds and 12 units provided data on interval cancers (8,166 screening visits). Relevant outcomes • cancer detection rate • mode of detection • tumour size • tumour stage • node status • interval tumours Verification at the different surveillance units was not reported, however it appears to have been pathological for positive results and follow-up in those with negative results. There was a total of 9,075 women/years of follow-up. The verification of interval tumours is also not discussed.	Cancer detection: 83 cancers were detected per 9,075 women years of follow-up. (only 13 units were able to provide data for incident rounds and 12 for interval cancers) Overall 9.1 per 1,000 women under surveillance. 42 were diagnosed at the prevalent round and 41 at the incident rounds. The cancer detection rate was 4.78 per 1,000 prevalent screens and 4.52 per 1,000 incident screens. Mode of detection: Data on the mode of detection was available for 67 of the 83 surveillance-detected cancers. Of the prevalent cancers, 47% were palpable and 85% were visible on mammography. Of the incident cancers, 62% were palpable and 100% were visible on mammography. No raw data for this. Tumour size, stage and node status: Complete data was available for only 75 of the 83 invasive cancers. Incomplete data was available for 5 and no data was available for 3 tumours.	Limitations include: Partly retrospective study. Considerable amount of data was missing or only available from certain centres (especially concerning interval cancers and the pathological features of the tumours) which may have biased the results. In most cases the denominator was taken from the total data available, not including the missing cases. Considerable variations in type and frequency of surveillance offered between units. No fixed age for the initiation of surveillance. Possibly different referral criteria in different centres and little characteristics given of the women selected in each centre, which may have led to selection bias and may go against combining all this data in such a way. No mention of past history of breast cancer, Tamoxifen use, BSO or OCP or HRT use. No mention of blinding to the women's risk status. Very little information about verification practices and verification bias is likely. Authors' conclusions: The authors recognise and discuss the limitations of their study, especially the retrospective design. The cancer detection rate from surveillance of these women under the age of 50 with a significant family history of breast cancer is comparable to that observed in population screening of women aged 50-64 in the NHSBSP.

Table 11. Primary research studies appraised investigating the accuracy and efficacy of mammographic surveillance compared to usual care on outcomes from breast cancer (continued)

Source	Study design Evidence grading	Comparison interventions and dates of testing	Sample	Outcomes and Verification	Results	Comments
Macmillan, R D. (2000) Nottingham, UK Results from 22 breast units in the UK. *Continued*		Comparison is also made of the tumour pathology and the Nottingham Prognostic Index (NPI) between the study cohort, the NHSBSP for women aged 50-64, and440 breast cancer cases presenting symptomatically before age 50 to the Nottingham City Hospital between 1993 and 1997. (This symptomatic cohort appears to overlap with the one utilised by Kollias et al. 1998) Finally, a comparison of incident cancer detection rate is made between the units in the study that use annual surveillance versus biennial surveillance. No mention of the method of classifying images or the cut-off for an abnormal result.	No other method or model was used.		The data was not stratified by mode of surveillance. The tumours detected at the prevalent round had a mean size of 19.9mm, 11 were *in situ* (28%) and 31 invasive (data missing for 2). Of the invasive tumours, 17 (59%) were node negative and 12 (41%) were node positive. The tumours detected at the incident rounds had a mean size of 13.9mm, 6 were *in situ* (17%) and 35 were invasive (data missing for 5). Of the invasive tumours, 19 (63%) were node negative and 11 (37%) were node positive. Interval tumours: There were 20 interval tumours. The interval tumours had a mean size of 19.4mm, 3 (16%) of them were *in situ* and 17 were invasive (data missing for 1). Of the invasive tumours, 9 (56%) were node negative and 7 (44%) were node positive. Comparisons: The incidence was given as 11.3/1,000/year (invasive = 9.1/1,000/year). This was compared to a population incidence of 2/1,000/year (invasive 1.8/1,000/yearr)	The pathological characteristics of the tumours identified appear better than those seen in women of a similar age without surveillance who present symptomatically, but not as good as those detected in the NHSBSP. These findings are encouraging but the limitations of the study must be borne in mind. The findings of the study provide evidence to support a programme of evaluation of surveillance for women with a significant history of breast cancer. The group who undertook this study proposed a fully prospective observational study with standardised protocols. Reviewers' conclusions: As discussed by the authors, the limitations of this study need to be taken into consideration i.e. the partly retrospective nature, the differences in protocols between centres and the incompleteness of the data. In addition, the comparison with cancers presenting symptomatically in a population without surveillance of a similar age is not directly comparable. The symptomatic women have not been assessed for risk, however this is more likely to underestimate the

Table 11. Primary research studies appraised investigating the accuracy and efficacy of mammographic surveillance compared to usual care on outcomes from breast cancer (continued)

Source	Study design Evidence grading	Comparison interventions and dates of testing	Sample	Outcomes and Verification	Results	Comments
Macmillan, R D. (2000) Nottingham, UK Results from 22 breast units in the UK. *Continued*					and this equates to an average relative risk for women in this study of 5.65 (invasive 5.1). The cancer detection rates for Prevalent cancers were 4.8/1,000 for the study group, 6/1,000 for the NHSBSP and 1.5/1,000 for the Gothenburg trial. For incident cancers this was 4.5/1,000, 4.6/1,000 and 1.7/1,000 respectively, and for interval cancers the rates were 2.5/1,000, unknown and 0.4/1,000 respectively. The pathological characteristics of the surveillance-detected tumours were better than those seen in women of a similar age who presented symptomatically but not as good as those seen in the NHSBSP. There was a significant difference in the proportion of DCIS. The NPI again showed the screen-detected tumours to have a more favourable prognosis than those seen in women of a similar age who presented symptomatically, but not as favourable as those seen in the NHSBSP. A significant difference was not able to be demonstrated between annual and biennial surveillance, although the cancer detection rate was higher in the former. (5.71/1,000 versus 3.64/1,000) (p=0.15)	differences in pathology between the symptomatic and surveillance detected tumours (as these symptomatic women are likely to be high risk as <50 years at diagnosis). However, the study does suggest that the cancer detection rate is as good as that of the NHSBSP, that tumours are detected at a slightly more favourable stage with a better prognostic index than those that are unscreened, and that there is advantage to utilising mammography in addition to CBE. The proposed prospective study would remove many of the aforementioned limitations and allow a better assessment of such surveillance.

Table 11. Primary research studies appraised investigating the accuracy and efficacy of mammographic surveillance compared to usual care on outcomes from breast cancer (continued)

Source	Study design Evidence grading	Comparison interventions and dates of testing	Sample	Outcomes and Verification	Results	Comments
Kerlikowske et al. (2000) Multi-centre in 7 sites in USA (California, New Hampshire, New Mexico, Vermont, Washington State (x2) and Colorado)	Cross-sectional study (CX P1 Q2)	Screening protocol Looked at the prevalent surveillance only. One mammographic exam per woman was included in the pooled analysis. If there were results from >1 exam then only the earliest one was included. 2 standard views were taken per breast. CBE not part of protocol as were excluded if had a palpable mass. Dates of surveillance from April 1985 to November 1997 (different units gave data for different periods during this overall time) Comparisons were made between the women with a family history of breast cancer and those of the same age group without such a history. There was also a comparison incorporated of the efficacy of XRM in addition to CBE, as all these women had no palpable mass and	Sample no = 389,533 women (of all risk groups), 50,834 (13%) were defined as having a family history. Mean age (at diagnosis or entry to the programme was not reported Recruited retrospectively from 7 screening registries in 6 US centres. Inclusion criteria: • aged 30 to 69 years; • referred for screening mammography between 1985 to 1997; • only the first/earliest mammographic image for each women was included. Exclusion criteria: • previous diagnosis of breast cancer; • palpable mass by history or physical examination;	Relevant outcomes: • cancer detection rate • tumour stage • sensitivity • specificity • PPV Verification of a positive (abnormal) result was done by linkage with physicians, a pathology database, or a radiology database (depending on the protocol in the area) to get pathology results of biopsies. (Probability matching software was used for this). Women with lobular carcinoma *in situ* only, were not considered to have breast cancer. 12 months between the examination and diagnosis or non-diagnosis was used as the cut-off for true and false negatives and true and false positives.	Cancer detection rate: 1,650 cases of breast cancer were identified in the total study population with 309 of these in women with a family history of breast cancer (6 per 1,000 women under surveillance) Cancer detection rate among women with a family history was 6.1 per 1,000 examinations as compared with 4 per 1,000 examinations in women without a family history. These were all meant to be first round/prevalent examinations although it was suggested from the data from 5 registries that a high proportion of women had self reported previous use of mammography (81.7% and 80.2% in women with and without a family history of breast cancer respectively). Tumour stage: 70 (23%) tumours were *in situ* and 239 (77%) were invasive in the women with a family history. This compares with 315 (23%) *in situ* tumours and 1,026 (77%) invasive tumours in women without a family history.	Limitations include: The accuracy of the data depends on the completeness of reporting to the SEER programme, tumour registries and pathology laboratories. Data were limited to a certain area, so if women moved outside this area and then developed a tumour, this would not be detected. Selection bias is hard to assess as it is not known why these women were referred for surveillance and what the characteristics of the sample are other than the number with a family history. No mention of Tamoxifen use, BSO or OCP or HRT use. Verification bias could be a factor, although only one image was considered and a period of12 months following surveillance was considered to determine the outcome. If annual surveillance was proposed for women at high risk, then this length of follow-up would reach to the next surveillance round and therefore be sufficient to verify a negative result. No blinding of radiologist to the risk status but might be less influenced in this study design as it included women of all risk groups and were not specifically referred for this purpose, so the radiologist may not delve much into their family history. Authors' conclusions: The rate of cancer detection was 1.3 to 2 times higher among women with a family history of breast cancer than in those of a similar age without such a history.

Table 11. Primary research studies appraised investigating the accuracy and efficacy of mammography screening compared to usual care on outcomes from breast cancer (continued)

Source	Study design Evidence grading	Comparison interventions and dates of testing	Sample	Outcomes and Verification	Results	Comments
Kerlikowske et al. (2000) Multi-centre in 7 sites in USA (California, New Hampshire, New Mexico, Vermont, Washington(x2) State and Colorado) *Continued*		therefore the tumours detected would not have been detected without the mammographic surveillance.	• if ZIP code outside the catchment area for Surveillance Epidemiology and End Results (SEER) programme or state tumour registry to minimise incomplete follow-up. Risk stratified retrospectively – considered to have a family history if they had at least one first-degree relative (mother, sister or daughter) with breast cancer.		Comparisons: The cancer detection rate is 1.5 (1.3-2) times higher in women with a family history of breast cancer than in those without such a history ($p<0.001$). The rate increases with age in both groups. This detection rate is all in women with normal CBE and therefore shows the importance of utilising mammography in addition to CBE in surveillance. Sensitivity: The overall sensitivity of surveillance (including women with and without a family history) was 80.9% (CI, 78.9--82.8%). This increased significantly with age (from 63.2% for ages 30-39 years to 83.3% for ages 60-69 years for women with a family history ($p=0.006$) and from 69.5% for ages 30-39 years to 87.7% for ages 60-69 years among women without a family history ($p=0.001$).The sensitivity did not differ significantly between women with and without a family history ($p=0.1$).	This is probably a reflection of the prevalence in this high-risk population. When stratified by age group, the cancer detection rate in women with a family history of breast cancer is similar to that among women a decade older without such a history. The results concern the ability of mammography to detect breast cancer in women with and without a family history of breast cancer. They do not provide information on the efficacy of mammography in reducing breast cancer mortality rates. The findings suggest that further information is required to determine whether mammography is sufficiently accurate and beneficial to support a recommendation for surveillance in this high-risk group. Reviewers' conclusions: This study must be considered in light of the above limitations. It is unclear whether the comparison group of women without a family history of breast cancer is representative of a comparison with women of average risk as it is unclear why these women (if under 50 years) would have been referred; there may have been selection bias. It is also unclear if the rounds included were prevalent or incident as women had often had previous imaging.

Table 11. Primary research studies appraised investigating the accuracy and efficacy of mammography screening compared to usual care on outcomes from breast cancer (continued)

Source	Study design Evidence grading	Comparison interventions and dates of testing	Sample	Outcomes and Verification	Results	Comments
Kerlikowske et al. (2000) Multi-centre in 7 sites in USA (California, New Hampshire, New Mexico, Vermont, Washington(x2) State and Colorado) *Continued*					In women <50 years, the sensitivity was significantly lower for invasive cancer than for all breast cancer (DCIS + invasive) (68.6% versus 74.9%, p=0.04). However, in women >50 years, the difference was not significant (83 versus 83.8, p>0.2). Specificity: The specificity of surveillance was lower among women with a family history than those without (87.7% versus 89.4%, p<0.001). The specificity was lower and homogenous across age groups in women with a family history and higher and not homogenous in women without such a history. PPV. The PPV was found to increase with age and also was higher in women with a family history of breast cancer than in those without (p=0.001).	The results indicate that cancer detection rate is higher in women with a family history of breast cancer, due to the higher prevalence in these high-risk women. Despite the supposed aggressiveness of tumours in these women there appears to be no difference in the percentage of *in situ* and invasive tumours between women with and without a family history. However, factors such as tumours size and node status have not been taken into consideration and could very well alter this finding. There is definitely benefit in utilizing mammography in addition to CBE in the detection of breast caner in high risk and low risk women.

Table 11. Primary research studies appraised investigating the accuracy and efficacy of mammography screening compared to usual care on outcomes from breast cancer (continued)

Source	Study design Evidence grading	Comparison interventions and dates of testing	Sample	Outcomes and Verification	Results	Comments
Kerlikowske et al. (2000) Multi-centre in 7 sites in USA (California, New Hampshire, New Mexico, Vermont, Washington(x2) State and Colorado) *Continued*					<table><tr><td rowspan="2">Age</td><td colspan="2">PPV</td></tr><tr><td>Family History</td><td>No Family History</td></tr><tr><td>30-39</td><td>1.9 (0.8-2.9)</td><td>1.2 (0.9-1.6)</td></tr><tr><td>40-49</td><td>2.5 (1.9-3.0)</td><td>1.8 (1.6-2.0)</td></tr><tr><td>50-59</td><td>4.1 (3.2-5.0)</td><td>3.3 (3.0-3.6)</td></tr><tr><td>60-69</td><td>6.7 (5.3-8.0)</td><td>5.6 (5.1-6.1)</td></tr></table> The PPV (for all age groups) is 3.7 for those with a family history versus 2.9 for those without a family history. This is a significant difference (p=0.001). Unfortunately the figures for the measures of accuracy cannot be checked as there is not sufficient data. The only data given for true positives and false negatives is not stratified by family history. In addition, it appears that some of the presented figures do not correlate with each other.	

Table 11. Primary research studies appraised investigating the accuracy and efficacy of mammography screening compared to usual care on outcomes from breast cancer (continued)

Source	Study design Evidence grading	Comparison interventions and dates of testing	Sample	Outcomes and Verification	Results	Comments
Nixon et al. (2000) Sweden	Analysis of subgroup of data from a randomised controlled trial – but the comparisons made in the study are those of a prospective cohort study (both receiving the same intervention). III-2 (CX P1 Q2)	Surveillance protocol was a randomised trial of mammographic screening versus no screening. In this paper comparisons are drawn between women in this study with a family history of breast cancer and those with no family history. The study was conducted in two counties but the data on family history was only collected in one county, in the women who attended screening (not in those that did not attend screening and not in the controls). The inter-screening interval was 24 months in women aged 40-49 years at randomisation and 33 months in women aged 50-74 years. Mammographic parenchymal pattern was classified by Wolkfe's method (data on this available for 22,438 women with a positive family history).	Sample no = 29,179. Women had family history data recorded when attending screening. This was 76% of the study group in this county. 3,226 women (11%) had a positive family history. No mean age is given. The results are stratified by age groups – 40-49 years and 50-74 years. Recruited as subgroup from data from Swedish two-county randomised trial of mammographic screening. Risk stratification has not been described, other than saying that these women had a family history of breast cancer. However, 'family history' is not defined.	Relevant outcomes: • cancer detection rate • mode of detection • tumour size • node status • interval tumours Verification of a positive screen was through pathology. Verification of a negative screen was through follow-up. Verification of interval tumours was not discussed.	Cancer detection: Overall, 45 tumours were detected by the screening in the women with a family history (1.4% of women over the 7-8 years). This compared to 228 in the 25,953 women without a family history of breast cancer (0.9% of these women over the 7-8 years) Mode of detection is presumably mammography – there is no discussion of the role CBE may have played in cancer detection. Tumour size and node status: (family history group) 32 (71%) tumours were <20mm in size and 10 (22%) were ≥20mm. 9 (20%) were node positive and 29 (64%) were node negative (the remainder are not commented on or perhaps were *in situ* – as these parameters refer to invasive tumours only). In those without a family history, 152 (66%) were <20mm and 48 (21%) were ≥20 mm, 35 (15%) were node positive and 149 (65%) were node negative.	Limitations include: Selection bias should be minimised by the RCT design in the overall study between those receiving and not receiving screening. However, in this subgroup study it is only those who received screening who are examined and are compared, according to having a family history or not. Those with family history taken were a select group who actually attended the screening, and this may have introduced bias. The sample in this subgroup study was also from only 1 county within Sweden and perhaps less representative of the wider population. There are few characteristics of this group given regarding mean age and other risk factors for breast cancer such as a personal history of disease, exogenous hormone use, or risk reducing strategies such as chemoprevention or prior preventative surgery. Unclear if the prevalent screen for all these women was within this screening programme or if some of them had been screened prior to this programme. Prevalent and incident results are not separated out.

Table 11. Primary research studies appraised investigating the accuracy and efficacy of mammographic surveillance compared to usual care on outcomes from breast cancer (continued)

Source	Study design Evidence grading	Comparison interventions and dates of testing	Sample	Outcomes and Verification	Results	Comments
Nixon et al. (2000) Sweden *Continued*		Comparison was made between women with a family history and those without a family history and between two age groups (40-49 years and 50-74years). Screening occurred over 7 to 8 years, but the actual dates are not given.			Interval tumours: Overall, 15 interval tumours arose in women with a family history of breast cancer and 70 in those with no family history. In those with a family history, 9 of the tumours were <20mm and 6 were >20mm. 11 were node positive and 4 were node negative. In those without a family history, 26 of the invasive tumours were <20mm and 32 were >20mm. 31 of the invasive tumours were node negative and 20 were node positive. Comparisons: Overall, there was a higher cancer detection rate and interval cancer rate in women with a family history of breast cancer compared with those without such a history. However, the percentage of screen-detected to interval cancers in both groups was similar. In the younger age group, women with a family history had a slightly higher percentage of interval tumours to screen-detected tumours. However, this difference was not statistically significant.	Verification bias is likely, and the average length of follow-up is not given. It would have been good to be able to compare the outcomes of those with a family history who were screened and unscreened. Unfortunately family history data was not accumulated for those who were not screened. Authors' conclusions: There appears to be no seriously decreased propensity for screen detection in women with a family history of breast cancer, although in women aged under 50 years there is a slightly higher proportion of interval cancers associated with family history. In this study the mean sojourn time is also estimated for those with and without a family history. There is quite a degree of uncertainty in these estimations. However, the results suggest that the screening interval used in this study is too long. It is proposed that a policy of annual surveillance for women with a family history of breast cancer is a reasonable one. Also, that this annual protocol should still be employed once women are over 50 years if they have a family history as the sojourn time does not increase with age in women with a family history of breast cancer as it does in women without such a history.

Table 11. Primary research studies appraised investigating the accuracy and efficacy of mammographic surveillance compared to usual care on outcomes from breast cancer (continued)

Source	Study design Evidence grading	Comparison interventions and dates of testing	Sample	Outcomes and Verification	Results	Comments
Nixon et al. (2000) Sweden *Continued*					The differences between those with and without a family history with respect to lymph node status, tumour size and grade were not statistically significant either. Unfortunately there was no comment on detection of cancer by CBE and therefore no conclusion can be drawn about the advantages of mammography versus CBE.	Reviewers' conclusions: This study suggests that mammographic surveillance of women at high risk is efficacious. It also verifies the idea that tumours in women at high risk of breast cancer develop more rapidly and therefore surveillance intervals are required to be shorter. Some of the data presented is difficult to interpret. There was no report on how many tumours were invasive and how many were *in situ*. The numbers in the tables on the characteristics of the tumours do not add up consistently and therefore there must also be some which were missing this data or did not have it collected. There is also some uncertainty about the characteristics of the sample group and the use, or not, of CBE. In addition, the process of risk stratification or what was meant by a "family history' is not explicit and therefore lessens the external validity. This may all be a result of this being a subgroup analysis of a larger trial and not being conducted specifically to examine women at high risk of breast cancer.

Table 11. Primary research studies appraised investigating the accuracy and efficacy of mammographic surveillance compared to usual care on outcomes from breast cancer (continued)

Source	Study design Evidence grading	Comparison interventions and dates of testing	Sample	Outcomes and Verification	Results	Comments
Myles et al. (2001) UK	Prospective cohort Study III-2 (CX P2 Q2)	Surveillance protocol involved annual mammography, although this was reduced in latter years to every 2 years due to pressures on resources. Also CBE was used, and some women had US – although it does not specify how many, how this was used (diagnostically or for surveillance) or the results of this. Comparison: The incidence in the study was compared with the incidence predicted by the Claus risk tables Dates of surveillance – 1987 to June 1999 (data on interval cancers is only complete up to July 1998)	Sample no = 2,998 women with a moderate family history. Mean age not given (age range 19-71) Recruited from a family history clinic in Manchester, UK. Risk stratified by Claus et al. (2001) model – the women included had sufficient family history to indicate a moderate increased risk of breast cancer but not sufficient to warrant gene mutation analysis. A cut-off of risk for inclusion or exclusion is not given. An incidence of between 0.21 and 0.23 was calculated from the Claus scores and then compared to the observed incidence,	Relevant outcomes • cancer detection rate; • mode of detection; • Interval cancer; Verification of a positive result appears to have been by pathology, although no pathology results are presented. Verification of a negative result is through follow-up. Verification of interval cancers is not discussed.	Cancer detection: During the programme, 41 cancers were detected. 15 were detected at the prevalent round, out of 2,998 surveillance screens (5 tumours per 1,000 surveillance screens). 26 were detected at incident rounds, out of 5,278 surveillance screens (4.9 tumours per 1,000 surveillance screens). Mode of detection: Presumably all detected by mammography; no mention is made of CBE. Interval tumours: 9 interval tumours occurred, 5 between the prevalent and subsequent round and 4 after the 2nd round onwards. The sensitivity which is documented is an estimate arrived at through modelling. It was not possible to calculate the sensitivity without more information on the study. The 70% of tumours that were assumed to be surveillance-detected were then applied to data from the Swedish Two County Trial which describes a 59% lower fatality rate from screen-detected tumours.	Limitations include: Few characteristics are presented about the cohort. No details of a personal history of disease, exogenous hormone use, or risk reducing strategies such as chemoprevention or prior preventative surgery. No comment on blinding of radiologists. They are a relatively low risk cohort, compared with other studies, not warranting genetic testing. Verification bias is likely The role of CBE and US in this study is not clear. The actual surveillance protocol was vague as some surveillance was done annually and some 2 yearly. This is likely to make a big difference to the outcomes (especially interval tumours), but is not accounted for. Authors conclusions Early indications are that the programme is likely to be effective. Further follow-up; analysis of tumour size, node status and malignancy grade, and subsequent mortality from breast cancer is required to confirm this.

Table 11. Primary research studies appraised investigating the accuracy and efficacy of mammographic surveillance compared to usual care on outcomes from breast cancer (continued)

Source	Study design Evidence grading	Comparison interventions and dates of testing	Sample	Outcomes and Verification	Results	Comments
Myles et al. (2001) UK *Continued*					This data was used to model the projected decrease in mortality from this surveillance. The estimate of the reduction of mortality due to the programme was made at 41%. This estimate contains considerable uncertainty. Comparison: The incidence predicted by the Claus tables was 3.73 and the cancer detection rate is just over 1 year's incidence.	Reviewers' conclusions: This study suggests that mammographic surveillance will detect tumours in women at high risk of breast cancer. However, there were still a considerable number of interval cancers occurring. It is not possible to comment on whether shortening the interval between surveillance may improve this as it is unclear how many rounds were done annually and how many biennially. There is also no suggestion if surveillance detection of tumours was advantageous in terms of being at an earlier stage and more treatable. This is because there was no information on intermediate outcomes such as the tumour characteristics. The article concluded by saying that an attempt was being made to collate this information and look at a more in-depth assessment of this programme. Unfortunately, the cancer detection information is not presented by modality and no comparison can be made between mammography and CBE detection rates.

Table 11. Primary research studies appraised investigating the accuracy and efficacy of mammographic surveillance compared to usual care on outcomes from breast cancer (continued)

Source	Study design Evidence grading	Comparison interventions and dates of testing	Sample	Outcomes and Verification	Results	Comments
Brekelmans et al. (2001) Erasmus University, Rotterdam, The Netherlands (Rotterdam Family Cancer Clinic)	Combined retrospective and prospective cohort study III-2 (C1 P1 Q2)	Surveillance protocol involved: CBE biannually (yearly in some moderate risk women). Mammography yearly and recommended monthly BSE – the minimum age being 25 years (although younger in women from families with a young age at onset). Post-1995, MRI is optional and used in women with dense breasts and in mutation carriers. Additional investigation with US and FNA is performed when clinically indicated. Comparison was made between the study sample and the expected number of cancers in a population aged 40 to 50 according to the National Cancer Registry. The cancer detection by the different modalities was also considered. .	Sample *n* = 1198 women divided in to 3 risk groups. Prevalent examination for 399 women Incident examination for 799 women, for 386 the previous imaging was in another hospital and for 413 the information on the previous imaging could not be found. Recruited from a large cancer clinic. Mean age overall at first surveillance was 38 years (range 21 to 70 years). Also given by risk category and is presented below. The criteria for risk stratification are described in full in the original article. They utilise BRCA1 and BRCA2 testing and referenced to tables from Claus et al. (2001). Therefore they include family history factors alone. All women with >15% risk of breast cancer.	Relevant outcomes: • cancer detection rate • mode of detection • tumour size • tumour stage • node status • interval tumours • mortality • Sensitivity Verification of tumours was done pathologically. Verification of negative results consisted of follow-up. Verification of interval tumours was not explicit. Median follow-up period of 3 years (range 0-22 years). Total follow-up = 3,607 person years. Moderate risk: 1,193 person/years High risk:2,146 person/years Mutation carriers 268 person/years	Cancer detection: 26 tumours were detected in total by surveillance, 3 were prevalent and 23 were incident (5 were mutation carriers, 17 were in the high-risk group and 4 in the moderate risk group), so the cancer detection rate for screening was 7.2 per 1,000 person years (invasive +DCIS). Cancer detection rates are given as: Moderate risk: 3.3/1,000 (1.1-8.6) person/years High risk: 8.4/1,000 (5.4-13.2) person/years Mutation carriers: 33/1,000 (17-63) person/years However, these figures are only for invasive tumours and also include the interval tumours. If recalculated with invasive tumours and DCIS but excluding interval tumours these rates are: Moderate risk: 3.3/1,000 person/years High risk: 7.9/1,000 person/years Mutation carriers 18.6/1,000 person/years	Limitations include: No blinding of radiologists to risk status mentioned. Verification bias is likely. Selection bias – there is little data on the characteristics of the sample and whether all women who were invited to screening took part. Also, there are no details of whether women had a past history of breast cancer, had undergone risk reduction strategies such as BSO or taking Tamoxifen, or had increased risk due to exogenous hormones (OCP/ HRT). The selective use of MRI and US may have confounded the results, however it was possible to separate tumours detected by different surveillance modalities. Authors' conclusions: Our results and those of others show that it is clearly possible to identify young women at high familial risk: the number of breast cancers detected in our population was on average 7 times greater than expected in an average risk population of comparable age. Our study sample was large enough and the follow-up period long enough to calculate age-specific parameters and results for three separate genetic risk groups: proven BRCA mutation carriers and women at a high or moderate risk of breast cancer.

Table 11. Primary research studies appraised investigating the accuracy and efficacy of mammographic surveillance compared to usual care on outcomes from breast cancer (continued)

Source	Study design Evidence grading	Comparison interventions and dates of testing	Sample	Outcomes and Verification	Results	Comments
Brekelmans et al. (2001) Erasmus University, Rotterdam, The Netherlands (Rotterdam Family Cancer Clinic) *Continued*		The study dates are a bit unclear; data was entered prospectively into a database from 1995 onwards, but retrospectively collected from before 1995. The study was concluded in January 2000	They are divided into Carriers, n=128(60-85% lifetime risk of breast cancer) mean age 37 (21-63) (113 BRCA1 and 15 BRCA2) High risk, *n* = 621 (30-50% lifetime risk of breast cancer), mean age 38 (22-70) Moderate risk, *n* = 449 (15-30% lifetime risk of breast cancer), mean age 38 (25-70)		Mode of detection: 9 were detected by mammography alone and 12 were detected with CBE and mammography. In addition, 3 were detected by MRI alone (indication was dense breast tissue in 1 and the other 2 were mutation carriers) and one by MRI and CBE. One was detected by CBE alone. Cancer detection rate with mammography was 5.8/1,000 p/years. CBE was 3.9/1,000 p/years. Combined XRM and CBE was 6/1,000 p/years. The cancer detection was also stratified by age groups and was found to increase with age, from 2.2/1,000 p/years in the <40 year group to 8.5/1,000 p/years in the 40-49year group, to 28.3/1,000 p/years in the >50 year group. (these figures have been recalculated to include DCIS and exclude interval tumours). Tumour size, stage and node status: (unfortunately, these data contain the interval tumour characteristics and they are not possible to separate out)	In those at high risk/mutation carriers <40 years there may need to be more intensive surveillance or the use of additional modalities such as MRI. This is currently under investigation in several countries including the Netherlands. With respect to detection rates we found, as expected, clear trends with age and genetic risk groups. The percentage of DCIS and of node-positive tumours was comparable with the results of the Dutch National Breast Screening Programme. Reviewers' conclusions: It appears from the comparison with tumours arising in the 40-50 year old population that the groups selected were at increased risk and therefore were stratified appropriately. It appears that in the mutation and high-risk group, the surveillance interval (although annual) may still be too long as all interval cancers occurred in these groups. The study demonstrates the value of using mammography in addition to CBE in the surveillance of women at high risk of breast cancer and also suggests that in the higher risk groups additional modalities may be of benefit. The presentation of cancer detection rates from surveillance that included interval cancers and excluded *in situ* cancers was misleading. However, there was sufficient raw data provided to recalculate these results appropriately.

Table 11. Primary research studies appraised investigating the accuracy and efficacy of mammographic surveillance compared to usual care on outcomes from breast cancer (continued)

Source	Study design Evidence grading	Comparison interventions and dates of testing	Sample	Outcomes and Verification	Results	Comments
Brekelmans et al. (2001) Erasmus University, Rotterdam, The Netherlands (Rotterdam Family Cancer Clinic) *Continued*					4 tumours were *in situ* and the rest were invasive. Of the invasive, 28 were ductal carcinomas, 2 were lobular and 1 was medullary. 10 tumours were sized <10 mm (2 in the carrier group, 7 in the high- risk group and 1 in the moderate risk group). 8 were 10-15mm and 11 were >15mm (range 16-40mm). The size was not able to be identified in two. 11 were node–positive – 2 prevalent, 6 incident and 3 interval. There were more node-positive tumours in proven carriers and the youngest age group, but these results were not statistically significant. Sensitivity: CBE 40% XRM 60% Both CBE and XRM 66% Interval tumours: There were 9 interval tumours, 4 in the carrier group and 5 in the high-risk group. The interval cancer rate was 2.5 /1,000 p/years. The interval cancers were all invasive and 3 were node-positive. Unfortunately, details of size are presented for all tumours together and it is not possible to elucidate these characteristics for the interval tumours.	This study also suggests equivalent sensitivity to the Dutch Breast Screening Programme and suggests benefit of XRM in addition to CBE.

Table 11. Primary research studies appraised investigating the accuracy and efficacy of mammographic surveillance compared to usual care on outcomes from breast cancer (continued)

Source	Study design Evidence grading	Comparison interventions and dates of testing	Sample	Outcomes and Verification	Results	Comments
Brekelmans et al. (2001) Erasmus University, Rotterdam, The Netherlands (Rotterdam Family Cancer Clinic) *Continued*					Interval tumour rate = 2.5 per 1,000 person/years The time interval from last negative surveillance result to diagnosis ranged from 8 weeks to 10 months. Mortality: 3 patients died, 2 of metastatic breast cancer and 1 of another cause. Comparison: The ratio of observed (in study) to expected (in National Cancer Registry) breast cancers was 7 overall. This was 23.7 in the carrier group, 7 in the high-risk group and 2.7 in the moderate-risk group) The sensitivity of this surveillance regimen overall is 74% (26/35) (63%, if the 4 surveillance detected tumours that were causing symptoms before detection are removed). The sensitivity is also stratified by the risk groups and by age. For risk group the results are 100% (4/4) in the moderate group, 77% (17/22) in the high-risk group and 56% (5/9) in the mutation carrier group.	

Table 11. Primary research studies appraised investigating the accuracy and efficacy of mammographic surveillance compared to usual care on outcomes from breast cancer (continued)

Source	Study design Evidence grading	Comparison interventions and dates of testing	Sample	Outcomes and Verification	Results	Comments
Brekelmans et al. (2001) Erasmus University, Rotterdam, The Netherlands (Rotterdam Family Cancer Clinic) *Continued*					For age group the results are 63% (5/8) for women <40 years, 73% (8/11) for women 40-49 years and 81% (13/16 for women aged ≥ 50 years. The differences between age groups and risk groups are not statistically significant. The sensitivity of the modalities of surveillance are: CBE 40% 14/35 XRM 60% 21/35 CBE + XRM 66% 23/35 (MRI 74% 26/35) The comparison of results with the Dutch Breast Screening Programme shows similar results.	

Table 11. Primary research studies appraised investigating the accuracy and efficacy of mammographic surveillance compared to usual care on outcomes from breast cancer (continued)

Source	Study design Evidence grading	Comparison interventions and dates of testing	Sample	Outcomes and Verification	Results	Comments
Gui et al. (2001) London, UK.	Prospective cohort study III-2 (C1 P1 Q2)	Surveillance protocol: Annual mammography and CBE from age 35 years. If the index case developed breast cancer <40 years, surveillance in the unaffected relative started 5 years younger than the earliest diagnosis within the family (with a lower limit of 25 years). Comparisons were made between a group of women at standard risk, and a group of women at moderate/high risk of breast cancer. Comparison was also made between these groups and data from the Office of National Statistics on the population incidence of breast cancer and data on cancer detection rates in the NHSBSP. Comparisons were also drawn between mammography alone or in combination with CBE. Dates – commenced in June 1993..	Sample no = 2,578 women, 1,500 at standard risk and 1,078 at moderate/high risk of breast cancer. Median age at start of study was 44 in the moderate/high risk group and 48 in the standard risk group. Median age (at diagnosis) in the moderate/high risk group= 45 years (range 26-66 years) (median age at diagnosis in the group at standard risk was 54.4 years, range 38-63 years)(p=0.03) Recruited from Breast Diagnostic Unit (BDU) which they were referred to by GP due to a family history of breast cancer. Inclusion criteria • pre-existing patients of the BDU; • known to be cancer-free as of June 1993.	Relevant outcomes: • cancer detection rate; • mode of detection; • tumour size; • tumour stage; • node status ; • interval tumours; • mortality; • sensitivity. Verification of a positive result was done pathologically. Verification of a negative result consisted of follow-up – the mean follow-up period calculates out to be 3.9 years (5902 women/yrs) for the standard risk group and 4 years (4327.8 women/yrs) for the mod/high risk group. (Total = 10229.8 women/yrs)	Cancer detection: 31 cancers were detected in total – 12 in the standard risk group and 19 in the moderate/high risk group. Cancer detection rate was therefore 2 per 1,000 woman/years and 4.4 per 1,000 woman/years for the standard and moderate/high risk groups respectively. Overall, it was 3 per 1,000 woman/years. However, these rates are misleading as they include interval cancers† All were incident as entry criteria stipulated that the women had received surveillance before at the BDU and were known to be cancer-free. Mode of detection: In the standard risk group, 6 tumours were detectable by both CBE and mammography, 4 were detectable by CBE alone and 2 were detectable by mammography alone. In the moderate/high risk group, 6 tumours were also detectable by CBE and mammography, 10 were detectable only by CBE and 3 were detectable only by mammography.	Limitations include: The results are misleading in the way that they are presented, including interval tumours. There are few characteristics given about the samples and whether all who were eligible to participate did participate. There is no information on risk reduction (BSO or Tamoxifen) or risk factors such as HRT or OCP use). No mention of blinding of radiologists to risk. Verification bias is likely. Authors' conclusions: The relatively short follow-up of this and other studies suggests that no firm conclusions can be drawn on the survival benefit of surveillance for moderate/high risk women. However, the NHSBSP detects breast cancer in the incident screen of about 3.8 cancers per 1,000 visits in women over 50 years screened by three yearly mammography. If this is considered acceptable then the results of this study suggest that surveillance women under the age of 50 years who are at moderate/high risk of breast cancer has a similar breast cancer detection rate. The authors also emphasise the need for CBE and mammography. Approximately 80% of the invasive cancers detected were 20mm or less in size and this is likely to translate to a survival benefit with longer follow-up

Table 11. Primary research studies appraised investigating the accuracy and efficacy of mammographic surveillance compared to usual care on outcomes from breast cancer (continued)

Source	Study design Evidence grading	Comparison interventions and dates of testing	Sample	Outcomes and Verification	Results	Comments
Gui et al. (2001) London, UK. *Continued*		End date unclear, but there were 5,902 women years of follow-up in the standard risk group and 4,327.8 women years of follow-up in those at moderate/high risk No comment on method used to classify screens and the cut-off for an abnormal examination.	Risk stratified by the Claus model (family history factors, including age of diagnosis). Divided into 2 groups: women at standard risk (lifetime risk < 1:6 (=16%)) and women at moderate/high risk (lifetime risk ≥1:6 (=16%)). A more detailed explanation of stratification is in original article. Within the group at moderate/high risk 233 (21.6%) had a predicted lifetime risk of >1:4 (25%)		Tumour size, stage and node status: Standard risk group had all invasive tumours, the tumour size ranged from 10 to 30 mm and 6 were node-negative, 4 were node-positive and 2 had unknown node status. The moderate/high risk group had 17 invasive tumours detected (2 which were of unknown pathological type) and 2 *in situ*. The tumour size ranged from 10-30mm (2 also unknown). Lymph node status was negative for 9 tumours, positive for 6 tumours and unknown for 2. Interval tumours: The data on interval tumours is difficult to interpret. It is stated that of the 26 cancers that were detectable by CBE, 17 (65%) were detected by patients and 9 (35%) were found by the clinician. The patient-detected clinical abnormalities presented as 15 interval cancers, while 2 patients waited for their routine clinical appointment to report their findings. This appears to be a high number of interval cancers, however it is not stipulated how many were in the standard group and how many in the high-risk group.	However, no firm conclusions can be drawn at this time. In conclusion, the study supports the effectiveness of a surveillance programme for women with a family history, selected according to prior probability. Reviewers' conclusions This study confirms women selected by the Claus model to be at moderate to high risk have a higher incidence of breast cancer. Regarding surveillance, the results are somewhat misleading as the cancer detection rate appears to be calculated with the inclusion of interval tumours. If these 17 tumours are removed from the total, the overall cancer detection rate for both risk categories is 1.4 per 1,000 visits. This would be much lower than the NHSBSP. The cancer detection rate for the moderate/high risk group, once the interval cancers are removed, might be higher than this but unfortunately it is not possible to calculate this as the interval cancers are not ascribed to one group or another. The results also emphasise the importance of using CBE in addition to mammography. The cancer detection rate of CBE is 2.5 per 1,000 woman years compared with 1.6 per 1,000 woman/yrs for mammography and 3 per 1,000 woman/years for CBE + mammography. This is again misleading as 17 of the tumours detected by CBE were interval tumours. The recalculated rate for CBE is 0.9 per 1,000 woman/years.

Table 11. Primary research studies appraised investigating the accuracy and efficacy of mammographic surveillance compared to usual care on outcomes from breast cancer (continued)

Source	Study design Evidence grading	Comparison interventions and dates of testing	Sample	Outcomes and Verification	Results	Comments
Gui et al. (2001) London, UK. *Continued*					It is misleading that these tumours are included in the cancer detection rates and other results. If the patient-detected tumours are removed, the cancer detection rate overall drops from 3 per 1,000 women/years to 1.4 per 1,000 women/years. Unfortunately there is not enough data to calculate this by risk group or by modality of surveillance. Mortality: One woman in each group has died of metastasized disease. Comparisons: The relative risk of developing breast cancer in the moderate/high risk group compared with the standard risk group was 2.6 (95% CI, 1.2-5.8). The relative risks for incidence of the study groups compared with the population incidence from the national statistics was 1.1 (95%CI, 0.6-1.8) for the standard risk group and 2.8 (95% CI 1.7-4.2) for the mod/high risk group. The comparison is made of the NHSBSP cancer detection rate of 3.8 per 1,000 to the rate of 4.4 per 1,000 in this study and used as justification for this surveillance.	Unfortunately, it is not possible to calculate the mammography cancer detection rate with the interval tumours removed as the data is not given (some have been included as detected by both and must have had diagnostic mammography). This suggests that in fact mammography might be more beneficial than CBE. The only conclusion that can really be drawn is that CBE combined with mammography detects the most tumours. In light of this the results and conclusions of this study must be treated with some caution.

Table 11. Primary research studies appraised investigating the accuracy and efficacy of mammographic surveillance compared to usual care on outcomes from breast cancer (continued)

Source	Study design Evidence grading	Comparison interventions and dates of testing	Sample	Outcomes and Verification	Results	Comments
Gui et al. (2001) London, UK. *Continued*					However, this is a false comparison as the rate form the study is including interval tumours. Sensitivity: The sensitivity was reported as: Mammography 55% CBE 84% Both 100% However, these figures are misleading as they include the interval cancers. The recalculated figure for both modalities combined is 45% and for CBE is 29%. Unfortunately, it is not possible to recalculate the sensitivity for mammography as some of the interval cancers must have had mammograms after detection by examination but the figure of how many of these were positive and how many were negative is not given. When stratified by risk group the sensitivities (including interval tumours) are: Standard risk CBE 83% XRM 66% Moderate/high risk CBE 84% XRM 47%	

Table 11. Primary research studies appraised investigating the accuracy and efficacy of mammographic surveillance compared to usual care on outcomes from breast cancer (continued)

Source	Study design Evidence grading	Comparison interventions and dates of testing	Sample	Outcomes and Verification	Results	Comments
Hou et al. 2002 Taiwan	Prospective cohort Study III-2 (C1 P1 Q2)	Surveillance protocol was annual CBE, mammography and US. Pre-menopausal women received surveillance during the 2nd week of the menstrual cycle to minimise the occurrence of breast densities or enhancing masses related to the menstrual cycle. 4-view film mammograms were conducted and reviewed by one radiologist. (US performed with a 7.5 MHz frequency transducer probe – the US results will be commented on in a later chapter). Dates – May 1994 to August 2001. No comparisons were made in this study BIRADS was used to classify the images and a cut-off of 4 or above was abnormal.	Sample no = 935 women Mean age (at surveillance)= 48.6years (range 35-75) Recruited as relatives of breast cancer cases in hospital. Inclusion criteria: • >35 years old; • female relatives of breast cancer patients (mothers, daughters, grandmothers, sisters). Exclusion criteria: • pregnant or lactating; • past history of breast cancer • known metastatic diseases Risk stratified – no specific risk stratification process carried out. Just all relatives of breast cancer patients.	Relevant outcomes • cancer detection rate; • mode of detection; • tumour stage; • node status; • interval tumours; • 5-year overall survival and event-free survival (free from cancer related death and tumour spread) • sensitivity; • specificity. Verification of positive result by any of three surveillance modalities was through biopsy and pathology results. Verification of a negative result was through follow-up. Median follow-up was 41.8 months (range 12-82 months) Verification of interval cancers.	Cancer detection rate: 21 cancers were detected, giving an overall cancer detection rate of 22% per 1,000 women under surveillance. Of the women with tumours, 1 was a BRCA1 mutation carrier, 2 were BRCA2 mutation carriers and the other 18 were mutation status unknown. Mode of detection: CBE detected 7 tumours. Mammography detected 11 tumours. Tumour size, stage and node status: 16 were invasive cancers, 2 were DCIS, 2 were mucinous carcinomas and 1 was a medullary carcinoma. Mean tumour size was 12mm. 7 were node-positive and 14 were node-negative. 1 interval cancer was reported. Five-year overall survival was 90.4% and the disease-free survival rate was 80.9%. The documented sensitivities were: CBE 33.3%	Limitations include: Verification bias is likely. Lead-time bias and length bias are likely in terms of the survival data. This population was not explicitly risk stratified and it is difficult to assess their overall risk of breast cancer. There are no characteristics of the overall group of women under surveillance, other than being relatives of breast cancer patients and the mean age. It is unclear if they have any additional risks for breast cancer. Only a prevalent round was examined and it is likely that the cancer detection rate would be higher in this round than in subsequent rounds. There is no mention of how interval cancers are verified as being true interval cancers. Authors' conclusions: Based on a higher sensitivity of sonography for detecting breast cancer in the high-risk group in our study, sonography is superior to mammography and physical examination of the breasts in the surveillance of women at high risk for breast cancer in Taiwan. If sonography will replace mammography as a surveillance tool, it needs further research. Otherwise, the low cost of US and convenience for women who live in rural areas suggests that sonography will be a useful tool for screening breast cancer in Taiwanese women in the high-risk group and in countries with a low incidence of breast cancer.

Table 11. Primary research studies appraised investigating the accuracy and efficacy of mammographic surveillance compared to usual care on outcomes from breast cancer (continued)

Source	Study design Evidence grading	Comparison interventions and dates of testing	Sample	Outcomes and Verification	Results	Comments
Hou et al. 2002 Taiwan *Continued*					Mammography 52.4% The documented specificities were: CBE 83.5% Mammography 82.2% If calculated with the interval tumour in the denominator the sensitivities are: CBE 31.8% (95% CI, 13.9-54.9%). XRM 50% (95% CI, 28.2 to 71.9). The calculated specificities are: CBE 99.4% (95% CI, 98.7-99.8%). XRM 99.6% (95% CI, 98.9-99.9%). No measures of statistical significance are presented in this paper.	Reviewers' conclusions: This study suggests that sonography is much more effective in the surveillance of women at high risk of breast cancer than mammography or CBE. However, these findings are specific to this population and are not generalisable. As discussed by the authors, the sensitivity of mammography is likely reduced by the higher proportion of Asian women with smaller denser breasts, which are less fatty, and also the overall lower incidence of breast cancer in this Taiwanese population. Sonography may be a useful modality of surveillance in these women, and especially in rural areas or areas without access to MRI, however it is unlikely to achieve such good results in a Western population.

Table 11. Primary research studies appraised investigating the accuracy and efficacy of mammographic surveillance compared to usual care on outcomes from breast cancer (continued)

Source	Study design Evidence grading	Comparison interventions and dates of testing	Sample	Outcomes and Verification	Results	Comments
Scheuer et al. (2002) USA	Prospective cohort Study III-2 (C1 P2 Q2)	Surveillance advised was: SBE monthly from age 18 years. CBE 2-4 times per year from age 25 years. Mammography annually from age 25 years. Some also received US and MRI, but there is little information on these modalities. Comparisons are made between the different modalities used, primarily CBE and mammography (as US and MRI not used in a consistent/surveillance test manner) Dates of study recruitment were from May 1995 to October 2000.	Sample no = 165 women were followed. These were recruited from 251patients (who consented to participate) out of 267 BRCA1 or BRCA2 mutation carriers, identified from patients who received genetic test results at a single cancer centre. The 165 were those who had not had and did not choose to have a bilateral mastectomy. Two- thirds of the 251 were BRCA1 and one-third BRCA2. 59.4% of the 251 had a prior personal history of breast cancer. 5.2% of the 251 women had undergone risk reducing bilateral oophorectomy. 8 individuals declined to participate and 8 were lost to follow-up. Mean age of testing for mutations was 47.7yrs (range 24 to 79 yrs)	Relevant outcomes: • cancer detection; • mode of detection; • tumour size, stage and node status; • Sensitivity. Verification of the self-reported history was performed by case note review. All tumours were confirmed by a pathology report. Negative results were verified by follow-up, but it does not appear that the case notes were reviewed for verification of negative disease status. If the patient could not be contacted for either follow-up, then their physician was contacted for follow-up details. Mean follow-up was 24.8 months (range 1.6-66) .	Cancer detection: 7 breast tumours were identified at surveillance (42 per 1,000 women under surveillance) 6 out of the 7 had a prior history of breast cancer). Mode of detection: 6 tumours were detected by radiographic means – 5 by mammography and 1 by MRI in a woman with an unremarkable mammogram and US. 6 tumours were found by physical examination between radiographic surveillance intervals – 5 by SBE (interval tumours) and 1 by CBE. Tumour size, stage and node status: Of the tumours radiographically detected, 3 were invasive and 3 were non-invasive. All of the 3 invasive tumours were less than 20mm in size. 1 was lymph node positive. The one tumour identified by CBE and not mammography was 25mm in size, invasive and lymph-node positive (it was mammographically visible at the time of detection; but this was at a time when mammography was not scheduled).	Limitations include: The determination of outcomes was self-reported and there may have been misclassification and recall bias. However this would have been reduced by the fact that check-ups of medical records were done for verification and pathological reports were reviewed for all tumours diagnosed. Verification bias is likely as it appears that self-reported negative results did not undergo case note review. It is not very clear whether the tumours identified by mammography were also palpable or not. This is a high-risk group of all BRCA1 or BRCA2 carriers and also with a high proportion of women with a prior history of breast cancer. There was also a proportion that had had BSO and this would reduce their risk. Therefore, it was hard to judge the risk of this group as a whole. Relatively short follow-up. Quite small sample. Authors' conclusions: The detection of early-stage tumours in this series was achieved despite a low sensitivity of radiographic breast cancer surveillance. More frequent mammographic examination, breast US and MRI offer potential options to improve sensitivity of breast cancer surveillance in genetically predisposed individuals.

Table 11. Primary research studies appraised investigating the accuracy and efficacy of mammographic surveillance compared to usual care on outcomes from breast cancer (continued)

Source	Study design Evidence grading	Comparison interventions and dates of testing	Sample	Outcomes and Verification	Results	Comments
Scheuer et al. (2002) USA *Continued*			Recruited out of 1,865 patients who received genetic test results from a cancer centre	There were 344 woman/years of follow-up in those that had not had prior bilateral mastectomy. There were 221 women/years of follow-up in those who had not undergone bilateral oophorectomy	Interval tumours: 5 interval tumours were identified by SBE. Mammograms had been unremarkable within 6 to 10 months in 5 cases and the remaining woman had deferred mammography due to pregnancy and last received surveillance 1.5 years before diagnosis. Of these tumours, all were ≤18mm in size and only one was lymph-node positive. Sensitivity: Sensitivity for mammography was 42% (5/12) – which is much lower than other trials. Sensitivity for CBE and mammography combined was 50% (6/12). Unfortunately, it is not possible to calculate the sensitivity of CBE as there were no figures of which of the mammographically detected tumours were also palpable at CBE.	Larger prospective trials comparing frequency and modalities of cancer surveillance as well as the role of risk-reducing operations are necessary to determine optimal management of patients at hereditary risk for these malignancies. Reviewers' conclusions: This is an interesting study, although it suffers somewhat from its design. There is not clear information produced on mammographic surveillance as there was a use of MRI and US which is not fully reported. The method of gathering self-reported information is also likely to have introduced misclassification bias and verification bias as the negative results reported were less thoroughly verified than the positive results. This was also a very high risk population with all BRCA1 or BRCA2 mutation carriers, with a high rate of prior personal history of breast cancer. However, there were also women included who had undergone risk-reducing surgery and/or were taking Tamoxifen, and this will have confounded the results. Lastly, it was not altogether clear which women were complying with the advised surveillance regime. Compliance was 83.3% for any BSE, 97.4% for any CBE and 93.4% for any mammography, but less then this for the recommended intervals. The study suggests that mammography is superior in the detection of tumours to CBE.

Table 11. Primary research studies appraised investigating the accuracy and efficacy of mammographic surveillance compared to usual care on outcomes from breast cancer (continued)

Source	Study design Evidence grading	Comparison interventions and dates of testing	Sample	Outcomes and Verification	Results	Comments
Scheuer et al. (2002) USA *Continued*						However, it is difficult to judge as information is not presented on how many of the mammographically detected tumours were also detected by CBE. There were still a considerable number of interval tumours arising. This suggests, as discussed by the authors, that more regular surveillance and other modalities of screening need to be investigated for this high risk population.,

Table 11. Primary research studies appraised investigating the accuracy and efficacy of mammographic surveillance compared to usual care on outcomes from breast cancer (continued)

Source	Study design Evidence grading	Comparison interventions and dates of testing	Sample	Outcomes and Verification	Results	Comments
Trecate et al. (2003) Italy (NB: Podo is an author on this one as well but we cannot find any further reports from the Podo et al trial.)	Prospective cohort study III-2 (C1 P2 Q3)	Surveillance protocol: Outlined in full in the paper and was dependent on age group. CBE was performed every 6 months for all ages. Mammography was annual and commenced at 25 years with bilateral one-view, and then increased to bilateral double-view from 30 years and over. Double-view was performed in craniocaudal and mediolateral oblique projections. One-view was performed in the mediolateral oblique projection for younger women. Annual US was performed alone from 20-25 years, then with mammography from 25-35 years, then 6 months after mammography from 35-40 years and above 40 years only if requested by the radiologist. US was performed with either 7.5 or 10-12 MHZ probes (ATL HDI 3500, Philips).	Sample no = 23 women at high risk of breast cancer (2 cases did not get US). No average age of women given, range was 30-61 years. Inclusion criteria • BRCA1 or BRCA2 mutation carrier or 1 in 2 probability to be a mutation carrier (on the basis of positive mutational analysis in close relatives). With a negative or positive personal history for breast or ovarian cancer OR • high risk for breast cancer according to criteria specified in paper. Risk stratification: As above, either BRCA1 or BRCA2 carrier, 1 in 2 probability of being a carrier or >50% risk of carrying a susceptibility gene for familial breast cancer on basis of family history.	Relevant outcomes: • cancer detection rate; • mode of detection; • tumour size and stage. Verification of positive results was with pathology and verification of negative results was with follow-up. There is no mention of the mean length of follow-up.	Cancer detection: 4 breast cancers were detected overall. Mode of detection: 3 were detectable by CBE but none of the tumours were detected by mammography. Tumour size and stage: All 4 tumours were invasive: 2 ductal invasive carcinomas, 1 lobular invasive carcinoma and 1 which was mixed ductal and lobular. 2 occurred in mutation carriers and 2 in women at high risk through family history. Only 2 tumours had the size recorded and these were 10mm and 30mm. No record of nodal status was given. Tumours were not stratified by mode of detection or compared to tumours in a population that did not receive surveillance. There was no mention of interval tumours.	Limitations include Small sample size. There are few characteristics given of the women selected other then their risk assessment. There is no information on how they were selected and the characteristics of any women who did not agree to participate. There is no mention of mean age, reproductive history, exogenous hormone use or preventative strategies (i.e. Tamoxifen use or BSO). There is also no indication of which women were having prevalent or incident imaging and for how long they were followed up in the study. There is likely verification bias and this is more likely, the shorter the follow-up period. Authors' conclusions: The authors' conclusions only relate to the MRI component of the study and do not refer to XRM. Breast MRI demonstrated to be a very useful technique for investigating breast disease. It is not influenced by breast density and does not use ionising radiation. For these reasons, it has been proposed to support mammography in the surveillance of BRCA-mutated patients. Moreover, according to the reported results, breast MRI seems very helpful in the high-risk patients group.

Table 11. Primary research studies appraised investigating the accuracy and efficacy of mammographic surveillance compared to usual care on outcomes from breast cancer (continued)

Source	Study design Evidence grading	Comparison interventions and dates of testing	Sample	Outcomes and Verification	Results	Comments
Trecate et al. (2003) Italy *Continued*		MRI was performed annually for all ages for 2 years during the study. A Siemens Vision 1.5 was used with a dedicated double coil. One pre-contrast image and 5 post-contrast images were taken. The contrast agent was Gd-DTPA at 0.1mmol/kg. The method of interpreting the MRI or mammography is not presented. The study was conducted over a 7-month period, however the exact dates are not given.	The latter refers to at least 3 cases of breast cancer before 60 years of age, at least 3 cases of breast cancer before 60 years of age and ovarian cancer at any age, or at least 3 cases of breast cancer before 60 years of age and male breast carcinoma at any age. 5 of the women had a personal history of breast cancer, 1 for ovarian cancer and 1 for ovarian and breast cancer. 1 had had a mastectomy, but the others had conservative surgery combined with radiation therapy.			We believe the breast MRI can be very useful within this kind of surveillance, with a less invasive approach to the disease. In the case of confirmed good diagnostic results, it could be proposed to be used every other year as an alternative to mammography. Reviewers' conclusions This study suggests that CBE detected more tumours than XRM in women at high risk of breast cancer. However, the sample is very small, as is the number of tumours detected, and it is difficult to know how long the women were followed up for and this would affect the reliability of the results. There could be false negatives that had not yet come to light. There is also a specific method of risk stratification in this study, which includes women with a personal history of breast cancer (although only if they are BRCA1 or 2 mutation carriers), and this will affect the generalisability of the study. In addition the results are not presented in a very clear manner and it is difficult to determine the overall sensitivity and specificity for all the surveillance modalities utilized, which would have been valuable information.

Table 11. Primary research studies appraised investigating the accuracy and efficacy of mammographic surveillance compared to usual care on outcomes from breast cancer (continued)

Source	Study design Evidence grading	Comparison interventions and dates of testing	Sample	Outcomes and Verification	Results	Comments
Kriege et al. 2004 NEJM	Prospective multicentre cohort study Grade III-2 (C1 P1 Q2)	Clinical breast examination: performed by an experienced physician every six months. Imaging studies performed annually by radiologists. XRM: oblique and cranio-caudal views and if necessary, compression views or magnifications. MRI: dynamic breast MRI with gadolinium-containing contrast medium according to a standard protocol. Whenever possible, both imaging investigations were performed on the same day or in the same time period, between days 5-15 of the menstrual cycle. BIRADS was used to classify the tumours and the results are presented according to 2 cut-offs, 3 and over and 4 and over.	1,952 recruited and 1,909 women with a genetic risk for breast cancer. Mean age 40 years (range 19-72). Within the group of 358 carriers of pathogenic mutations, 276 had BRCA1 mutation, 77 had a BRCA2 mutation, 1 woman had BRCA1 and BRCA2 mutations, 2 had a PTEN mutation and 2 had a TP53 mutation. Inclusion criteria: Cumulative lifetime risk of breast cancer 15% or more owing to a familial or genetic predisposition and age 25-70 years. Women could be tested at an age younger than 25 if they had a family history of breast cancer being diagnosed before the age of 30 years since testing began at an age 5 years younger than that at which the youngest family member was found to have cancer.	Outcomes of relevance: • cancer detection • mode of detection • sensitivity • specificity • AUC • tumour characteristics The results of each exam were blinded so that the two examinations were not linked. When one of the imaging exams was a BI-RADS 3 or 0 ('need additional imaging evaluation') further investigation by USS with or without fine-needle aspiration was advised, or MRI or XRM was repeated. When one of the two exams was BI-RADS 4 or 5, a cytologic or histologic evaluation of a biopsy specimen was performed. When the results of XRM and MRI were negative but the findings on CBE were rated as uncertain or suspicious, additional investigations were also performed.	Cancer detection: 51 malignant tumours (44 invasive breast cancers, 6 DCIS and 1 non-Hodgkin's lymphoma) arose. 45 of the breast tumours were screen-detected and 5 were interval tumours. The figures were all calculated including the 5 interval tumours but excluding 5 tumours that did not have sufficient data. It is not possible to recalculate these without the interval tumours as it is not clear, once stratified what groups they would be in. Mode of detection: 3 tumours were detected by CBE (cut-off of 'suspicious'). 11 tumours were detected by XRM at BIRADS cut-off of 4m and 18 at a BIRADS cut-off of 3. Sensitivity (95% CI): CBE 6.7% (1.4 to 18.3%) XRM 24.4% (12.9 to 39.5) BIRADS 4 XRM 40% (25.7 to 55.7) BIRADS 3 Specificity (95% CI): CBE 99.9% (99.8 to 99.9%) XRM 99.6% (99.4 to 99.8%) BIRADS 4 XRM 94.9% (94.3 to 95.6%0 BIRADS 3	Of the 1,952 women included, 8 withdrew from the study before the first visit and another 35 were excluded because they ultimately proved not to be carriers in a family with a proven mutation and therefore had a less than 15% lifetime risk of developing breast cancer. Of the 1,909 remaining women, 88 (4.6%) left the study or were lost to surveillance before October 2003. 65 of these 88 women underwent prophylactic mastectomy. Another 89 women (4.7%) remained under surveillance but later refused surveillance by MRI because of claustrophobia or other reasons. The characteristics of the women are given and include the number with previous surveillance, menopausal status, HRT and OCP use and BSO (7%). Of the 20 cancers not detected by XRM or CBE, 11 of the 19 invasive tumours were smaller than 10mm and only 1 was associated with a positive node. Larger tumours (>2cm diameter) were found more often in women with BRCA1, BRCA2, PTEN, and TP 53 mutations than in the other 2 risk groups in the study, suggesting that more frequent surveillance is needed in these two groups. Authors' conclusions: The authors' conclusions mostly relate to the entire surveillance strategy and especially to MRI.

Table 11. Primary research studies appraised investigating the accuracy and efficacy of mammographic surveillance compared to usual care on outcomes from breast cancer (continued)

Source	Study design Evidence grading	Comparison interventions and dates of testing	Sample	Outcomes and Verification	Results	Comments
Kriege et al. 2004 NEJM *Continued*		Comparisons made between the modalities of surveillance and of tumour characteristics between the study population and 2 control populations of women with symptomatic tumours that did not receive surveillance. Dates of study were November 1999 to October 2003.	Women with symptoms of breast cancer or a personal history of breast cancer were excluded. Recruited from 6 familial cancer clinics in the Netherlands.	The diagnosis of malignant tumours was based on the results of a histologic examination. One of the investigators, an expert pathologist, reviewed all the biopsy specimens that formed the basis for the diagnosis of breast cancer. Mean follow-up was 2.7 years (range 0.1-3.9 years).	PPV: CBE 50% (11.8 to 88.2%) XRM 47.8 (26.8 to 69.4%) BIRADS 4 XRM 8% (4.8 to 12.3%) BIRADS 3 NPV: CBE 98.95 (98.6 to 99.2%) XRM 99.1% (98.8 to 99.4%) BIRADS 4 XRM 99.3% (99.0 to 99.5%) BIRADS 3 Area under ROC: XRM:0.686 Tumour characteristics: There were 44 invasive tumours and 6 DCIS. The number of tumours less than 10mm in size was significantly higher in the study cohort than in symptomatic women without surveillance in both the National Cancer Registry control group ($p<0.001$) and the genetic study control group ($p=0.04$). Lymph nodes were negative in 66.7% (28/42) of the study cohort. This was also significantly higher in the study cohort than the number of node negative tumours in the National Cancer Registry control group ($p<0.001$) and the genetic study control group ($p=0.001$).	The surveillance programme used in this study, especially MRI, can detect breast cancer at an early stage in women at risk for breast cancer. However a drawback of MRI is that it has a lower specificity than XRM and as a result, MRI will generate more findings judged as uncertain, which require short-term follow-up or additional investigations. Reviewers' conclusions: A generally well conducted study with conclusions drawn from the data presented above, and the respective surveillance tests performed either on the same day or within a short period of the first screening test undertaken. The results for CBE and XRM suggest that XRM is more sensitive than CBE for the surveillance of women at high risk of breast cancer, but has equivalent specificity, PPV and NPV at BIRADS 4. Lowering the BIRADS cut-off to 3 increases the sensitivity of XRM but decreases the specificity and the PPV. This is due to a higher number of false positive examinations that arise at a lower cut-off.

Table 11. Primary research studies appraised investigating the accuracy and efficacy of mammographic surveillance compared to usual care on outcomes from breast cancer (continued)

Source	Study design Evidence grading	Comparison interventions and dates of testing	Sample	Outcomes and Verification	Results	Comments
Warner et al. (2004) Ontario and Montreal, Canada	Prospective cohort Study III-2 (C1 P2 Q2)	Study protocol: CBE biannually and mammography, US and MRI all performed annually. 4 modalities all performed the same day. Commencing at least 1 year after the woman's last mammogram. CBE coded as normal, suggestive of benign disease, indeterminate, or suspicious of malignancy. Indeterminate CBE exams were repeated after 3 months. Mammography was conventional 4-view film. Further views done when necessary. MRI was performed with 1.5 T magnet (Signa, General Electrical Medical Systems). The first 38 patients in the first year were done in a single-turn elliptical coil after a bolus injection of 0.1mmol/kg of Gd-DTPA. Images were taken in the coronal plane. For the remaining patients, a phased-array coil arrangement was used..	Sample no. = 236 female BRCA1 and BRCA2 mutation carriers. Mean age at first surveillance 46.6 years (range 25-65 years). Mean age of diagnosis was 47.4 years (33.4-63 years). Recruited from familial cancer clinics Inclusions: • BRCA1 or BRCA2 mutation carrier. Exclusions: • past history of unilateral breast cancer if the contralateral breast not intact; • pregnant or lactating women (participation deferred); • history of bilateral breast cancer, currently undergoing chemotherapy or known to have metastatic disease;	Relevant Outcomes: • cancer detection rate; • mode of detection; • tumour stage, size and node status; • interval cancers; • mortality; • sensitivity; • specificity; • PPV; • NPV; • ROC curves. NB: the PPV and specificity do not include in the denominator women who had additional diagnostic studies that did not result in biopsy. Verification of positive results was by pathology. Biopsy was undertaken if there was suspicion from any of the 4 modalities of surveillance. Verification of a negative result was through follow-up.	Cancer detection: 22 cancers were detected in 21 women (1 bilateral). (7 of these women had previous breast cancer). Mode of detection: 2 were detected by CBE (9.1%), 8 by mammography (36%). 2 cancers (9.1%) were detected by mammography alone. Tumour stage, size and node status: 6 tumours were DCIS and 16 were invasive (15 infiltrating ductal and 1 invasive lobular). The mean size of the invasive tumours was 11mm at the first surveillance round and 13mm at the second round (overall range 5mm-60mm). 15 cases were node sampled and 2 were node positive. Interval cancers: There was only 1 interval cancer, detected in a 40 year old BRCA1 mutation carrier 7 months after her 3rd screen. Retrospectively, this tumour was visible on MRI and on mammography at last surveillance visit.	Limitations include: Likely verification bias. Selected participants are very high risk, being proven mutation carriers and also including those with a prior history of breast cancer. It is not clear which images were incident and which were prevalent, and which tumours were detected at which round. A large number of women had had prior mammography. No mention of whether women had had risk reducing measures such as bilateral salpingo oophorectomy or Tamoxifen. Was quite high level of attrition in the study and the characteristics of those women are not outlined. This may have introduced bias. Authors' conclusions: Relate to overall surveillance strategy as a whole. This study of BRCA mutation carriers demonstrates that the addition of annual MRI and US to mammography and CBE significantly improves the surveillance for detecting early breast cancers. The use of US did detect additional tumours, but had a high false-positive rate and in light of this its benefit remains to be seen. There was no observed benefit from CBE over and above the 3 imaging modalities.

Table 11. Primary research studies appraised investigating the accuracy and efficacy of mammographic surveillance compared to usual care on outcomes from breast cancer (continued)

Source	Study design Evidence grading	Comparison interventions and dates of testing	Sample	Outcomes and Verification	Results	Comments
Warner et al. (2004) Ontario and Montreal, Canada *Continued*		This provided sagital images. US used a 7.5MHz transducer (the first 7 patients did not receive US) Each imaging modality was read independently by a radiologist and scored on the 5 point BIRADS scale. All lesions with a score of 4 or 5 were biopsied. Pre-menopausal women had surveillance performed mid-menstrual cycle to avoid changes due to cyclical hormonal variation. Radiologists were blinded to the results of CBE, 31 women left the study before completing 3 rounds, 16 underwent bilateral mastectomy, 3 were too large for MRI machine, 3 stopped due to pregnancy, 4 developed metastatic cancers, 4 were lost to follow-up and 1 did not wish to continue participating.	• Women weighing >91kg (technical reasons). Risk stratification not really performed as only BRCA. mutation carriers included (all very high risk group). There were 137 (58%) BRCA1 mutation carriers and 99 (42%) BRCA2 mutation carriers. 31% were Ashkenazi Jews. 30% had a history of breast cancer, 9% a history of ovarian cancer and 60% had no history of cancer or a history of another type of cancer. 85% of the women (*n*=205) had had mammography within the last 15 months and therefore this was an incident rather than a prevalent round for them. 45% were pre-menopausal and 55% were post-menopausal.	All patients were followed up for a minimum of 1 year from the date of the last surveillance examination.	Another woman, who elected to have a bilateral mastectomy after breast cancer was found, had a 2mm focus of DCIS in the contralateral breast which had not shown up on surveillance 2 months earlier. Mortality: All 22 patients who had tumours diagnosed were still alive and disease-free at the time the article was written. It was felt that the cancers detected on the second surveillance round were of an earlier stage. The 2 node-positive tumours were detected in the first round. However, it was not exactly clear that the first round was really a prevalent round as a high percentage of women had had prior mammography. It was found that false-positives and false-negatives decreased from the first to the second and then to the third round of surveillance. The measures of accuracy are therefore presented by the modality of surveillance and by the year of the surveillance. These can be seen in the paper, but overall values for the 3 years are reported here.	MRI is likely to become the cornerstone of breast cancer surveillance for BRCA1 and BRCA2 mutation carriers, but it is necessary to demonstrate that this surveillance tool lowers breast cancer mortality before it can be recommended for general use. Reviewers' conclusions: This study demonstrates a greater efficacy and higher sensitivity of XRM than CBE in the surveillance of BRCA1 and BRCA2 mutation carriers for breast cancer. The overall findings relate to the surveillance strategy as a whole (including US and MRI). As the authors suggest, the findings do not answer whether this translates into reduced mortality. However, the tumours detected did seem to be of an earlier stage and smaller size, with only 2 tumours node-positive. The results of this study are limited to the very high risk population of women who are proven BRCA1 or BRCA2 mutation carriers and including those with a personal history of breast cancer. It may therefore not be generalisable to all women with an increased risk of breast cancer due to a family history. Further studies with larger numbers and longer follow-up, and including women of other risk groups are required.

Table 11. Primary research studies appraised investigating the accuracy and efficacy of mammographic surveillance compared to usual care on outcomes from breast cancer (continued)

Source	Study design Evidence grading	Comparison interventions and dates of testing	Sample	Outcomes and Verification	Results	Comments
Warner et al. (2004) Ontario and Montreal, Canada *Continued*		All participants underwent the first round, but only 58% the second and 36% the third (a total of 120 women were still undergoing surveillance when the paper was written). No direct comparisons were made in this study, except between modes of detection. Dates of surveillance were between Nov 1997 and March 2003.			Sensitivities of combinations of modalities: XRM + CBE = 45% Measures of accuracy of individual modalities: Sensitivity (95% CI): CBE = 9.0% (1 to 29%) XRM = 36% (17.1 to 59.3%) Specificity: XRM= 99.8% (98.7 to 99.9%) PPV: XRM= 89% (51.7 to 99.7%) NPV: XRM= 97% (94.8 to 98.3%) AUC: XRM= 0.77 CBE = 0.48 Mamm + CBE = 0.77	

Table 11. Primary research studies appraised investigating the accuracy and efficacy of mammographic surveillance compared to usual care on outcomes from breast cancer (continued)

Source	Study design Evidence grading	Comparison interventions and dates of testing	Sample	Outcomes and Verification	Results	Comments
Murday et al. (2004) London and Glasgow, UK	Retrospective cohort study III-2 (C1 P1 Q2)	Surveillance protocol: Annual 2-view mammography and CBE. Also encouraged to perform monthly BSE (adherence to the recommended surveillance interval was commented on as being poor, although it improved with surveillance duration). In addition to the 192 women having mammographic surveillance, 23 had US only. Of these women, 19 were <35 years and 4 were in the age range for mammography but chose US. There were 5 symptomatic, 14 prevalent and 9 incident rounds. Comparisons were made between the groups at different risk levels according to their family history, and also between the different modalities used for surveillance. Comparisons were also made with a control group of probands' sisters .	Sample no =192 women 6 had identified BRCA1mutation in their families. 61 had greater than 50% chance of having a high-risk gene in their family, 35 had a 20-50% risk and 90 had a chance of less than 20% of carrying a high-risk mutation. Mean age (at diagnosis) is not provided but calculates out as 39.9 years (range 29-48 years) Recruited from family cancer clinics in South Thames region. Inclusion criteria • <50 years age at first appointment (prior to Jan 1996); • attended family cancer clinics in South Thames region before 1996; • have a risk estimate that was high enough for surveillance to be recommended under the age of 50 years*.	Relevant outcomes: • cancer detection rate; • mode of detection; • tumour size; • tumour stage; • node status ; • interval tumours. • Specificity • PPV Verification of a positive result was by further investigation and pathology. Diagnostic details were obtained from radiology management systems, hospital information systems, electronic patient records, breast unit notes and case notes. (The entire cohort was checked for diagnostic results regardless of their surveillance results). Negative results also had notes reviewed for diagnostic test results and the follow-up was therefore confirmation.	Cancer detection: 9 cancers were diagnosed by the surveillance (including BSE), 3 at the prevalence round and 3 at the incident round and 3 by BSE (interval cancers). The cancers detected by BSE were said to be part of the 'advised' surveillance but are actually interval tumours. Mode of detection: 6 were visible on mammography (although 4 had no definite features of malignancy seen), 3 of these were also palpable at CBE, and the other 3 were not. Tumour size, stage and node status: Tumour size was up to 90mm, although it is not recorded for 4 tumours. 1 tumour was *in situ* (incident), 1 was *in situ* and invasive and the other 4 were invasive. 2 tumours had lymph node spread (1 detected on prevalent round and the other on incident round, and both in women at high calculated risk of breast cancer) but the others did not.	Limitations include: Retrospective study; Quite small sample; Few characteristics of women presented so cannot judge selection bias. No mention of the use of risk reducing strategies such as BSO or Tamoxifen. Verification bias likely, although may have been reduced by also reviewing diagnostic test results in those with negative surveillance results. The definition of interval cancer is not very clear. This depends on whether BSE is considered to be a part of the surveillance programme. Also, two tumours were never detectable by mammography and therefore were not considered interval tumours but, if detectable by BSE, they should have been detectable by CBE (if not real interval tumours) at the surveillance visit (although that isn't something that can be verified retrospectively). The comparisons between risk groups are interesting, but could only be applied to other surveillance programmes if the same model of risk stratification was adopted. Authors' conclusions: The results of the surveillance were felt to be disappointing with only 3 tumours detected at the incident rounds and 3 presenting as interval tumours. It is suggested that this emphasises the importance of BSE in women at high risk of breast cancer.

Table 11. Primary research studies appraised investigating the accuracy and efficacy of mammographic surveillance compared to usual care on outcomes from breast cancer (continued)

Source	Study design Evidence grading	Comparison interventions and dates of testing	Sample	Outcomes and Verification	Results	Comments
Murday et al. (2004) London and Glasgow, UK *Continued*		They had to be younger than 50 years of age and to have not been diagnosed with cancer at the time of their siblings' first appointment. Consent was sought from the sisters and then the consultants filled in a postal questionnaire about their sisters. This included details of any surveillance they were having and any cancer that occurred with them since the consultants' entry into this surveillance programme. Dates Jan 1996 to Jan 2001.	• have had annual 2-view mammography and CBE in the same hospital where they had had genetic counselling (to allow easy access to results); • previously unaffected by cancer. Risk stratified by national guidelines published in the UK. (*for surveillance, women required a minimum of 1 first-degree relative with breast cancer diagnosed at less than 40 years or 2 relatives diagnosed at less than 60 years, of which one was a first-degree relative.	The process of verification of Interval tumours was not particularly specified. There were 280 person/years of follow-up. This does not appear to correlate with the number of mammograms performed. This may be because some of the mammograms were symptomatic or recall examinations.	Interval tumours: 3 interval tumours were detected by BSE. 2 were never detectable on mammography, 1 was in a BRCA1 mutation carrier, and the other 2 in women with a Cyrillic calculated risk that put them in the medium and high-risk groups. Two were invasive, 1 was *in situ* and none were node positive. In comparing the groups at different risk, it was found that the majority of cancers detected were in the high-risk and BRCA family groups, and all cancers were in groups with over a 20% calculated risk of having a high-risk gene in their family. Comparison with the control group of consultants' sisters had a poor response from the postal questionnaire, with only 45% of those eligible to fill it in returning it. This cohort comprised 90 unaffected sisters. During the follow-up, 3 of the sisters developed breast cancer under the age of 50 years. There is no comment on what, if any, surveillance this cohort were receiving.	This study also suggests that surveillance below the age of 50 years may be unnecessary in families with a low chance of having a BRCA1 or BRCA2 mutations, but it is important for high-risk women to undergo surveillance at least annually and possibly commence at less than 35 years. It is suggested that those at high risk/BRCA carriers should perhaps undergo surveillance even more frequently, perhaps biannually and from an even earlier age. Reviewers' conclusions: The limitations of this study need to be taken into consideration. However, it does emphasise the importance of stratifying risk within women at high risk of breast cancer in order to determine whether surveillance is effective and warranted or not, especially when the negative effects of surveillance are considered. This applies not only to individuals, but the inclusion of women for whom surveillance is not truly warranted may reduce the evaluated efficacy of the surveillance programme. Also, the increased numbers may mean that surveillance is delayed in those who really need it.

Table 11. Primary research studies appraised investigating the accuracy and efficacy of mammographic surveillance compared to usual care on outcomes from breast cancer (continued)

Source	Study design Evidence grading	Comparison interventions and dates of testing	Sample	Outcomes and Verification	Results	Comments
Murday et al. (2004) London and Glasgow, UK *Continued*			Risk was also recalculated using the Claus (family history) model, a cyrillic computer program and modified by Bayes thereon. This was an attempt to take into account the number of unaffected females in the family, without which the risk is overestimated.		They were found not to be a good control group as they were generally younger and difficult to age match and also, although not receiving as much surveillance, were generally (60%) having some kind of follow-up. Sensitivity: The sensitivity of mammography and CBE was reported as 78% (7/9). However, this included 1 interval tumour in the calculation (which retrospectively was visible on the last screen). Without this being included, the sensitivity is 67% (6/9). Specificity: 84% PPV: 9% Unfortunately, there were not adequate data presented to be able to check the calculation of these figures. Presumably the PPV would be affected by including interval tumours in the true positives also. The sensitivity of CBE alone would have been 33% (3/9) and of XRM alone would have been 66% (6/9). No tumours were detected by CBE that were not visible on mammography.	In this study, only one-third of women actually received the recommended interval between their prevalence and first incident round. This is a further limitation of the study as, although this reflects a real-life situation, the interval cancer rate may have been lower if surveillance intervals had been adhered to. The study shows that mammography is more effective at detecting cancer in women at a high risk of breast cancer with a cancer detection rate twice that of CBE. No tumours were detected on CBE that were not visible on mammography

Table 11. Primary research studies appraised investigating the accuracy and efficacy of mammographic surveillance compared to usual care on outcomes from breast cancer (continued)

Source	Study design Evidence grading	Comparison interventions and dates of testing	Sample	Outcomes and Verification	Results	Comments
Banks et al. (2004) UK multi-centre at 10 breast screening units	Prospective cohort Study III-2 (CX P1 Q2)	Surveillance consisted of mammography done 3-yearly Comparisons were made between the overall accuracy of mammography and the various subgroups, including high-risk women. (also smokers, previous breast surgery, previous OCP use, BMI, regular exercise, alcohol intake). Dates of testing were June 1996 to March 1998. Images were classified as abnormal if the patient was referred for further investigation.	Sample no = 122,355 women overall underwent surveillance. Of these10,959 had a high risk of breast cancer due to a family history. Women were aged between 50-64 years. Inclusion criteria: • 50-64 years • no past history of cancer except non-melanoma skin cancer. Women were excluded if results were positive but did not have cancer and were asked to come back earlier than 3 years as were no longer considered to be on routine surveillance. Risk stratification was not thoroughly explained. Family history consisted of having a mother or sister with breast cancer. Recruited from 10 breast screening units that were involved in the million women study.	Relevant outcomes: • sensitivity • specificity These were calculated for the overall cohort and then in various subgroups, including one of high-risk women. Verification of a positive result was through pathology and ICD code. If diagnosed within 3 months of surveillance it was considered surveillance detected. Verification of a negative result was through follow-up. If detected between 3-12 months after surveillance it was considered not to have been surveillance detected. 12-month follow-up of medical records and the NHS central register.	The overall sensitivity for mammography was 86.6% and the specificity was 96.8%. For women at high risk of breast cancer the sensitivity was 83.8% (74.6-90) and the specificity was 97.3% (96.9-97.5%).	Limitations include: Verification bias likely. Few characteristics of the high-risk women described, including little information on their risk status. Unclear if have had any risk-reducing surgery or are on Tamoxifen, or if taking HRT. It is likely that this group may not be very representative of women at high risk altogether as they are receiving 3-yearly normal screening and therefore may not be a very high risk group. The age range of these women is also much older than the age when most women at high risk would begin surveillance and therefore the results do not translate to surveillance in all women at high risk of breast cancer. It is thought that the decrease in sensitivity in mammography found in some studies of the surveillance of women at high risk of breast cancer with mammography is due to these women being younger and therefore having denser breast tissue. This would not have shown up in the age range in this study. It was also stated that the accuracy of mammography did not alter with age; however the age groups examined were all within the 50-64 age group. Authors' conclusions: The authors' conclusions regarding high-risk women were that sensitivity and specificity did not vary significantly according to family history.

Table 11. Primary research studies appraised investigating the accuracy and efficacy of mammographic surveillance compared to usual care on outcomes from breast cancer (continued)

Source	Study design Evidence grading	Comparison interventions and dates of testing	Sample	Outcomes and Verification	Results	Comments
Banks et al. (2004) UK multi-centre at 10 breast screening units Continued						Reviewers' conclusions: This study is interesting but the conclusions are drawn on a subset of women at high risk. The issue for sensitivity of mammography in women at high risk is more related to the fact that they require surveillance at a much younger age than the group examined in this study. There is also quite a lack of detail regarding the women at high risk in this study or the surveillance that was received, as this group was only a small focus of this study. Studies that specifically look at high-risk women as their main focus are more likely to be reliable for this data. There was also a discrepancy between the 3-yearly surveillance interval and the 12-month follow-up, which meant that interval tumours would be underestimated.

Table 11. Primary research studies appraised investigating the accuracy and efficacy of mammographic surveillance compared to usual care on outcomes from breast cancer (continued)

Source	Study design Evidence grading	Comparison interventions and dates of testing	Sample	Outcomes and Verification	Results	Comments
Halapy et al. (2004) Ontario, Canada	Retrospective cohort Study III-2 (C1 P1 Q2)	The Ontario Breast Cancer Screening Protocol was biennial screening with 2-view mammography and CBE. Women considered to be at high risk (e.g. family history of breast or ovarian cancer) are recalled in 1 year. Comparisons are drawn between 3 risk groups: those with no family history, those with moderate family history and those with a strong family history. Comparisons are also made between age groups and between different modalities of screening. Dates of screening were Jan 1996 to Dec 1997.	Sample no = 143,574 women altogether with the majority (121,825) in the no family history group. Mean age (at diagnosis or entry to programme) is not given. Recruited from the Ontario Breast Screening programme (OBSP). Inclusion criteria: • age of 50 years or older; • had an OBSP screen (first or rescreen) between Jan1996 and Dec1997; • resident in Ontario; • no history of breast cancer or augmentation mammoplasty; • free from acute breast symptoms. Exclusions: • women who only had CBE (*n*=77);	Relevant outcomes • cancer detection rate; • mode of detection; • tumour size; • tumour stage; • node status; • PPV. Verification of positive screens was obtained from the OBSP recall process and through linkage with the Ontario Cancer Registry. All cancers are confirmed by a pathology report. Negative screens were verified by follow-up – although was only 1 year in this study.	Cancer detection: The rates were stratified by risk group and age and can be seen in more detail in the original paper. Cancer detection in risk groups: Strong family history, 65 tumours; Moderate family history,133 tumours; No family history, 788 tumours. In initial (prevalent) screens the cancer detection rate was 8.3 per 1,000 (95% CI 7.5-9.1) for those with no family history, 9.7 (95% CI, 7-12.4) for those with a moderate family history and 10.6 (95% CI, 6.7-14.5) for those with a strong family history. For the rescreens (incident screens) the rate was 5.3 per 1,000 (4.8-5.8), 9.0 (7.1-10.9) and 7.7 (5.2-10.2) respectively for those with no history, a moderate history and a strong history of breast cancer. ($p<0.05$ for strong versus no family history and $p<0.001$ for a moderate family history versus no family history). The cancer detection rates increased with increasing age and are consistently higher in the 70-plus age group for all degrees of family history and for both the prevalent and incident screens.	Limitations include: Retrospective cohort. Conducted in a screening, not surveillance population. No blinding to risk status and it is discussed that this may have influenced nurse practitioner referral patterns. Only looks at women over 50 years age; will have different characteristics to younger women, e.g. breast density, tumour characteristics. Family history: self reported and not verified, could lead to misclassification. No ability to assess interval tumours in this study as only looked at one year; this has been assessed in a later paper (2005). Few characteristics of sample, although did collect data on HRT use but not on reproductive histories or prior risk reduction strategies (BSO and Tamoxifen). Did ask about previous screens, where data was missing were excluded from calculations that involved these aspects. Verification bias is likely. In the comparisons made it is a little unclear on the screening received i.e. whether those in the strong family history category would have been considered at high risk and subsequently screened more frequently. This would not alter the comparisons of prevalent screens, but may affect comparisons in the incident screen results if this group were having more frequent screening than the other groups.

Table 11. Primary research studies appraised investigating the accuracy and efficacy of mammographic surveillance compared to usual care on outcomes from breast cancer (continued)

Source	Study design Evidence grading	Comparison interventions and dates of testing	Sample	Outcomes and Verification	Results	Comments
Halapy et al. (2004) Ontario, Canada *Continued*			• If women had more than one screen done in that period, the second screen was excluded (3,924 screens). Only the first one included. Risk stratified by self reported information on the type of cancer, which relative and the age of their diagnosis. 'Strong' family history was defined as 2 or more first-degree relatives diagnosed with breast or ovarian cancer at any age, 1 first-degree relative diagnosed with breast cancer <50 years or 1 first-degree relative diagnosed with breast and ovarian cancer at any age (this is similar to OBSP criteria for annual screening of women at high risk).		Note: Women currently using HRT had no significant association between family history and cancer detection rate but those not currently using it did have a significant association. Mode of detection: The cancer detection rate is presented by screening modality and shows that mammography consistently detects more tumours than CBE. It is unclear how many tumours are detected by each screening modality alone and how many are detected by both. The cancer detection rate for CBE ranges from 4 to 6.1 per 1,000 for the prevalent screens and 1.9 to 3.1 per 1,000 for the incident screens (depending on risk group). The cancer detection rate for mammography ranges from 7.9-9.8 per 1,000 for the prevalent screen and 5-7.3 for the incident screens (dependent on age group). Unfortunately, the results are all stratified by risk group and prevalent or incident screen. The figures are all presented as rates and there are no raw figures so it is not possible to calculate overview figures or to check these calculations.	Authors' conclusions: By determining screening outcomes for women with a family history, a preliminary indication is given of whether screening is as effective in women with a family history as in those without is achieved. Because much attention is focused on younger women (less than 50 years of age) with a family history, it is important to point out that we see evidence of continuing increased risk with age in these women. Greater cancer detection rates with high proportions of tumours with a good prognosis in women with a family history indicate that these women may have the potential to benefit from regular breast cancer screening. However, to completely evaluate the performance of breast cancer screening using interim indicators of effectiveness and to identify appropriate management guidelines, the other outcomes such as sensitivity are required. Comparisons of such outcomes will be reported in a following paper (2005). Reviewers' conclusions: This study demonstrates that women with a family history of breast cancer have higher cancer detection rates at screening, which reflects their increased risk of breast cancer. It also emphasises that increasing age remains an important risk factor also within each family history risk category.

Table 11. Primary research studies appraised investigating the accuracy and efficacy of mammographic surveillance compared to usual care on outcomes from breast cancer (continued)

Source	Study design Evidence grading	Comparison interventions and dates of testing	Sample	Outcomes and Verification	Results	Comments
Halapy et al. (2004) Ontario, Canada *Continued*			'Moderate' family history was defined as 1 first-degree relative diagnosed with breast cancer at 50 years or older or 1 first-degree relative with ovarian cancer but not breast cancer at any age. 'No' family history was defined as no first-degree female relatives diagnosed with breast and/or ovarian cancer.		Tumour size, stage and node status: In those with no family history, 123 (15.6) were *in situ* tumours and 665 (84.4) were invasive. The proportion that were invasive but <20mm in size was 83.5%. Of the invasive tumours 73.7 (70-77.4) were node-negative. In those with a moderate family history, 6 (4.5%) were *in situ* and 127 (95.5%) were invasive. The proportion that were invasive but <20mm in size was 86.6%. Of the invasive tumours 80.6% (72.8-88.4) were node-negative. In those with a strong family history, 15 (23%) tumours were *in situ* and 50 (77%) were invasive. The proportion that were invasive but <20mm in size was 88%. Of the invasive tumours, 67.7% (51.9-83.4) were node-negative. The only statistically significant difference in tumour characteristics is that women with a moderate family history had fewer *in situ* tumours than women with no family history. PPV (95% CI). Stratified by degree of family history and prevalent or incident screen:	The study is interesting as it has looked at the potential confounders of previous mammography outside the breast screening programme and also HRT use. The results also suggest that in women at increased risk due to a family history there is a similarly high proportion of *in situ* and node-negative tumours detected by screening as there is in those without a family history. Thereby suggesting an equal efficacy of the programme in women at high risk as in those who are at average risk. In this study it was interesting that those at moderate risk had a significantly smaller proportion of *in situ* tumours detected compared to the other two groups. This was attributed to small numbers of statistical chance. However, perhaps these women did not qualify for annual screens like the strong history group, but may warrant more regular screening. It is not possible to know from these results. The results of this study, however, do not help with the question of when surveillance for those with a family history of breast cancer should begin and how effective this surveillance would be in a younger population. The results are not comparable to a younger population due to the difference in breast density and tumour development and characteristics that affect screening efficacy.

Table 11. Primary research studies appraised investigating the accuracy and efficacy of mammographic surveillance compared to usual care on outcomes from breast cancer (continued)

Source	Study design Evidence grading	Comparison interventions and dates of testing	Sample	Outcomes and Verification	Results	Comments
Halapy et al. (2004) Ontario, Canada *Continued*					Moderate family history Prevalent CBE 6.4% (3.9 to 8.8%) XRM 9.9% (7.2 to 12.6%) Incident CBE 7.4% (4.8 to 9.9%) XRM 15.9% (12.7 to 19.2%) Strong family history Prevalent CBE 7.3% (3.9 to 10.85) XRM 12.0% (7.7 to 16.4%) Incident CBE 6.1% (3.1 to 9.0%) XRM 14.5% (10.1 to 19.0%) The PPV is also seen to increase with age in all family history groups. No *p* values are given and no raw data to calculate them	It is of interest that the subgroup of women with a family history, who had not had previous mammography outside the screening programme, was especially likely to have cancer detected. It is hypothesised that this suggests that they should have undergone surveillance earlier than when they first came to the programme. In relation to the modalities used, it appears that mammography is far more effective at detecting tumours than CBE alone. It is unfortunate that there is not raw data to verify these figures.

Table 11. Primary research studies appraised investigating the accuracy and efficacy of mammographic surveillance compared to usual care on outcomes from breast cancer (continued)

Source	Study design Evidence grading	Comparison interventions and dates of testing	Sample	Outcomes and Verification	Results	Comments
Halapy et al. (2005) Ontario, Canada (Reports further results of their previous study, but reports on different outcomes in this paper.)	Retrospective cohort Study III-2 (C1 P1 Q2)	Ontario Breast Cancer Screening Protocol was biennial screening with 2-view mammography and CBE. Women considered to be at high risk due to a family history of breast or ovarian cancer are recalled in 1 year. Comparisons are drawn between 3 risk groups, those with no family history, those with moderate family history and those with a strong family history, and between different age groups and different screening modalities. (When comparing one modality with another, cancers detected by another modality are considered as interval cancers). Dates of screening were 1996 to 1997.	Sample no = 115,460 women altogether – majority (85.4%) in the no family history group. There were 16,813 women with a moderate or strong family history. Mean age (at diagnosis or entry to programme) is not given. Recruited fro the Ontario Breast Screening programme (OBSP) Inclusion criteria • age 50-69 years; had an OBSP screen (first or rescreen) between Jan1996 and Dec1997; • resident in Ontario; • no history of breast cancer or augmentation mammoplasty; • free from acute breast symptoms. Exclusions: • women who only had CBE (*n*=58);	Relevant outcomes: • cancer detection rate • mode of detection • interval cancers • sensitivity • specificity Verification of positive screens was obtained from the OBSP recall process and through linkage with the Ontario Cancer Registry. All cancers are confirmed by a pathology report. Negative screens verified by follow-up, although was only 12 months post-screening in this study. Interval cancers were defined as any cancer presenting between 2 regular screening examinations. This included cancers with positive screening tests that were not diagnosed at assessment, but	Cancer detection: 604 women were diagnosed with cancer (invasive breast cancer) from screening. There were 110 tumours diagnosed in women with a moderate or strong family history. Cancer detection rate for women with a family history was 6.5 per 1,000 women. Mode of detection: For the 115,460 women who received mammography, 571 tumours were detected = 4.9 per 1,000 women. For the 114,911 women who had CBE, there were 279 tumours detected = 2.4 per 1,000 women. All of the 16,813 women with a family history of breast cancer had mammography and 105 tumours were detected (6.2 per 1,000). Of these women, only 16,721 had CBE performed and 49 tumours were detected (2.9 per 1,000). Interval cancers: Interval cancer rate was defined as the number of women with a diagnosis of an interval cancer per 10,000 person-years at risk within 12 months of the screen date. It only includes invasive tumours.	Limitations include: Retrospective cohort. Screening population, not surveillance. Only looks at women over 50 years age who will have different characteristics to younger women e.g. breast density, tumour characteristics. Family history self-reported and not verified, could lead to misclassification. No blinding to risk status and was commented that this may have affected the nurse practitioner referral patterns. Few characteristics of sample, although did collected data on HRT use but not on reproductive histories. Did ask about previous screens. No information on risk reduction strategies such as BSO or Tamoxifen. Verification bias is likely. In the comparisons made it is a little unclear on the screening received, whether those in the strong family history category would have been considered at high risk and subsequently screened more frequently. This would not alter the comparisons of prevalent screens, but may affect comparisons in the incident screen results if this group were having more frequent screening than the other groups. Authors' conclusions: By determining screening outcomes for women with a family history, a preliminary indication is given as to whether screening is as effective in these women as in women without a family history.

Table 11. Primary research studies appraised investigating the accuracy and efficacy of mammographic surveillance compared to usual care on outcomes from breast cancer (continued)

Source	Study design Evidence grading	Comparison interventions and dates of testing	Sample	Outcomes and Verification	Results	Comments
Halapy et al. (2005) Ontario, Canada (Reports further results of their previous study, but reports on different outcomes in this paper.) *Continued*			• if women had more than one screen done in that period, the second screen was excluded (3,536 screens); only the first one included. Risk stratified by self-reported information on the type of cancer, which relative and the age of their diagnosis. 'strong' family history was defined as 2 or more first-degree relatives diagnosed with breast or ovarian cancer at any age, 1 first-degree relative diagnosed with breast cancer <50 years or 1 first-degree relative diagnosed with breast and ovarian cancer at any age. (This is similar to OBSP criteria for annual screening of women at high risk.}	were subsequently diagnosed within 12-month follow-up (missed in assessment interval cancers).	61 women were diagnosed with an invasive interval cancer within 12 months of their screening examination. 48 (78.7%) were in the group with no family history, 7 (11.5%) were in the moderate risk group and 6 (9.8%) were in the strong family history group. Comparisons: Interval cancer rates increased across family history groups and were greatest in women with a strong family history. However, the difference among the groups was not significant. The interval cancer rates (and 95%CIs) were 4.9 (3.5-6.3), 6.4 (1.7-11.1) and 10.5 (2.1-18.8) per 10,000 person years in the no family history, moderate history and strong history groups respectively. The interval cancer rates were higher for CBE alone than for mammography alone across all family history groups, e.g.: Moderate family history: Mammography interval tumour rate of 8.2 per 10, 000 person/years (2.9-13.6) CBE interval tumour rate of 46.8 per 10,000 person/years.	Our study found that screening with both modalities or with mammography only was able to identify a large proportion of invasive breast cancers in women with a moderate or strong family history, indicating their potential to benefit from regular breast cancer screening. However, because interval cancer rates in women with a family history, especially those with a strong family history, were already quite high after 12 months of follow-up, screening with one-year intervals may be important, even in an older population of women with a family history. Reviewers' conclusions: This study must be interpreted with the above limitations. It is not generalisable to the younger population of women at high risk of breast cancer. It also does not include the detection of *in situ* cancer, which would alter the results. It does suggest that mammography, or mammography plus CBE is more effective for detecting cancer in women at high risk then CBE alone. There are still relatively high interval cancer rates using both CBE and mammography and this suggests that a more intensive surveillance protocol may be required for these women (at a high risk from family history) or the addition of other modalities of surveillance

Table 11. Primary research studies appraised investigating the accuracy and efficacy of mammographic surveillance compared to usual care on outcomes from breast cancer (continued)

Source	Study design Evidence grading	Comparison interventions and dates of testing	Sample	Outcomes and Verification	Results	Comments
Halapy et al. (2005) Ontario, Canada (Reports further results of their previous study, but reports on different outcomes in this paper.) *Continued*			'Moderate' family history was defined as 1 first-degree relative diagnosed with breast cancer at 50 years or older or 1 first-degree relative with ovarian cancer but not breast cancer at any age. 'No' family history was defined as no first-degree female relatives diagnosed with breast and/or ovarian cancer.		This is due to the method of analysis where tumours not detected by that modality were considered as interval tumours and just reflects the lower detection rate of CBE alone. There were no statistically significant results found from the multivariate analysis for age, HRT and a prior screen outside of the breast screening programme. Sensitivity of screening with CBE and mammography: (95% CI) Moderate Family History 50-59yrs 89.5% (79.7 to 99.2%) 60-69yrs 93.6% (86.6 to 100%) Total study group 91.8% (85.9 to 97.6%) Strong Family History 50-59yrs 86.7% (69.5 to 100%) 60-69yrs 82.6% (67.1 to 98.1%) Total study group 84.2% (72.6 to 95.8%) Specificity of screening with CBE and mammography: (95% CI) Moderate family history: 50-59 years 88.55 (87.7 to89.4%) 60-69 years 90.8% (90.0 to 91.5%) Total study group: 89.6% (89.0 to 90.2%)	

Table 11. Primary research studies appraised investigating the accuracy and efficacy of mammography screening compared to usual care on outcomes from breast cancer (continued)

Source	Study design Evidence grading	Comparison interventions and dates of testing	Sample	Outcomes and Verification	Results	Comments
Halapy et al. (2005) Ontario, Canada (Reports further results of their previous study, but reports on different outcomes in this paper.) *Continued*					Strong family history: 50-59 years 88.2% (87.1 to 89.4%) 60-69 years 91.0% (89.9 to 92.0%) Total study group: 89.5% (88.8 to 90.3%) Specificity of screening with CBE and mammography: In general, the specificity did not differ according to family history. The sensitivity and specificity of mammography and CBE were also examined individually. The sensitivity was higher for mammography than CBE– ranging from 76.3 to 89.4% across different family history groups for mammography and 40-42.8 across different family history groups for CBE. However, the specificity was found to be comparable between the 2 screening modalities, across all family history groups Sensitivity by modality (95% CI): Moderate family history: CBE 40.0 (29.6 to 50.4%) XRM 89.4 (82.9 to 96.0%)	

Table 11. Primary research studies appraised investigating the accuracy and efficacy of mammographic surveillance compared to usual care on outcomes from breast cancer (continued)

Source	Study design Evidence grading	Comparison interventions and dates of testing	Sample	Outcomes and Verification	Results	Comments
Halapy et al. (2005) Ontario, Canada (Reports further results of their previous study, but reports on different outcomes in this paper.) *Continued*					Strong family history: CBE 40.5% (24.7 to 56.4%) XRM 76.3% (62.8 to 89.8%) Specificity by modality (95% CI): Moderate family history: CBE 94.8% (94.4 to 95.2%) XRM 93.9% (93.4 to 94.3%) Strong family history: CBE 94.0% (93.4 to 94.6%) XRM 94.6% (94.0 to 95.2%) No statistically significant results were found after multivariate analysis for age/HRT use or having a prior screen outside of the breast screening programme. The above figures are all based on invasive tumours only, and do not include DCIS.	

Table 11. Primary research studies appraised investigating the accuracy and efficacy of mammographic surveillance compared to usual care on outcomes from breast cancer (continued)

Source	Study design Evidence grading	Comparison interventions and dates of testing	Sample	Outcomes and Verification	Results	Comments
Kuhl et al. (2005b) Germany	Prospective cohort study III-2 (C1 P2 Q2)	Surveillance protocol: Biannual CBE and US and annual XRM and MRI. If abnormalities found on CBE or US at round without XRM or MRI, these additional modalities were used to further investigate this. Surveillance commenced at 30 years or 5 years before the youngest family member affected with the disease. (NB in first 2 years, women under 30, or 30-39 years with dense breasts did not receive XRM, but this was subsequently abandoned and all women received XRM,) Mammography (XRM): Annual conventional film screen XRM performed with at least 2 views per breast (medio-lateral oblique and caudal-cranial), obtained and interpreted in accordance with German radiological practice guidelines. Diagnoses coded according to the BI-RADStm diagnostic categories on a 5-point	Sample no = 529 (out of 590 eligible women, 49 were lost to follow-up after 1 surveillance round and 12 were also excluded as they had a clinical abnormality at initial examination). Inclusion criteria: • asymptomatic women; • personal history of breast cancer included provided that the patient had not undergone bilateral mastectomy, had not received chemotherapy within the previous 12 months and had no metastases. (139 women were included with a personal history of breast cancer.) Exclusion criteria: • clinical signs of breast cancer; • chemotherapy within the previous 12 months;	Relevant outcomes: • cancer detection; • mode of detection; • tumour size; • tumour stage; • node status; • interval tumours; • sensitivity; • specificity; • PPV; • NPV. Verification of a positive result was achieved by histology (for positive imaging studies). Verification of a negative result was achieved by follow-up (for negative imaging studies). If a breast cancer was identified clinically (by palpation) between surveillance rounds or at the 6-month clinical visit, the imaging studies of the previous round were considered false negative. Mean follow-up was 5.3 years (range 2-7 years).	Cancer detection: A total of 43 breast cancers were identified in 41 patients. Eleven of these women had a prior history of breast cancer; 40 of these were said to be detectable by imaging. Mode of detection: CBE identified only one tumour (also detected on imaging) XRM identified 14 tumours (only 1 was diagnosed by XRM that wasn't diagnosed by MRI). Tumour size, stage and node status: Of the 21 cancers detected by XRM and US, 16 were invasive and the rest were DCIS. The invasive cancers had a mean size of 13.9mm and 5 were node-positive. Unfortunately, characteristics not presented separately for these modalities. Interval tumours: The paper states that 40 out of 43 tumours in this cohort were detected by imaging. However, a sentence in the discussion states that the rate of interval cancers was 2% in this cohort.	Limitations include: CBE and the imaging studies were performed within a time frame of 8 weeks. Few sample characteristics presented, such as OCP or HRT use, or the use of preventative strategies such as tamoxifen or BSO. Verification bias is likely. Reporting of interval tumours is unclear. lack of blinding to the results of the CBE. Author's conclusions: The authors' conclusions relate to the overall surveillance strategy (including US and MRI). If US is used in combination with XRM, it can help compensate for some but by far not for all of the shortcomings of XRM, and it causes a substantial number of false-positive diagnoses. If MRI is used for surveillance, XRM proved to be of limited and ultrasound of no additional value. US may however be useful to bridge the relatively long time interval between annual surveillance rounds. Propose that in view of the insufficient diagnostic accuracy of XRM and USS, breast MRI should be considered an integral part of surveillance programmes for women at high familial risk, in particular in documented carriers of pathogenic BRCA mutations.

Table 11. Primary research studies appraised investigating the accuracy and efficacy of mammographic surveillance compared to usual care on outcomes from breast cancer (continued)

Source	Study design Evidence grading	Comparison interventions and dates of testing	Sample	Outcomes and Verification	Results	Comments
Kuhl et al. (2005b) Germany *Continued*		scale (1, negative; 2, benign; 3, probably benign; 4, suspicious abnormality; 5, highly suggestive of malignancy). Breast MRI: Standard dynamic axial contrast-enhanced breast MRI of both entire breasts was performed on a 1.5T system (NT/INTERA; Philips, Best, the Netherlands) after injection of 0.1mmol/kg body weight gadopentetate dimeglumine (Magnevist, Schering, Berlin, Germany) Ultrasound (US): performed with 7.5-to 13-MHz probes (Siemens Elegra, GE logic 500 and ATL HDI 5000; Siemens, Erlangen, Germany). The entire breast was systematically examined by the physician who interpreted the study. Diagnoses were scored on a5-point scale identical to the XRM BI-RADStm categories	• women having undergone bilateral mastectomy. Recruited from high risk clinics in a single Gynaecology Department Risk Stratification According to definition of the Consortium on Familial Breast and Ovarian Cancer of the German Cancer Aid corresponding to a lifetime risk of breast cancer of at least 20%. (two or more cases of breast cancer on the same side of the family, including at least two cases with onset before age 50 years, or with breast or ovarian cancer, irrespective of age, families with at least one case of breast cancer diagnosed before 35 years, families with three or more cases of breast cancer on the same side of the family, and women who met the criteria for high familial risk,	A total of 1,542 annual surveillance rounds were completed. Verification of last surveillance round was by continued surveillance in 428 women, telephone interview in 52 women and for 6 women who had prophylactic mastectomy it was by pathology of the specimen. XRM: BIRADS of 4 or 5, biopsy was recommended irrespective of finding in US or MRI. BIRADS 3 was managed by 6-months follow-up until receiving a BIRADS 2 or biopsy clarification. US categorised as BIRADS 3 managed by short-term (6 months) US follow-up. BIRADS 4 or 5 managed by US-guided biopsy (14G, semi-automatic or automatic biopsy gun) except for the following constellation: if an US finding that was suspicious was clearly benign on XRM or MRI no biopsy was performed.	This translates to 10 tumours if the 2% is of the total study population or 1 tumour if it is 2% of the total tumours detected. The latter is most likely as it would alter the figures least but it is unclear. Comparisons: When stratified by risk groups, the detection rates at both the prevalent and incident rounds were much higher in the mutation carriers than the other 2 risk groups, but these differences are not statistically significant. Sensitivity: CBE 2.3% (0.1 to 12%) n = 1/43 XRM 32.6% (19 to 48.5%) n = 14/43 When stratified by risk groups, XRM becomes less sensitive as the lifetime risk of breast cancer increases, with a sensitivity of 25%, for the mutation carrier group. Sensitivity of XRM by risk group: Risk 20% 50% Risk 21-40% 25% Mutation carriers 25%	Reviewers' conclusions: This study shows a higher sensitivity and cancer detection rate for XRM versus CBE in the surveillance of this high-risk population. Unfortunately, the other measures of accuracy can only be calculated for XRM. The limitations of this study must be taken into account in the interpretation.

Table 11. Primary research studies appraised investigating the accuracy and efficacy of mammographic surveillance compared to usual care on outcomes from breast cancer (continued)

Source	Study design Evidence grading	Comparison interventions and dates of testing	Sample	Outcomes and Verification	Results	Comments
Kuhl et al. (2005b) Germany *Continued*		Each imaging study was read and scored independently by a different radiologist who had substantial experience with the respective imaging technique. The readers were informed about the clinical findings from CBE and the risk status of the patient but were blinded to the results of the respective other imaging modalities. Comparisons are made between the 3 risk groups and the different modalities of screening. Dates of study were February 1996 to February 2002.	irrespective of the result of mutational analysis) In women without a personal history of breast cancer the Claus tables were also used to quantify risk. Women were then stratified into 3 risk groups for analysis: • mutation carriers; • high lifetime risk (20-40%); • moderate lifetime risk (20%).	MRI: Suspicious scores (4 or 5) were managed by magnetic resonance-guided biopsy. For findings categorised as BIRADS 3 short-term follow-up after 6 months was recommended with further management corresponding to that of XRM BIRADS 3 lesions BIRADS 3 categories in all imaging that received short-term follow-up were not considered positive for the calculation of outcomes. Invasive cancer and DCIS were considered a malignant diagnosis but LCIS and atypical ductal hyperplasia were considered to be benign.	Specificity (95% CI) XRM 96.8% (95.7 to 97.7%) *n* = 1364/1409 Stratification by risk group does not appear to affect the specificity. PPV XRM 23.7% (1 to 29%) *n* = 14/59 The PPV increases with the increasing risk of breast cancer, this will be affected by the higher incidence in women at higher risk. NPV XRM 97.9% (97.0 to 98.6%) *n* = 1,364/1,393. Unfortunately, the specificity, PPV and NPV cannot be calculated for CBE as there is inadequate data.	

Table 11. Primary research studies appraised investigating the accuracy and efficacy of mammographic surveillance compared to usual care on outcomes from breast cancer (continued)

Source	Study design Evidence grading	Comparison interventions and dates of testing	Sample	Outcomes and Verification	Results	Comments
Elmore et al. (2005) USA, multi-centre across 5 states. (Washington, Oregon, California, Massachusetts and Minnesota)	Matched case-control study III-2 (CX P2 Q2)	Surveillance protocol: Mammography and/or CBE; unclear what surveillance intervals were being used. Comparison made with matched control subjects, between levels of risk and between the 2 age groups. Comparisons were also made of surveillance with either or both modalities and then with CBE alone and with mammography alone. Inclusion criteria for controls: • free of breast cancer prior to index date; • within 1 year of case subject in age and in same age group (40-49 versus 50-65); • alive on the date that the matched case subject had died; • were enrolled continuously in a health plan (no gaps longer than 3 months) during the index period and were active health plan members at the time of the matched case subject's diagnosis.	Sample no = 3,752 women. Cases = 1,351 women who had died of breast cancer (random sample of those eligible). Controls = 2,501 women that were matched to the cases for age and risk level (201 cases that only 1 control could be found to match). Mean age not given; 2 age categories, 40-49 years and 50-65 years. Cases and controls had very similar demographic and clinical characteristics e.g. ethnicity, reproductive features, co-morbidities. Recruited from 6 health plans in 5 states in USA. Inclusion criteria for cases: • breast cancer diagnosis between 1983 to 1993 (includes DCIS and invasive carcinoma as defined by ICD-9);	Relevant outcomes: • Mortality/ occurrence of surveillance among those who had died of breast cancer. Verification: Surveillance history was abstracted from the medical records (for 2 years and also for 3 years prior to the index date) as were details of risk category. The data on breast cancer characteristics and treatment were collected from NCI Surveillance, Epidemiology and End Results (SEER) cancer registries or from tumour registries at 4 sites. Double review done by blinded reviewers of randomly selected medical records to check validity of abstracted data.	In women at increased risk of breast cancer in the 40-49 year age group, 169 out of 216 cases (78.2%) had received surveillance. This compared with 303 out of 376 (80.6%) of controls. In the 50-65 year group of women at increased risk of breast cancer, 158 out of 195 cases (81%) had received surveillance. This compared with 296 out of 347 controls (85.3%). Odds ratios for fatal breast cancer in relation to surveillance are presented for surveillance by either CBE or mammogram, by CBE alone and by mammography. This is further stratified into the 2 age groups and 2 risk groups. In general, surveillance gives an odds ratio (OR) <1, however all but one OR have 95% confidence intervals that include the value of 1 and are therefore non significant. The one that does not is for surveillance by CBE in women at an increased risk aged 50-65 years and is 0.61 (95% CI, 0.39-0.97).	Limitations include: Case-control design subject to bias and confounding. The high-risk group was only a small subgroup in this study. Risk stratification was basic and included a past history of breast biopsy as a risk factor. Documentation of reproductive history and history of previous breast biopsies was made between cases and controls (as potential confounders) and there were no significant differences. However, no documentation of exogenous hormonal use (HRT/OCP), other potential risk factors or risk reduction strategies (BSO or tamoxifen). Comparisons were not made of demographics/characteristics of those that were selected as cases or controls and those that were not. However, as the sampling was random one would expect that they were representative samples of the whole population. Possibility of differential and non-differential misclassification. Authors' conclusions: We observed no appreciable association between breast cancer mortality and surveillance history, regardless of whether surveillance took place during a woman's 40s, 50s or 60s.

Table 11. Primary research studies appraised investigating the accuracy and efficacy of mammographic surveillance compared to usual care on outcomes from breast cancer (continued)

Source	Study design Evidence grading	Comparison interventions and dates of testing	Sample	Outcomes and Verification	Results	Comments
Elmore et al. (2005) USA, multi-centre across 5 states (Washington, Oregon, California, Massachusetts and Minnesota) *Continued*		Exclusions for controls • risk level not the same as the matched case once medical records reviewed; • complete medical chart not available; • abstraction not necessary because 2 controls already found for the case. Attempted to find 2 matched controls for every case subject. Dates 1983 to 1998.	• between 40 and 65 years; • died from breast cancer or causes possibly relating to breast cancer from 1983 to 1998; • were enrolled continuously in a health plan (no gaps longer than 3 months) during the index period and were active health plan members at the time of diagnosis. Exclusions for cases: • medical chart information was not available; • medical chart was not reviewed because the funded study period had ended or for other reasons; • no eligible control subjects found	Clinicians, blinded to case-control status, reviewed cases where it was unclear if was surveillance or diagnostic mammogram. Multiple sensitivity analyses were performed to see if they would alter the results obtained, but no significant differences occurred. Family history data was abstracted from notes for 10 years prior to, but not including, the index date. This was to avoid better family history data being collected in cases versus controls due to diagnosis/suspicion of breast cancer.		Our results suggest that surveillance might be efficacious only among women who are at increased risk for breast cancer, although the differences in the estimated efficacy of surveillance according to women's risk levels were well within the limits of chance. The non-randomised study design and potential limitations of the available data argue for a cautious interpretation of these results. Nevertheless, it is possible that the efficacy of surveillance for breast cancer may be lower in community settings than in randomised control trials or as the treatment of breast cancer Improves. Reviewers' conclusions: As suggested by the authors, this study must be interpreted with great caution due to its design. For the purpose of this review, its utility is also lessened by the method of risk stratification, which was very simplistic and included suspicion of breast disease as well as family history factors, and by the lack of information available about the surveillance that these women actually underwent.

Table 11. Primary research studies appraised investigating the accuracy and efficacy of mammographic surveillance compared to usual care on outcomes from breast cancer (continued)

Source	Study design Evidence grading	Comparison interventions and dates of testing	Sample	Outcomes and Verification	Results	Comments
Elmore et al. (2005) USA, multi-centre across 5 states (Washington, Oregon, California, Massachusetts and Minnesota) *Continued*			Increased risk was defined as a family history of breast cancer or a breast biopsy noted in the medical records before the index date (defined as date of first suspicion of breast abnormalities in the cases and the same date was used for matched control subjects). The index period was the 3 year period before and including the index date. On this basis women were classified at being at increased risk or average risk of breast cancer.			

Table 11. Primary research studies appraised investigating the accuracy and efficacy of mammographic surveillance compared to usual care on outcomes from breast cancer (continued)

Source	Study design Evidence grading	Comparison interventions and dates of testing	Sample	Outcomes and Verification	Results	Comments
Gui et al. (2006) London	Prospective cohort Study III-2 (C1 P1 Q2)	Surveillance protocol (for women at moderate and high risk): Annual CBE and mammography from the age of 35 years. Ultrasound examination was sometimes used in addition if there was uncertainty. After the age of 50 years, women at a moderate risk underwent surveillance every 18 months while women at high risk continued to receive surveillance annually until the age of 69 years. The standard risk women were discharged to the NHSBSP unless there was any clinical indication for continued follow-up. Dates of study were patients already attending the breast diagnostic unit over 4 months from Jan to April 2003, and patients were followed until Dec 2004. Comparisons were drawn between the different risk categories and between the cancer detection rate in the study and that in older women in the UK NHSBSP.	Sample no = 1,132 women attended the incident round (1,170 visits (includes recall visits and results visits, so not all surveillance visits). Women were risk stratified in to standard risk (lifetime risk less than 17%), moderate risk (lifetime risk 17-30%) and high risk (lifetime risk over 30%). There were: 137 high-risk women; 803 moderate risk women; 192 standard risk women. Only 406 were completely asymptomatic and had no clinical signs. A lot of the others had nodularity or breast pain. Median age at diagnosis was 52 (range 45-69). This was 63 years in the standard risk group, 54 years in the moderate risk group and 51years in the high risk group.	Relevant outcomes: • cancer detection rate; • mode of detection; • tumour size; • tumour stage; • node status; • interval tumours; • mortality; • sensitivity; • specificity; • PPV; • NPV; Verification of positive results was with pathology and verification of negative results was with follow-up. Verification of interval tumours is not discussed. Follow-up for patients was at least one routine 12 or 18 monthly surveillance interval in the high and moderate risk groups respectively.	Cancer detection rate 15 cancers were detected on incident rounds, 7 during the first round (active study period) and 8 during the second round. The cancer detection rate is just given for the active study period and is 6.2 breast cancers per 1,000 women under surveillance. Of the 7 tumours, 2 were in the standard risk group, 3 were in the moderate risk group and 2 were in the high risk group. The respective cancer detection rates are therefore 10.4, 3.7 and 14.6 per 1,000 women under surveillance. Mode of detection: 4 tumours had clinical findings on CBE and all were identifiable on mammography. Tumour size, stage and node status: Unfortunately, these results are not presented for surveillance-detected tumours and interval tumours separately. For all tumours that occurred, 13 (76.5%) were invasive, 4 (23.5) were *in situ*. The median invasive tumour size was 15mm (range 7-28mm) and the median DCIS size was 4mm (range 2-30mm).	Limitations included: No mention of blinding to risk status of women. Not all women received mammography (some declined but their characteristics are not described). There are few characteristics of the women in these cohorts other than risk level and age. There is no mention of other risk factors such as HRT or OCP use and there are no characteristics given of women who did not participate in the study. There was also no mention of past history of breast cancer and the use of risk reduction strategies such as BSO or Tamoxifen use. Verification bias is likely. The role of ultrasound in this study is unclear; it is stated that 594 US scans were performed to evaluate clinical symptoms further, but it appears that no cancers were picked up on ultrasound that were not diagnosed by CBE or mammography. It is unclear how the results of the US examinations correlated with the results of the mammography and CBE and whether the use of US may have confounded these results. Authors' conclusions: Surveillance for women at increased breast cancer risk is effective. Early detection and recall rates are comparable to those of older women attending the NHSBSP.

Table 11. Primary research studies appraised investigating the accuracy and efficacy of mammographic surveillance compared to usual care on outcomes from breast cancer (continued)

Source	Study design Evidence grading	Comparison interventions and dates of testing	Sample	Outcomes and Verification	Results	Comments
Gui et al. (2006) London *Continued*		Comparisons are also made between CBE and XRM surveillance.	The mean age of diagnosis was 58.8 years in the standard risk group, 55.6 years in the moderate risk group and 49.7 years in the high risk group. Mean age of entry to the study was 54 years for the standard risk group, 49 years for the moderate risk group and 47 years for the high risk group. Recruited consecutive women attending the breast diagnostic unit for their routine incident surveillance screen. Risk stratified by criteria set out in the article into standard, moderate and high risk. The criteria for classification were based on guidance provided by the National Institute for Clinical Excellence (NICE) and purely related to family history criteria.		Two of the 13 patients with invasive cancer were node positive (15.4%). Interval tumours: There were 2 interval tumours, One was in the standard risk group and 1 in the moderate risk group. Mortality: There were 4 deaths overall in the cohort but none were related to breast cancer. Comparisons: The cancer detection rate overall was reported as 6.2 per 1,000 women under surveillance and therefore was more than would be expected in the UK NHSBSP (3.8 per 1,000 women screened). The rate in this study included some women at standard risk and this would have reduced/ underestimated the cancer detection rate. The following results are all for the first surveillance screen only; this was still an incident round as the participants were all already receiving surveillance at the facility.	The study data also supports the fact that women in the moderate risk category should perhaps continue intensive surveillance after the age of 50 years and not returned to the NHSBSP as is current practice. Reviewers' conclusions: This study does suggest that surveillance of women at high risk of breast cancer is effective. The study design is confused by the use of US scanning, which is not thoroughly discussed and may have confounded the results. There is comment made about how the cancer detection rate is still high among women in the moderate risk category, with the mean age of diagnosis being 55.6 years, and that this suggests that they should continue intensive surveillance into their 50s. However, little comment is made of the fact that the cancer detection rate in the standard risk women is even greater, with the mean age of diagnosis being 58.8 years, and there is still one interval cancer in this group. However, this may be a product of selection bias in this group. It is stated in the methodology that the standard risk women in this study were those that had not yet been discharged back to the NHSBSP and this was usually because of a clinical indication for continued follow-up. Therefore, it can probably be assumed that these women were not representative of women at standard risk in general.

Table 11. Primary research studies appraised investigating the accuracy and efficacy of mammographic surveillance compared to usual care on outcomes from breast cancer (continued)

Source	Study design Evidence grading	Comparison interventions and dates of testing	Sample	Outcomes and Verification	Results	Comments
Gui et al. (2006) London *Continued*					The sensitivity of mammography was 85.7% (6/6+1). The specificity of mammography was 98.8% (816/816+10). The positive predictive value of mammography was 37.5% (6/6+10). The negative predictive value of mammography was 99.9% (816/816+1).	

Table 11. Primary research studies appraised investigating the accuracy and efficacy of mammographic surveillance compared to usual care on outcomes from breast cancer (continued)

Source	Study design Evidence grading	Comparison interventions and dates of testing	Sample	Outcomes and Verification	Results	Comments
Maurice et al. (2006) Manchester, UK.	Prospective cohort study III-2 (CX P1 Q2)	Surveillance protocol: Mammography and CBE at 12-18 monthly intervals, (2-view from 1999 onwards and at baseline and 1-view for follow-up images prior to 1999). Commencing at presentation to the clinic, but not before the age of 35 years unless there were relatives in the family who were affected at an early age. Then surveillance commenced 5 years before the earliest age breast cancer was diagnosed, but not before 30 years. All women were offered instruction on BSE. Comparison: Data was collected from 1991-2002 from women aged less than 50 years who presented symptomatically with breast cancer at the same breast unit. They were excluded if they had had any form of surveillance. Also compared detection rates to those in women 10-15years older in the NHSBSP. .	Sample no = 3,016 women at high risk of breast cancer including 32 known carriers of BRCA1 or BRCA2 mutations. Mean age at diagnosis was 45 years (family history group). Recruited from a family history clinic in Manchester. Inclusion criteria: • asymptomatic • age <50 years • assessed at 1 in 6, or greater, lifetime risk of breast cancer. Risk stratified by the Claus et al. (2001) tables.	Relevant outcomes: • cancer detection rate; • tumour size; • tumour stage; • node status; • interval cancers; • mortality. Verification of positive results is not explicit but appears to have been pathology. Verification of a negative result is follow- up. 10,826 woman/years of follow-up. Therefore the average follow-up time = 3.6 years Verification of interval tumours is not discussed. Interval cancers were included from the 2 years after a normal surveillance result; however the screen detected cases in those 2 years are not included.	Cancer detection In the Family history group there were 62 breast cancers that occurred – 45 were surveillance detected (17 were in BRCA 1 or 2 mutation carriers) 19 were prevalent round cancers, 26 were incident cancers and 17 were interval cancers. (The annual incidence rate was 3.97 per 1,000 woman years). Tumour size, stage and node status Nine of the prevalent tumours were invasive (47%) and 10 were *in situ* (53%). Size of the invasive tumours was <20mm in 78% and 20-50mm in 22%. 8 of the invasive tumours had node sampling and 7 (88%) of these were node negative. Significantly more prevalent carcinomas were *in situ* than incident and interval tumours (p=0.013). Unfortunately the characteristics of the incident and interval tumours are presented conjointly and cannot be separated out. There were 34 invasive tumours (79%) and 9 (21%) *in situ*. This was a significant difference between the prevalent and incident tumours (p=0.013).	Limitations No blinding mentioned to risk status of women No mention of past history of breast cancer or use of risk reducing strategies such as BSO and Tamoxifen. All women, in both groups were followed up by the same group of surgeons and received the same treatment regimes. The only difference was the increase in risk reducing contra lateral mastectomy in the Family history group – this may have biased survival/mortality comparisons between the 2 groups. Selection bias – there was few characteristics given of the 2 cohorts involved. The symptomatic cohort were not risk stratified and so it is unknown how they compare to the family history group and whether they would be women who would qualify for surveillance or not. Their risk status, depending what it is, could bias the results in either direction. One would imagine they might be a very high risk group to have developed tumours at a young age and are perhaps even higher risk than those in the family history group. There is a considerable amount of information missing from the outcomes in the symptomatic group which, if known, could change the outcomes considerably.

Table 11. Primary research studies appraised investigating the accuracy and efficacy of mammographic surveillance compared to usual care on outcomes from breast cancer (continued)

Source	Study design Evidence grading	Comparison interventions and dates of testing	Sample	Outcomes and Verification	Results	Comments
Maurice et al. (2006) Manchester, UK. *Continued*		Dates of study were January 1991 and December 2002 No mention of classification system for images or cut-off		All women were followed up for recurrence, death from breast cancer or other death until the end of 2004.	24 tumours (71%) were <20mm in size, 9 (26%) were 20-50mm and 1 (3%) was >50mm. Thirty 3 had node sampling and of those, 20 (61%) were node negative. Interval cancers: There were 17 interval cancers. Mortality: Two of the women with prevalent tumours, and 2 of the women with incident/interval tumours died. This was 10.5% and 5% respectively. Comparisons: The symptomatic women had known histology for 1,000 tumours, and unknown for 108 tumours (9.7%). Of those with known histology, 918 (92%) invasive tumours and 82 (8%) *in situ*. Tumour size was not known for 213 (20.8%) of the cases. Of those known, it was <20mm for 321(39%), 20-50mm for 414 (51%) and >50mm for 78 (10%). Nodes had been sampled in 939 (90.6%). For those sampled, 441 (47%) were negative and 498 (53%) were positive. In the symptomatic group 216 women (19%) died.	The tumours that have unknown size and histology and have not had there nodes sampled are hypothetically more likely to be smaller and less invasive and if this was known, could reduce the difference seen between the 2 cohorts. Likely verification bias: Lead time is adjusted for in the survival analysis. Authors' conclusions: This study has shown that the tumours detected in women under 50 with a family history of breast cancer in a surveillance programme were significantly smaller, less likely to be node-positive and less likely to be invasive than tumours that present symptomatically in similar aged women not receiving surveillance. This resulted in a survival advantage for women in the surveillance programme over a group of women of the same age who were not exposed to surveillance, even taking into account the lead time. The study has also confirmed that the reported detection rates for women in the FHC setting are comparable to those in population screening programmes for women 10-15 years older.

Table 11. Primary research studies appraised investigating the accuracy and efficacy of mammographic surveillance compared to usual care on outcomes from breast cancer (continued)

Source	Study design Evidence grading	Comparison interventions and dates of testing	Sample	Outcomes and Verification	Results	Comments
Maurice et al. (2006) Manchester, UK. *Continued*					All of these tumours' attributes and outcomes are significantly different between the family history group and the symptomatic group. The family history /surveillance group had less invasive tumours, more *in situ*, more small tumours and more that were node negative. The p values are: Histology p = <0.001 Size p = <0.001 Node status p = 0.013 Mortality p = 0.013 When compared with women 10-15 years older in the NHSBSP, the cancer detection rate was 5.97 per 1,000 at prevalent round and 4.84 at incident round compared with 5.5 per 1,000 and 4.6 per 1,000 respectively in the NHSBSP.	Reviewers' conclusions: The results of this study need to be interpreted with the above limitations in mind. The authors acknowledge that criticism could be made of their choice of control group, however, that this was the most appropriate available as RCTs are near impossible in this setting and no other adequate control group could be identified. It was felt that the biggest concern was that if familial cancer had a better prognosis than sporadic cancer as this would bias the results, however this is very unlikely and in fact the opposite hypothesis is most likely true, especially for BRCA1 carriers. However, the problem remains that it is unclear what risk group the symptomatic women really were, as they appear not to have been assessed for risk. There is also the missing data which, if present, may alter the results considerably. The only result which did not have a great deal of missing data was mortality. Therefore the survival analysis is likely to be a true representation, once adjusted for lead time. This does suggest that the surveillance programme was effective in reducing mortality in women at high risk of breast cancer.

Chapter 4: Accuracy and efficacy of ultrasound

SECONDARY RESEARCH

The search strategy identified no relevant reviews.

PRIMARY RESEARCH

The search identified four eligible primary research studies. Below is an overview of study designs and aspects of quality represented by these studies. Full details of the papers appraised, including methods, key results, limitations and conclusions, are provided in evidence **Table 14**. Studies are presented in chronological order of publication within the tables.

Study design and quality assessments

As discussed, an RCT would be the most robust method of comparing the usefulness for surveillance diagnostic tests. However, no such evidence was identified. Mammographic surveillance of women at high risk of breast cancer has become accepted practice on the basis of it demonstrating equivalent breast cancer detection rates to established screening programmes in women over 50 years of age, and on the basis of surveillance detected-tumours in high-risk women showing more favourable characteristics than tumours arising in women not receiving surveillance. It would be considered unethical to offer women at high risk surveillance by CBE or US alone, without XRM. Therefore the four studies which examined the accuracy and effectiveness of US surveillance in populations of high-risk women compared to CBE, also used XRM for surveillance. The XRM results will be discussed in the following chapter. All four eligible studies were graded evidence level III-2. They were all prospective cohort studies.

The studies were of moderate quality in design and conduct. There were several limitations that apply to all of them. They were all likely to be affected by verification bias. This is because the reference standard for diagnosis was different in the case of a positive surveillance result versus a negative result. Positive surveillance screens were followed by biopsy or surgical excision and histopathological confirmation. However, verification of negative surveillance screens was only possible through clinical follow-up over the surveillance interval. This follow-up would detect false negatives that arose as interval tumours but would not detect false negatives that did not present symptomatically. The duration of follow-up was not reported for one study (Podo et al. 2002), but in all the other studies the lower range of follow-up was equivalent to or greater than the surveillance interval and therefore all interval tumours should have been captured.

Another limitation is that ultrasound examination is extremely prone to inter-observer variability. This can affect both the internal and external validity of the study. However, this variability is also likely to be encountered in non-study, i.e. routine use settings. There were also different systems of classification used for US and differing cut-off points determining an abnormal examination. The level at which this is set would influence the outcomes of the study and also the ability to draw conclusions across studies. The radiologist was not always blinded to the women's risk status in these studies, or to the results of the other screening modalities. This knowledge may affect their degree of suspicion and therefore the thoroughness with which they carried out the examination.

Study setting

Three studies were undertaken in single centres (Hou et al. 2002; Kuhl et al. 2005b; Trecate et al. 2003) and one was multi-centred (Warner et al. 2004). Hou et al. (2002) recruited relatives of breast cancer patients in a hospital in Taiwan, Trecate et al. (2003) recruited participants from the National Cancer

Institute in Milan, Warner et al. (2004) recruited participants from familial cancer centres in Canada and Kuhl et al. (2005b) recruited participants from high risk breast clinics in one hospital in Germany. As discussed in the chapter on surveillance by XRM, the setting of the study usually determines the prevalence and spectrum of disease in the participant population (Deeks 2001). However, once again, this was also determined by the risk stratification that participants underwent.

Risk stratification

The methods of risk stratification varied between the studies. Due to the variety of risk stratification strategies and the differing methods employed in the studies, the remainder of the information on these studies is presented individually.

Hou et al. (2002)

Study sample

This prospective cohort study recruited 935 women who were relatives of breast cancer patients in a Taiwanese hospital. All participants were aged over 35 years and the mean age at screening was 48.6 years, with a range of 35-75 years. There was no specific risk stratification process, but all participants had at least one first-degree (mother, sister or daughter) or second-degree (grandmother) relative with breast cancer. Exclusion criteria were: pregnant or lactating women, a past history of breast cancer or known metastatic disease.

Interventions and comparators

Surveillance consisted of annual CBE, XRM and US. US was performed with a 7.5MHz frequency transducer probe. The BIRADS system was used for both US and XRM with scores of 4 and 5 leading to biopsy. It was not reported whether radiologists interpreting the US scans were blinded to the results of CBE. The median follow-up time was 41.8 months, with a range of 12-82 months. Comparisons were drawn between the different modalities of screening.

Outcomes

Cancer detection rate

Twenty-one cancers were detected overall. This gives a cancer detection rate of 22 per 1,000 women under surveillance. Seven of the tumours were detected by CBE, i.e. seven per 1,000 women under surveillance and 19 by US, i.e. 20 per 1,000 women under surveillance.

Sensitivity

The sensitivities presented in the paper are only calculated with the surveillance-detected (by all modalities) cancers as the denominator, they did not include the interval cancer. The documented figures are 33.3 per cent for CBE and 90.4 per cent for US. There are no confidence intervals documented. If calculated with the interval tumour as a false negative, the respective results are 32 per cent (95% CI, 13.9 to 54.9%) and 86.4 per cent (95% CI, 65.1 to 97.1%).

Specificity

It is unclear how the specificities have been calculated. The documented figures are 83.5 per cent for CBE and 86.3 per cent for US. There are no confidence intervals documented.

Tumour characteristics

Sixteen tumours were invasive, two were DCIS, two were mucinous carcinomas and one was a medullary carcinoma. The mean tumour size was 12mm and seven were lymph node positive. These characteristics are not documented as stratified by mode of detection.

Interval tumours

There was one interval tumour.

Survival

The five-year overall survival was 90.4 per cent and the disease-free survival rate was 80.9 per cent. This calculation was not adjusted for lead time bias or length bias.

In summary, this study suggested that US surveillance was more accurate and effective then CBE at detecting tumours in women at high risk of breast cancer in Taiwan. However, there were no measures of statistical significance documented in this study. It is noted that this finding may be related to Asian women generally having smaller denser breasts and that these findings may not be reproducible in a Western population.

Trecate et al. (2003)

Study sample

This prospective cohort study recruited 23 women at high risk of breast cancer from the National Cancer Institute of Milan, Italy. There was no age restriction and no average age of the cohort was given. The age range was 30-61 years. Risk stratification was specific to this study. The women included were either BRCA1 or BRCA2 mutation carriers, had a one in two probability of being a carrier or over a 50 per cent risk of carrying a susceptibility gene for familial breast cancer based on family history. Women with a personal history of breast cancer were included (six women).

Interventions and comparators

Surveillance depended on the age group of the women. All ages had CBE every six months. Mammography was annual and commenced at 25 years with bilateral one-view, and then increased to bilateral double-view from 30 years and above. Annual US was performed alone from 20-25 years, then with XRM from 25-35 years, then six months after XRM from 35-40 years and above 40 years only if requested by the radiologist. The US was performed with either 7.5 or 10-12MHz probes (ATL HDI 3500, Philips). Annual MRI was performed for all ages for two years during the study. The MRI results are reported in a subsequent chapter. The method of classifying the images was not documented. Follow-up was not documented. It was not reported whether the radiologists interpreting the US scans were blinded to the results of CBE. The study was conducted over a seven-month period but the dates were not given. It is unclear if this work may have been related to the study by Podo et al. (2002). Comparisons were made between the different modalities of surveillance.

Outcomes

Cancer detection rate

Four breast cancers were detected overall. This gives an overall detection rate (including MRI) of 170 per 1,000 women under surveillance. Three tumours were detectable by CBE (130 per 1,000 women under surveillance) but none of the tumours were detected by US.

Measures of accuracy

No measures of accuracy were calculated in this study.

Tumour characteristics

All four tumours were invasive. Only two tumours had the size recorded and these were 10mm and 30mm. No record of the lymph node status was documented. There was no stratification of tumour characteristics by modality of surveillance.

Interval tumours

No interval tumours were documented.

In summary, this study suggests that US is not an effective addition to surveillance for breast cancer in women at very high risk of breast cancer i.e. mostly mutation carriers with a high proportion having a personal history of breast cancer. The results are extremely limited by the very small sample size, small number of tumours detected and the lack of detail documented in the publication. The study focuses on very high risk women and may not be generalisable to all women at high risk of breast cancer.

Warner et al. (2004)

Study sample

This prospective cohort study recruited 236 female BRCA1 and BRCA2 mutation carriers from familial cancer centres in southern Ontario and Montreal in Canada. There were no age restrictions and the mean age at first surveillance was 46.6 years with a range of 25-65 years. Risk stratification was performed by all participants being BRCA1 or BRCA2 mutation carriers. This was therefore a very high risk group, 31 per cent of whom were of Ashkenazi Jewish descent. In addition, 30 per cent had a personal history of breast cancer. Exclusion criteria were: a past history of unilateral breast cancer if the contralateral breast was not intact, pregnant or lactating women, history of bilateral breast cancer currently undergoing chemotherapy or known to have metastatic disease and women weighing over 91kg (technical reasons). Thirty-one women left the study before completing three rounds, 16 underwent bilateral mastectomy, three were too large for MRI machine, three stopped due to pregnancy, four developed metastatic cancers, four were lost to follow-up and one did not wish to continue participating.

Interventions and comparators

Surveillance consisted of biannual CBE and annual XRM, US and MRI, all performed on the same day. Surveillance commenced at least one year after the woman's last mammogram. CBE was coded as normal, suggestive of benign disease, indeterminate, or suspicious of malignancy. Indeterminate CBE exams were repeated after three months. The MRI results are discussed in a subsequent chapter. US used a 7.5MHz transducer (the first seven patients did not receive US). All participants underwent the first screen, but only 58 per cent had the second and 36 per cent the third. BIRADS was used to classify the images and scores of 4 or 5 were biopsied. Each imaging study was read and scored independently by a radiologist who specialised in breast imaging, and the radiologists were blinded to the CBE results. All patients were followed up for a minimum of one year after their last screening examination. Comparisons were drawn between different modalities of surveillance.

Outcomes

Cancer detection rate

Twenty-two cancers were detected overall in 21 women. Seven of these women had a past history of breast cancer. This gives an overall cancer detection rate, including MRI, of 93 per 1,000 women under surveillance. Two were detected by CBE (8 per 1,000 women under surveillance) and seven by US (30 per 1,000 women under surveillance). Two tumours were detected by US alone.

Sensitivity

All the measures of accuracy in the paper are presented individually for each year of surveillance. These results have been combined to give overall results for the three rounds of surveillance. There was not enough raw data to calculate measures of accuracy for CBE.

The sensitivity of US was 33 per cent (95% CI, 14.6-56.9). The sensitivity of CBE was 9 per cent (1-29%).

Specificity

The specificity of US was 96 per cent (95% CI, 93.7-97.7%).

PPV

The PPV of US was 29 per cent (95% CI, 12.6-51.1%).

NPV

The NPV US was 97 per cent (94.5-98.2%).

AUC

The AUC for US was 0.65. The AUC for CBE is also given at 0.48. There were no confidence intervals documented for the AUCs.

Tumour characteristics

Sixteen tumours were invasive and six were DCIS. The mean size of the invasive tumours was 11mm at the first surveillance round and 13mm at the second round. Fifteen cases had lymph node sampling and two were node-positive. The tumour characteristics are not documented stratified by modality of surveillance.

Interval tumours

There was one interval tumour, detected seven months after a third screen.

Mortality

All 22 patients with tumours were still alive and disease-free at the time of publication of the article.

In summary, this study suggests superior efficacy and accuracy of US to CBE in detecting early breast cancer through the surveillance of high-risk women. The results of this study are limited to the very high risk population of women who are proven BRCA1 or BRCA2 mutation carriers, including those with a personal history of breast cancer. It is therefore not generalisable to all women with an increased risk of breast cancer due to a family history. Further studies with larger numbers and longer follow-up, and including women of other risk groups are required.

Kuhl et al. (2005b)

Study sample

This prospective cohort study recruited 529 women from high-risk clinics in a single hospital in Germany. There was no age restriction and the mean age of the whole cohort was 41.7 years, with a range of 27-59 years. Risk stratification was performed according to the Consortium on Familial Breast and Ovarian Cancer of the German Cancer Aid. All participants had over a 20 per cent lifetime risk of breast cancer. In women that did not have a personal history of breast cancer, the Claus tables were also used to stratify risk. Women with a personal history of breast cancer were included, provided the women had not had bilateral mastectomy, had not had chemotherapy within the last 12 months and had no metastases (139 women had a personal history). Another inclusion criterion was being asymptomatic.

Interventions and comparators

Surveillance consisted of biannual CBE and US and annual XRM and MRI. If abnormalities found on CBE or US at the round without XRM or MRI, these additional modalities were used to further investigate this. Surveillance commenced at 30 years, or five years before the youngest family member affected with the disease. (NB, in the first two years, women under 30, or 30-39 years with dense breasts did not receive XRM, but this was subsequently abandoned and all women received XRM though these data were not included in the calculation of accuracy measures). The MRI results are reported in a subsequent chapter. US was performed with 7.5-13MHz probes. Each imaging study was read and scored independently by a different radiologist who had substantial experience with the respective imaging technique. The radiologists were informed about the clinical findings from CBE and the risk status of the patient but were blinded to the results of the respective other imaging modalities. BIRADS was used to classify the images and scores of 4 or 5 went for biopsy. The mean follow-up time was 5.3 years, with a range of 2-7 years. The number of total annual surveillance rounds for which data on all three imaging modalities was available was 1,452, and this was used in the calculation of accuracy measures. Comparisons are made between the three risk groups and the different modalities of surveillance.

Outcomes

Cancer detection rate

A total of 43 tumours arose in 41 patients during the study period. Forty of these were detected by imaging. That gives a cancer detection rate for the overall surveillance strategy (including MRI) of 76 per 1,000 women under surveillance. Eleven (25%) of these patients had a prior history of breast cancer. CBE identified only one tumour (2 per 1,000 women under surveillance) which was also

detected on imaging. Seventeen tumours were detected by US (32 per 1,000 women under surveillance), two of these were at the half-yearly CBE and US screen and were not palpable.

Sensitivity

The overall sensitivity for US was 39.5 per cent (95% CI, 25.0 to 55.6%) and for CBE was 2.3 per cent (95% CI, 0.1 to 12.3%).

When stratified by risk groups, US became less sensitive as the lifetime risk of breast cancer increased, with a sensitivity of 67.7 per cent in women with a 20 per cent lifetime risk, 30 per cent in women with a 21-40 per cent lifetime risk and 25 per cent for the mutation carrier group. The average age of the women decreased as the risk of breast cancer increased, and this may have contributed to the decreasing sensitivity. However, the difference in age is only small and unlikely to account for the whole effect. The mean ages and age ranges in the three groups were 43.8 years (range 35-59), 40.3 years (31- 57) and 38.9 years (27-51) in the lifetime risk of 20 per cent (21-40%) and the mutation carriers respectively. A more aggressive nature of tumours, or a different histopathology i.e. prominent pushing margins are other factors that may have contributed to the decrease in sensitivity in the highest risk women.

Specificity

The overall specificity for US was 90.5 per cent (95% CI, 88.8-92.0%). There was insufficient data to calculate the specificity for CBE.

Stratification by risk group or by a past history of breast cancer does not appear to affect the specificity.

PPV

The overall PPV for US was 11.3 per cent (95% CI, 6.7-17.4%). There was insufficient data to calculate the PPV for CBE.

Stratification by risk group or by a past history of breast cancer does not appear to affect the PPV either.

Tumour characteristics

Thirty-four tumours were invasive and nine were DCIS. The characteristics of the tumours are presented together for XRM and US. Of the 21 cancers detected by XRM and US, 16 were invasive and the rest were DCIS. The invasive cancers had a mean size of 13.9mm and five were node-positive. There were no significant differences in the characteristics of the tumours detected by XRM or US (p values all > 0.05).

Interval tumours

The interval tumour rate is given as 2 per cent in this cohort. It is unclear if this is a percentage of the women under surveillance or of the tumours that arose. It was also documented that there was one interval cancer that arose between surveillance rounds. However, it was reported that 40 of the 43 cancers were detectable by imaging, which would suggest three interval cancers. These figures were reported in an unclear manner.

In summary, this study suggests that the addition of US to CBE does improve the cancer detection rate in the surveillance of women at high risk of breast cancer. There is not enough data to calculate measures of accuracy for CBE in this study. The data on interval tumours is somewhat unclear in its documentation. This study included women at high risk who had a personal history of breast cancer, but the majority of the results were not significantly different if stratified by personal history.

Summary

Four studies were identified of relevance to the accuracy and efficacy of surveillance with US of women at high risk of breast cancer. These were all prospective cohort studies. A total of 1,723 women underwent surveillance in the four studies. There was heterogeneity between the studies in terms of the surveillance strategies, screening intervals, the participants' risk status and age. Warner et al. (2004) recruited participants who were BRCA1 or BRCA2 mutation carriers. Trecate et al. (2003) also recruited mutation carriers of those who had over a 50 per cent chance of being a carrier. Hou et al.

(2002) recruited first-degree or second-degree relatives of patients with breast cancer and Kuhl et al. (2005b) recruited women with greater than a 20 per cent lifetime risk of breast cancer. Two studies included women with a personal history of breast cancer (Kuhl et al. 2005b; Trecate et al. 2003) and two excluded them (Hou et al. 2002; Warner et al. 2004).

In all the studies surveillance consisted of CBE, XRM and US. Three of the studies also included MRI in their surveillance strategies. The XRM and MRI results will be discussed in subsequent chapters. Ultrasound was carried out with 7.5MHz frequency probes in two studies (Hou et al. 2002; Warner et al. 2004), Trecate et al. (2003) used a 7.5MHz or 10-12MHz probe and Kuhl et al. (2005b) used a 7.5-13MHz probe. BIRADS was used to classify the images in three of the studies with biopsies being performed for lesions scoring over or equal to 4. There was no documentation of the classification system used by Trecate et al. (2003).

All four studies looked at the intermediate outcome of cancer detection rate and three calculated measures of accuracy (Hou et al. 2002; Kuhl et al. 2005b; Warner et al. 2004). All the studies documented their tumour characteristics, but these were not stratified by the modality of surveillance nor were they compared with tumours in populations that had not undergone surveillance. Hou et al. (Hou et al. 2002) calculated survival but did not adjust this for lead time. Warner et al. (2004) documented mortality but there was not long enough follow-up or a large enough sample for this to be meaningful data.

The results of the cancer detection rates and measures of accuracy are summarised in **Tables 12** and **13**.

Table 12. Cancer detection rates in surveillance of women at high risk of breast cancer with CBE and US

Study	Cancer detection rate overall	Cancer detection rate by modality	
		CBE	US
Hou et al. (2002)	22 per 1,000 w/s	7 per 1,000 w/s	20 per 1,000 w/s
Trecate et al. (2003)	170 per 1,000 w/s	130 per 1,000 w/s	0
Warner et al. (2004)	93 per 1,000 w/s	8 per 1,000 w/s	30 per 1,000 w/s
Kuhl et al. (2005b)	76 per 1,000 w/s	2 per 1,000 w/s	32 per 1,000 w/s

w/s = women under surveillance
The cancer detection rates reported by women under surveillance cannot be compared across studies due to the differing surveillance intervals and differing length of the studies.

Table 13. Measures of accuracy in surveillance of women at high risk of breast cancer with CBE and US

Study	Accuracy	CBE (95% CI)	US (95% CI)	*P* values
Hou et al. (2002)	Sensitivity	31.8% (13.9% to 54.9%)	86.4% (65.1% to 97.1%)	No *p* values were calculated in this study
	Specificity	99.4% (98.7% to 99.8%)	99.4% (98.7% to 99.8%)	
Trecate et al. (2003)	No measures of accuracy documented			
Warner et al. (2004)	Sensitivity	9.1% (1% to 29%)	33.3% (14.6% to 56.9%)	=0.05
	Specificity	N/R	96.0% (93.7% to 97.7%)	N/A
	PPV	N/R	29.2% (12.6% to 51.1%)	N/A
	NPV	N/R	96.7% (94.5% to 98.2%)	N/A
	AUC	0.48	0.65	
Kuhl et al. (2005b)	Sensitivity	2.3% (0.1% to 12.3%)	39.5% (25.0% to 55.6%)	<0.001
	Specificity	N/R	90.5% (88.8% to 92.0%)	N/A
	PPV	N/R	11.2% (6.7% to 17.4%)	N/A
	NPV	N/R	98% (97.1% to 98.7%)	N/A

N/R = not reported N/A = not applicable

These results show that the cancer detection rate for US is on the whole greater than that from CBE alone. The study by Trecate et al. (2003) does not support this, but this study had a very small sample size and number of tumours detected, and as such was unreliable. The results of two studies (Kuhl et al. 2005b; Warner et al. 2004) show US to be statistically significantly more sensitive than CBE in the

surveillance of women at high risk of breast cancer. Kuhl et al. (2005b) shows that the sensitivity of US decreases as the risk status of women increases, being especially low for mutation carriers. Warner et al. (2004) document a higher AUC for the ROC curve for US than for CBE. Hou et al. (2002) demonstrate no difference in specificity between US and CBE. The other studies do not have comparative data for the specificity of CBE. It is noted that the PPV of US is low, 29.2 per cent in Warner et al. (2004) and 11.2 per cent in Kuhl et al. (2005b). This is due to the high number of false - positives generated by US examination. This is of importance due to the anxiety and potential harm related to further invasive investigation of abnormal surveillance results. All the studies also documented the interval tumours. However, these were for the overall surveillance strategies which in three cases included MRI. It is likely that the interval tumours would have been greater if MRI had not been included in the protocol.

In conclusion, it appears that surveillance with US is more sensitive than surveillance with CBE alone in women at high risk of breast cancer. However, the sensitivity is still relatively low except for in the study by Hou et al. (2002). The higher sensitivity in this latter study is attributed to the population under surveillance being all Asian women and it is unlikely to be generalisable to a Western population. The sensitivity of US also decreases with the increasing risk of breast cancer (Kuhl et al. 2005b). The low sensitivity of surveillance with CBE and US alone suggest that there would be a high rate of interval cancers. This cannot be assessed due to the design of these studies. This would suggest that CBE and US alone are not adequate for the surveillance of women at high risk of breast cancer and that other modalities of screening are required in addition or instead of this strategy. The following chapter reviews the evidence for surveillance with US and XRM for women at high risk of breast cancer.

Table 14. Primary research studies appraised investigating the accuracy and efficacy of ultrasound surveillance compared to usual care on outcomes from breast cancer

Source	Study design Evidence grading	Comparison interventions and dates of testing	Sample	Outcomes and verification	Results	Comments
Hou et al. 2002 Taiwan	Prospective cohort Study III-2 (C1 P1 Q2)	Surveillance protocol: annual CBE, mammography and US. Pre-menopausal women underwent surveillance during the 2nd week of the menstrual cycle to minimise the occurrence of breast densities or enhancing masses related to the menstrual cycle. 4-view film mammograms were conducted and reviewed by one radiologist. US performed with a 7.5MHz frequency transducer probe. Dates were May 1994 to August 200. No comparisons were made in this study other than between modalities of surveillance. BIRADS was used to classify lesions and a cut off of 4 or above determined an abnormal result.	Sample no: 935 women. Mean age (at surveillance) = 48.6 years (range 35-75) Recruited inclusion criteria: • >35 years old; • female relatives of breast cancer patients (mothers, daughters, grandmothers, sisters). Exclusion criteria: • pregnant or lactating; • past history of breast cancer; • known metastatic diseases. No specific risk stratification process carried out' just all relatives of breast cancer patients.	Relevant outcomes: • cancer detection rate; • mode of detection; • tumour stage; • node status; • interval tumours; • 5-year overall survival and event-free survival (free from cancer related death and tumour spread; • sensitivity • specificity Verification of positive result, by any of 3 surveillance modalities, was through biopsy and pathology results. Verification of a negative result was through follow-up. Median follow- up was 41.8 months (range 12-82 months) Verification of interval cancers.	Cancer detection rate: 21 cancers were detected, giving an overall cancer detection rate of 22 per 1,000 women under surveillance. Of the women with tumours, 1 was a BRCA1 mutation carrier, 2 were BRCA2 mutation carriers and the other 18 were mutation status unknown. Mode of detection: CBE detected 7 tumours. mammography detected 11 tumours. US detected 19 tumours. Tumour size, stage and node status: 16 were invasive cancers, 2 were DCIS, 2 were mucinous carcinomas and 1 was a medullary carcinoma. Mean tumour size was 12mm. 7 were node-positive and 14 were node-negative. 1 interval cancer was reported Five year overall survival was 90.4% and the disease free survival rate was 80.9%.	Limitations included: Verification bias is likely. Lead time bias and length bias are likely in terms of the survival data. This population was not explicitly risk-stratified and it is difficult to assess their overall risk of breast cancer. There are no characteristics of the overall group of women, other than being relatives of breast cancer patients and the mean age. It is unclear if they have any additional risks for breast cancer. Only a prevalent round was examined and it is likely that the cancer detection rate would be higher in this round than in subsequent surveillance rounds, There is no mention of how interval cancers are verified as being true interval cancers. Authors' conclusions: Based on a higher sensitivity of sonography for detecting breast cancer in the high-risk group in our study, sonography is superior to mammography and physical examination of the breasts in the surveillance of women at high risk for breast cancer in Taiwan. If sonography will replace mammography as a surveillance tool, it needs further research.

Table 14. Primary research studies appraised investigating the accuracy and efficacy of ultrasound surveillance compared to usual care on outcomes from breast cancer (continued)

Source	Study design Evidence grading	Comparison interventions and dates of testing	Sample	Outcomes and Verification	Results	Comments
Hou et al. 2002 Taiwan *Continued*					The documented sensitivities were: CBE 33.3% Mammography 52.4% US 90.4% The documented specificities were: CBE 83.5% Mammography 82.2% US 86.3% If these figures are recalculated with the interval tumour included in the denominator the results are: Sensitivity: CBE 31.8% (95% CI, 13.9-54.9%) XRM 50% (95% CI, 28.2 - 71.8%) US 86.4% (95% CI, 65.1-97.1%) Specificity: CBE 99.4% (95% CI, 98.7-99.8%) XRM 99.6% (95% CI, 98.9-99.9%) US 99.4% (95% CI, 98.7-99.8%) The figures for specificity do not agree with those in the paper and it is not evident how they were calculated. There were no measures of statistical significance documented in this study.	Otherwise, the low cost of US and convenience for women who live in rural areas suggests that sonography will be a useful tool for breast cancer surveillance in Taiwanese women in the high-risk group and in countries with a low incidence of breast cancer. Reviewers' conclusions: This study suggests that sonography is much more accurate for the surveillance of women at high risk of breast cancer, than mammography or CBE. However, these findings are specific to this population and are not generalisable. As discussed by the authors, the sensitivity of mammography is likely reduced by the higher proportion of Asian women with smaller denser breasts, which are less fatty and also the overall lower incidence if breast cancer in this Taiwanese population. Sonography may be a useful surveillance modality in these women, and especially in rural areas or areas without access to MRI, however it is unlikely to achieve such good results in a Western population.

Table 14. Primary research studies appraised investigating the accuracy and efficacy of ultrasound surveillance compared to usual care on outcomes from breast cancer (continued)

Source	Study design Evidence grading	Comparison interventions and dates of testing	Sample	Outcomes and verification	Results	Comments
Trecate et al. (2003) Italy (NB: Podo is an author on this one as well but we cannot find any further reports from the Podo et al trial.)	Prospective cohort study III-2 (C1 P2 Q3)	Surveillance protocol: Outlined in full in the paper and was dependent on age group, CBE was performed every 6 months for all ages. Mammography was annual and commenced at 25 years with bilateral one-view, and then increased to bilateral double-view from 30 years and above. Double-view was performed in craniocaudal and mediolateral oblique projections. One-view was performed in the mediolateral oblique projection for younger women. Annual US was performed alone from 20-25 years, then with mammography from 25-35 years, then 6 months after mammography from 35-40 years and above 40 years only if requested by the radiologist. US was performed with either 7.5 or 10-12MHz probes (ATL HDI 3500, Philips).	Sample no = 23 women at high risk of breast cancer (2 cases did not get US). No average age of women given, range was 30-61 years. Inclusion criteria • BRCA1 or BRCA 2 mutation carrier, or 1 in 2 probability to be a mutation carrier (on the basis of positive mutational analysis in close relatives) with a negative or positive personal history for breast or ovarian cancer . OR • High risk for breast cancer according to criteria specified in paper. Risk stratification: As above, either BRCA1 or BRCA2 carrier, 1 in 2 probability of being a carrier or >50% risk of carrying a susceptibility gene for familial breast cancer on basis of family history.	Relevant outcomes: • cancer detection rate; • mode of detection; • tumour size and stage; Verification of positive results was with pathology and verification of negative results was with follow-up. There is no mention of the mean length of follow-up.	Cancer detection: 4 breast cancers were detected overall. Mode of detection: 3 were detectable by CBE but none of the tumours were detected by US examination (although 1 woman did not receive an US). Tumour size and stage: All 4 tumours were invasive: 2 ductal invasive carcinomas, 1 lobular invasive carcinoma and 1 which was mixed ductal and lobular. 2 occurred in mutation carriers and 2 in women at high risk through family history. Only 2 tumours had the size recorded and these were 10mm and 30mm. No record of nodal status was given. There was no mention of interval tumours	Limitations include: Small sample size. There are few characteristics given of the women selected other then their risk assessment. There is no information on how they were selected and the characteristics of any women who did not agree to participate. There is no mention of mean age, reproductive history, exogenous hormone use or preventative strategies (i.e. Tamoxifen use or BSO). There is also no indication of which women were having prevalent or incident surveillance screens and for how long they were followed up in the study. There is likely verification bias and this is more likely, the shorter the follow-up period. Authors' conclusions: The authors' conclusions relate to the overall surveillance strategy, including XRM and MRI. Breast MRI demonstrated to be a very useful technique for investigating breast disease. It is not influenced by breast density and does not use ionising radiation. For these reasons, it has been proposed to support mammography in the surveillance of BRCA mutated patients. Moreover, according to the reported results, breast MRI seems very helpful in the high risk patients group. We believe the breast MRI can be very useful within this kind of surveillance, with a less invasive approach to the disease.

Table 14. Primary research studies appraised investigating the accuracy and efficacy of ultrasound surveillance compared to usual care on outcomes from breast cancer (continued)

Source	Study design Evidence grading	Comparison interventions and dates of testing	Sample	Outcomes and Verification	Results	Comments
Trecate et al. (2003) Italy *Continued*		MRI was performed annually for all ages for 2 years during the study. A Siemens Vision 1.5 was used with a dedicated double coil. One pre-contrast image and 5 post-contrast images were taken. The contrast agent was Gd-DTPA at 0.1mmol/kg. The method of interpreting the MRI or mammography is not presented. The study was conducted over a 7-month period; however the exact dates are not given.	The latter refers to at least 3 cases of breast cancer before 60 years of age, at least 3 cases of breast cancer before 60 years of age and ovarian cancer at any age, or at least 3 cases of breast cancer before 60 years of age and male breast carcinoma at any age. 5 of the women had a personal history of breast cancer, 1 for ovarian cancer and 1 for ovarian and breast cancer. (1 had had a mastectomy, but the others had conservative surgery combined with radiation therapy).			In the case of confirmed good diagnostic results, it could be proposed to be used every other year as an alternative to mammography. Reviewers' conclusions: This study suggests that CBE is better than US at detecting tumours in women at high risk of breast cancer. However, the sample is very small, a small number of tumours were detected and it is difficult to know how long the women were followed up for and this would affect the reliability of the results. There could be false-negatives that had not yet come to light. There is also a specific method of risk stratification in this study, which includes women with a personal history of breast cancer (although only if they are BRCA1 or BRCA2 mutation carriers), and this will affect the generalisability of the study. In addition the results are not presented in a very clear manner and it is difficult to determine the overall sensitivity and specificity for all the surveillance modalities utilised, which would have been valuable information.

Table 14. Primary research studies appraised investigating the accuracy and efficacy of ultrasound surveillance compared to usual care on outcomes from breast cancer (continued)

Source	Study design Evidence grading	Comparison interventions and dates of testing	Sample	Outcomes and verification	Results	Comments
Warner et al. (2004) Ontario and Montreal, Canada	Prospective cohort Study III-2 (C1 P2 Q2)	Study protocol: CBE biannually and mammography, US and MRI all performed annually 4 modalities all performed the same day. (commencing at least 1 year after the woman's last mammogram) CBE coded as normal, suggestive of benign disease, indeterminate, or suspicious of malignancy. Indeterminate CBE exams were repeated after 3 months. Mammography was conventional 4-view film. Further views done when necessary. MRI was performed with 1.5 T magnet (Signa, General Electrical Medical Systems). The first 38 patients in the first year were done in a single-turn elliptical coil after a bolus injection of 0.1mmol/kg of Gd-DTPA. Images were taken in the coronal plane.	Sample no = 236 female BRCA1 and BRCA2 mutation carriers. Mean age at first surveillance 46.6 years (range 25-65 years). Mean age of diagnosis was 47.4 years (33.4-63 years) Recruited from familial cancer clinics. Inclusions: • BRCA 1 or BRCA2 mutation carrier. Exclusions: • past history of unilateral breast cancer if the contra lateral breast not intact; • pregnant or lactating women (participation deferred); • history of bilateral breast cancer, currently undergoing chemotherapy or known to have metastatic disease.	Relevant outcomes: • cancer detection rate; • mode of detection • tumour stage, size and node status; • interval cancers; • mortality; • sensitivity; • specificity; • PPV; • NPV; • ROC curves; NB: The PPV and specificity do not include in the denominator women that had additional diagnostic studies that did not result in biopsy. Verification of positive results was by pathology; biopsy was undertaken if there was suspicion from any of the 4 modalities of surveillance. Verification of a negative result was through follow-up.	Cancer detection: 22 cancers were detected in 21 women (1 bilateral). 7 of these women had previous breast cancer. Mode of detection: 2 were detected by CBE (9.1%). 7 by US (33%). 2 were detected by US alone (9.5% and not all women had undergone US testing). Tumour stage, size and node status: 6 tumours were DCIS and 16 were invasive (15 infiltrating ductal and 1 invasive lobular). The mean size of the invasive tumours was 11mm at the first round and 13mm at the second round (overall range 5mm-60mm). 15 cases were node sampled and 2 were node positive. Interval cancers: There was only 1 interval cancer, detected in a 40 year old BRCA1 mutation carrier 7 months after her 3rd screen.	Limitations included: Likely verification bias. Selected participants are very high risk, being proven mutation carriers and also including those with a prior history of breast cancer. It is not clear which rounds were incident and which were prevalent and which tumours were detected at which round. (A large number of women had had prior mammography). No mention of whether women had had risk reducing measures such as bilateral salpingo oophorectomy or Tamoxifen. There was quite a high level of attrition in the study and the characteristics of those women are not outlined. This may have introduced bias. Authors' conclusions: The authors' conclusions relate to the overall surveillance strategy including XRM and MRI. This study of BRCA mutation carriers demonstrates that the addition of annual MRI and US to mammography and CBE significantly improves the surveillance for detecting early breast cancers. The use of US did detect additional tumours, but had a high false-positive rate and in light of this its benefit remains to be seen. There was no observed benefit from CBE over and above the 3 imaging modalities.

Table 14. Primary research studies appraised investigating the accuracy and efficacy of ultrasound surveillance compared to usual care on outcomes from breast cancer (continued)

Source	Study design Evidence grading	Comparison interventions and dates of testing	Sample	Outcomes and verification	Results	Comments
Warner et al. (2004) Ontario and Montreal, Canada *Continued*		For the remaining patients, a phased-array coil arrangement was used. This provided sagital images. US used a 7.5MHz transducer (the first 7 patients did not receive US) Each imaging modality was read independently by a radiologist and scored on the 5-point BIRADS scale. All lesions with a score of 4 or 5 were biopsied. Pre-menopausal women had surveillance performed mid menstrual cycle to avoid changes due to cyclical hormonal variation. Radiologists were blinded to the results of CBE 31 women left the study before completing 3 rounds, 16 underwent bilateral mastectomy, 3 were too large for MRI machine, 3 stopped due to pregnancy, 4 developed metastatic cancers,	• Women weighing >91kg (technical reasons). Risk stratification not really performed as only BRCA mutation carriers included (all very high risk group). There were 137 (58%) BRCA1 mutation carriers and 99 (42%) BRCA2 mutation carriers. 31% were Ashkenazi Jews. 30% had a history of breast cancer, 9% a history of ovarian cancer and 60% had no history of cancer or a history of another type of cancer. 85% of the women (*n*=205) had had mammography within the last 15 months and therefore this was an incident rather than a prevalent surveillance screen for them. 45% were pre-menopausal and 55% were post-menopausal.	All patients were followed up for a minimum of 1 year from the date of the last examination.	Another woman, who elected to have a bilateral mastectomy after breast cancer was found, had a 2mm focus of DCIS in the contra lateral breast which had not shown up at surveillance 2 months earlier. Mortality: All 22 patients who had tumours diagnosed were still alive and disease-free at the time the article was written. It was felt that the cancers detected on the second round were of an earlier stage. The 2 node-positive tumours were detected in the first round. However, it was not exactly clear that the first round was really a prevalent round as a high percentage of women had had prior mammography. It was found that false-positives and false-negatives decreased from the first to the second and then to the third round of surveillance. The measures of accuracy are therefore presented by the surveillance modality and by the year of the screen. These can be seen in the paper, but overall values for the 3 years are reported here.	MRI-based surveillance is likely to become the cornerstone of breast cancer surveillance for BRCA1 and BRCA2 mutation carriers, but it is necessary to demonstrate that this surveillance tool lowers breast cancer mortality before it can be recommended for general use. Reviewers' conclusions: This study demonstrates that US is better than CBE alone at detecting tumours in women with a high risk of breast cancer. As the authors suggest, this study does not answer whether this translates into reduced mortality. However, the tumours detected did seem to be of an earlier stage and smaller size, with only 2 tumours node-positive. The results of this study are limited to the very high risk population of women who are proven BRCA1 or BRCA2 mutation carriers and including those with a personal history of breast cancer. It may therefore not be generalisable to all women with an increased risk of breast cancer due to a family history. Further studies with larger numbers and longer follow-up, and including women of other risk groups, are required.

Table 14. Primary research studies appraised investigating the accuracy and efficacy of ultrasound surveillance compared to usual care on outcomes from breast cancer (continued)

Source	Study design Evidence grading	Comparison interventions and dates of testing	Sample	Outcomes and verification	Results	Comments
Warner et al. (2004) Ontario and Montreal, Canada *Continued*		4 were lost to follow-up and 1 did not wish to continue participating. All participants underwent the first round, but only 58% the second and 36% the third (a total of 120 women were still undergoing surveillance when the paper was written). No direct comparisons were made in this study. Dates of surveillance were between Nov 1997 and March 2003.			Measures of accuracy of individual modalities: Sensitivity (95% CI): US = 33% (14.6 to 56.9%) CBE = 9% (1% to 29%) Specificity (95% CI): US = 96% (93.7 to 97.7%) PPV (95% CI): US = 29% (12.6 to 51.1%) NPV (95% CI): US = 97% (94.5 to 98.2%) AUC: CBE = 0.48 US = 0.65 Unfortunately, the specificity, PPV and NPV could not be calculated for CBE as there was not enough data.	

Table 14. Primary research studies appraised investigating the accuracy and efficacy of ultrasound surveillance compared to usual care on outcomes from breast cancer (continued)

Source	Study design Evidence grading	Comparison interventions and dates of testing	Sample	Outcomes and verification	Results	Comments
Kuhl et al. (2005b) Germany	Prospective Cohort study III-2 (C1 P2 Q2)	Surveillance protocol: Biannual CBE and US and annual XRM and MRI. If abnormalities found on CBE or US at round without XRM or MRI, these additional modalities were used to further investigate this. Surveillance commenced at 30 years, or 5 years before the youngest family member affected with the disease. (NB: in first 2 years, women under 30, or 30-39 years with dense breasts did not receive XRM, but this was subsequently abandoned and all women received XRM). Mammography (XRM): Annual conventional film screen XRM performed with at least 2 views per breast (medio-lateral oblique and caudal-cranial), obtained and interpreted in accordance with German radiological practice guidelines. Diagnoses coded according to the BIRADS diagnostic categories on a 5-point scale (1, negative; 2, benign; 3, probably benign; 4, suspicious abnormality; 5, highly suggestive of malignancy).	Sample no = 529 (out of 590 eligible women, 49 were lost to follow-up after 1 surveillance round and 12 were also excluded as they had a clinical abnormality at initial examination). Inclusion criteria: • asymptomatic women; • personal history of breast cancer included, provided that the patient had not undergone bilateral mastectomy, had not received chemotherapy within the previous 12 months and had no metastases (139 women were included with a personal history of breast cancer). Exclusion criteria: • clinical signs of breast cancer; • chemotherapy within the previous 12 months;	Relevant outcomes: • cancer detection; • mode of detection; • tumour size; • tumour stage; • node status; • interval tumours; • sensitivity; • specificity; • PPV; • NPV. Verification of a positive result was achieved by histology (for positive imaging studies). Verification of a negative result was achieved by follow-up (for negative imaging studies). If a breast cancer was identified clinically (by palpation) between surveillance rounds or at the 6-month clinical visit the imaging studies of the previous round were considered false - negative. Mean follow-up was 5.3 years (range 2-7 years). A total of 1,542 annual surveillance rounds were completed.	Cancer detection: A total of 43 breast cancers were identified in 41 patients (11 of these women had a prior history of breast cancer), 40 of these were said to be detectable by imaging. Mode of detection: CBE identified only one tumour (also detected on imaging) US identified 17 tumours (2 of these were at the half-yearly CBE and US and they weren't palpable). Tumour size, stage and node status: Of the 21 cancers detected by XRM and US, 16 were invasive and the rest were DCIS. The invasive cancers had a mean size of 13.9 mm and 5 were node-positive. Interval tumours: The paper states that 40 out of 43 tumours in this cohort were detected by imaging. However, a sentence in the discussion states that the rate of interval cancers was 2% in this cohort. This translates to 10 tumours if it is 2% of the total population or 1 tumour if it is 2% of the total number of tumours. The latter is more likely as this would alter the figures the least, but it is unclear	Limitations included: CBE and the imaging studies were performed within a time frame of 8 weeks. Few sample characteristics presented, such as OCP or HRT use, or the use of preventative strategies such as tamoxifen or BSO. Verification bias is likely. Unclear reporting of interval tumours lack of blinding to the results of the CBE . Author's conclusions: The authors' conclusions relate to the overall surveillance strategy including XRM and MRI. If US is used in combination with XRM, it can help compensate for some but by far not for all of the shortcomings of XRM, and it causes a substantial number of false-positive diagnoses. If MRI is used for surveillance, XRM proved to be of limited and ultrasound of no additional value. US may, however, be useful to bridge the relatively long time interval between annual surveillance rounds. Propose that in view of the insufficient diagnostic accuracy of XRM and USS, breast MRI should be considered an integral part of surveillance programmes for women at high familial risk, in particular in documented carriers of pathogenic BRCA mutations. Reviewer's conclusions: Similar to those of the authors above. US has a higher cancer detection rate and sensitivity than CBE alone..

Table 14. Primary research studies appraised investigating the accuracy and efficacy of ultrasound surveillance compared to usual care on outcomes from breast cancer (continued)

Source	Study design Evidence grading	Comparison interventions and dates of testing	Sample	Outcomes and verification	Results	Comments
Kuhl et al. (2005b) Germany *Continued*		Breast MRI: Standard dynamic axial contrast-enhanced breast MRI of both entire breasts was performed on a 1.5T system (NT/INTERA; Philips, Best, the Netherlands) after injection of 0.1mmol/kg body weight gadopentetate dimeglumine (Magnevist, Schering, Berlin, Germany) Ultrasound (US): Performed with 7.5-13-MHz probes (Siemens Elegra, GE logic 500 and ATL HDI 5000; Siemens, Erlangen, Germany). The entire breast was systematically examined by the physician who interpreted the study. Diagnoses were scored on a 5-point scale identical to the XRM BIRADS categories. Each imaging study was read and scored independently by a different radiologist who had substantial experience with the respective imaging technique. The readers were informed about the clinical findings from CBE and the risk.	• women having undergone bilateral mastectomy. • Recruited from high-risk clinics in a single gynaecology department Risk Stratification: According to definition of the Consortium on Familial Breast and Ovarian Cancer of the German Cancer Aid ,corresponding to a lifetime risk of breast cancer of at least 20% (two or more cases of breast cancer on the same side of the family, including at least two cases with onset before age 50 years, or with breast or ovarian cancer, irrespective of age, families with at least one case of breast cancer diagnosed before 35 years, families with three or more cases of breast cancer on the same side of the family, and women who met the criteria for high familial risk, irrespective of the result of mutational analysis).	Verification of last surveillance round was by continued surveillance in 428 women, telephone interview in 52 women and for 6 women who had prophylactic mastectomy it was by pathology of the specimen. XRM: BIRADS of 4 or 5, biopsy was recommended irrespective of finding in US or MRI. BIRADS 3 was managed by 6-months follow-up until receiving a BIRADS 2 or biopsy clarification. US categorised as BIRADS 3 managed by short-term (6 months) US follow-up. BIRADS 4 or 5 managed by US-guided biopsy (14G, semi-automatic or automatic biopsy gun) except for the following constellation: if an US finding that was suspicious was clearly benign on XRM or MRI no biopsy was performed.	Comparisons: When stratified by risk groups, the detection rates at both the prevalent and incident rounds were much higher in the mutation carriers than the other 2 risk groups, but these differences are not statistically significant. Sensitivity (95% CI): US 39.5% (25.0 to 55.6%) n =17/43 CBE 2.3% (0.1 to 12.3%) n = 1/43 When stratified by risk groups, US becomes less sensitive as the lifetime risk of breast cancer increases, with sensitivities of 25% for the mutation carrier group. This effect is not seen with MRI which maintains good sensitivity across all risk groups. US Sensitivity by risk group: Risk 20% 67.7% (22 to 96%) n = 4/6 Risk 21 to 40% 30.0% (12 to 54%) *n* = 6/20 Mutation carriers 25% (3 to 65%) *n* = 2/8 Specificity (95% CI): US 90.5% (88.8 to 92.0%) *n* = 1,275/1,409	Unfortunately the specificity, PPV and NPV for CBE cannot be calculated as there is not enough data, so comparisons of these cannot be made. The sensitivity of US decreases with the increasing risk status of women, being only 25% in mutation carriers. The limitations of this study must be taken into account in the interpretation

Table 14. Primary research studies appraised investigating the accuracy and efficacy of ultrasound surveillance compared to usual care on outcomes from breast cancer (continued)

Source	Study design Evidence grading	Comparison interventions and dates of testing	Sample	Outcomes and Verification	Results	Comments
Kuhl et al. (2005b) Germany *Continued*		status of the patient but were blinded to the results of the respective other imaging modalities. Comparisons are made between the 3 risk groups and the different modalities of screening. Dates of study were February 1996 to February 2002.	In women without a personal history of breast cancer the Claus tables were also used to quantify risk. Women were then stratified into 3 risk groups for analysis: • mutation carriers; • high lifetime risk (20-40%); • moderate lifetime risk (20%).	MRI: Suspicious scores (4 or 5) were managed by magnetic resonance-guided biopsy. For findings categorised as BIRADS 3 short-term follow-up after 6 months was recommended with further management corresponding to that of XRM BIRADS 3 lesions BIRADS 3 categories in all imaging that received short-term follow-up were not considered positive for the calculation of outcomes. Invasive cancer and DCIS were considered a malignant diagnosis but LCIS and atypical ductal hyperplasia were considered to be benign.	Stratification by risk group does not appear to affect the specificity. PPV (95% CI): US 11.3% (6.7 to 17.4%) *n* = 17/151) The PPV increases with the increasing risk of breast cancer, this will be affected by the higher incidence in women at higher risk. NPV (95% CI): US 98% (97.1 to 98.7) *n* = 1275/1301	

Chapter 5: Accuracy and efficacy of ultrasound and mammography

SECONDARY RESEARCH

The search strategy identified no relevant reviews.

PRIMARY RESEARCH

The search identified nine eligible primary research studies. Below is an overview of study designs and aspects of quality represented by these studies. Full details of the papers appraised, including methods, key results, limitations and conclusions, are provided in evidence **Table 17**. Studies are presented in chronological order of publication within the tables.

Study design and quality assessments

As discussed, the most robust method of comparing the usefulness for surveillance of diagnostic tests would be an RCT. However, no such evidence was identified. All of the nine eligible studies were graded evidence level III-2. Of these nine studies, eight were prospective cohort studies (Crystal et al. 2003; Hou et al. 2002; Kolb et al. 1998; Kuhl et al. 2005b; O'Driscoll et al. 2001; Podo et al. 2002; Trecate et al. 2003; Warner et al. 2004) and one was a retrospective cohort study (Sim et al. 2004). Seven of the studies were designed to compare the accuracy and effectiveness of US surveillance to mammography plus CBE, with or without MRI, in populations of high-risk women under surveillance with all these modalities. The MRI results will be discussed in subsequent chapters. The remaining two studies examined US as an additional test exclusively in women whose mammography and CBE examinations were normal (Crystal et al. 2003; Kolb et al. 1998).

These studies were of moderate quality in design and conduct. There are several limitations that apply across all the studies. They were all likely to be affected by verification bias. This is because the reference standard for diagnosis was different in the case of a positive surveillance result versus a negative result. Positive surveillance screens were followed by biopsy or surgical excision and histopathological confirmation. However, verification of negative surveillance screens was only possible through clinical follow-up over the surveillance interval. The duration of follow-up varied between four studies with an overall range of 8- 84 months and in three studies was not reported (Kolb et al. 1998; Podo et al. 2002; Trecate et al. 2003). It is possible that interval tumours may not have been detected in case of inadequate follow-up after surveillance and therefore the effectiveness of the test would be over-estimated. Ultrasound also poses difficulty in the verification of interval tumours as it is performed in real-time and it is more complex. Also, it is sometimes not possible to retrospectively review previous examinations to see if the lesion was visible but overlooked. In addition, ultrasound examination is extremely prone to inter-observer variability. This can affect both the internal and external validity of the study. There were a variety of systems of classification used for US and differing cut-off points determining an abnormal examination. The level at which this is set would influence the outcomes of the study and also the ability to draw conclusions across studies. The radiologists were not always blinded to the women's risk status in these studies, or to the results of the other modalities of surveillance. This knowledge may affect their degree of suspicion and therefore the thoroughness with which they carried out the examination.

The one retrospective study by Sim et al. (Sim et al. 2004) is inherently more prone to bias and confounding due to its design. This was a reanalysis of data from an existing retrospective cohort study which focused on MRI and XRM surveillance (Stoutjesdijk et al. 2001). The aim of the primary study did

not include an analysis of US examinations. It is therefore likely that selection bias was operating in determining which women underwent US scans. In fact, there must have been some indication for US, which means they may have been more diagnostic than surveillant. There was also a lack of information on the participants and the study protocol in this study.

Study setting

Six studies were undertaken in single centres (Hou et al. 2002; Kolb et al. 1998; Kuhl et al. 2005b; O'Driscoll et al. 2001; Sim et al. 2004; Trecate et al. 2003) and three were multi-centred (Crystal et al. 2003; Podo et al. 2002; Warner et al. 2004). The centres of recruitment were usually breast screening or genetic screening clinics. As discussed in the chapter on surveillance by XRM, the setting of the study usually determines the prevalence and spectrum of disease in the participant population (Deeks 2001). However, once again, this was also determined by the risk stratification that participants underwent.

Risk stratification

The methods of risk stratification varied between the studies. As with the preceding chapters, the risk stratification strategies and the rest of the information for these studies is presented individually.

Kolb et al. (1998))

Study sample

This prospective cohort study recruited 3,626 women, including 565 high-risk women, from a private medical centre in the USA. There was no age restriction and the mean age was 59.3 years for the whole cohort. Risk stratification was not clearly documented, but all high-risk women were said to have a primary family member (mother, daughter, sister or brother) with breast cancer. Other inclusion criteria were: having dense breasts (defined as BIRADS category 2-4) and having had normal CBE and XRM. This meant that US was performed as an adjunct test. There was a second high-risk group consisting of 478 women with a past history of breast cancer. It is therefore assumed that none of the family high-risk group had a past history of breast cancer.

Interventions and comparators

Surveillance consisted of one round of US in women with normal CBE and XRM. The US bandwidth was 5-10MHz, set at 7.5MHz. The mean follow-up period was not reported. The classification system for images was not documented. Comparison of the cancer detection rates was made between the risk groups and the comparison between XRM and CBE surveillance was implicit as all these women had normal XRM. Conducting this study of US as an adjunct test means that the results will over-estimate the performance of US as there are no results for tumours detected by XRM that would not be detected by US.

Outcomes

Cancer detection rate

One tumour was detected amongst 565 high-risk women who had normal results from CBE and XRM. The rate of additional cancers detected by the use of US was 1.8 per 1,000 women screened. No significant difference was found between risk groups, even when a group of women at high risk due to a past history of breast cancer (in which five tumours were detected) was added to the group of women at high risk from a family history (p=0.09).

In summary, there was limited information arising from this study. The results are limited by the small number of tumours detected. It suggested that in women at high risk with dense breasts there may be tumours identified by US, when used as an adjunct test, that were not identified by CBE or XRM. The efficacy of US as an adjunct test was probably increased by restricting the inclusion to women with dense breasts. The results of this study are not generalisable to populations of women with less dense breasts.

O'Driscoll et al. (2001)

Study sample

This prospective cohort study recruited 149 women at moderate risk of breast cancer from a clinical genetics department in the U.\K. There were no age restrictions and the mean age on entering surveillance was 42 years, with a range of 30-69 years. Risk stratification was performed by criteria specific to this study. Moderate risk was defined as a risk three times greater than the age-matched general population. Women at higher risk were not included as they were already involved in other studies. There were no other inclusion or exclusion criteria and it was not documented whether women with a past history of breast cancer were included or excluded.

Interventions and comparators

Surveillance consisted of one round of XRM and bilateral breast US. The first 29 surveillance screens were performed with a 7.5MHz probe and the rest were done with an 8-12MHz probe. The mean follow-up time was 13.7 months. The classification system for images was not documented. An experienced breast radiologist, blinded to the XRM report, performed the breast US. The two modalities of surveillance were compared.

Outcomes

Cancer detection rate

One tumour was detected by US out of 149 women under surveillance. The cancer detection rate was 6.7 per 1,000 women under surveillance. This tumour was not detected by XRM. There was only one abnormal mammogram out of the 149 women screened and this was a false-positive, proven histologically to be a fibroadenoma.

Tumour characteristics

The one tumour detected was an adenoid cystic carcinoma which measured 11mm in diameter.

Interval tumours

There was one interval cancer. This presented 10.5 months after a negative mammogram and ultrasound examination. The pathology was an invasive lobular carcinoma with extensive lobular carcinoma *in situ* and there was no lymph node spread.

In summary, this study suggests that in this population of women at high risk of breast cancer there was a tumour detectable by US that was mammographically occult. The findings are limited by this being a pilot study with a small sample size and small number of tumours detected.

Hou et al. (2002)

Study sample

This prospective cohort study recruited 935 women who were relatives of breast cancer patients in a Taiwanese hospital. All participants were aged over 35 years of age and the mean age at screening was 48.6 years, with a range of 35-75 years. There was no specific risk stratification process, but all participants had at least one first-degree (mother, sister or daughter) or second-degree (grandmother) relative with breast cancer. Exclusion criteria were: pregnant or lactating women, a past history of breast cancer or known metastatic disease.

Interventions and comparators

Surveillance consisted of annual CBE, XRM and US. US was performed with a 7.5MHz frequency transducer probe. The BIRADS system was used for both US and XRM with scores of 4 and 5 leading to biopsy. It was not reported whether the radiologist interpreting the images was blinded to the results of the

other imaging modalities. The median follow-up time was 41.8 months, with a range of 12-82 months. Comparisons were drawn between the different modalities of surveillance.

Outcomes

Cancer detection rate

There were 21 cancers detected overall. This gives a cancer detection rate of 22 per 1,000 women under surveillance. Seven of the tumours were detected by CBE (seven per 1,000 women under surveillance), 11 by XRM (12 per 1,000 women under surveillance) and 19 by US (20 per 1,000 women under surveillance).

Sensitivity

The sensitivities presented in the paper are only calculated with the surveillance detected cancers as the denominator, they did not include the interval cancer. The documented figures are 33.3 per cent for CBE, 52.4 per cent for XRM and 90.4 per cent for US. There are no confidence intervals documented. If calculated with the interval tumour as a false negative, the respective results are 32 per cent (95% CI, 13.9 to 54.9%) and 50 per cent (95% CI, 28.2% to 71.8%) and 86.4 per cent (95% CI, 65.1 to 97.1%).

Specificity

It is unclear how the specificities have been calculated. The documented figures are 83.5 per cent for CBE, 82.2 per cent for XRM and 86.3 per cent for US. There are no confidence intervals documented.

Tumour characteristics

Sixteen tumours were invasive, two were DCIS, two were mucinous carcinomas and one was a medullary carcinoma. The mean tumour size was 12mm and seven were lymph node positive. These characteristics are not documented stratified by mode of detection.

Interval tumours

There was one interval tumour.

Survival

The five-year overall survival was 90.4 per cent and the disease-free survival rate was 80.9 per cent. This calculation was not adjusted for lead time bias or length bias.

In summary, this study suggested that US surveillance was more accurate and effective than CBE and XRM at detecting tumours in women at high risk of breast cancer in Taiwan. There were no measures of statistical significance documented in this study. It is noted that this may be related to Asian women generally having smaller denser breasts and that these findings may not be reproducible in a Western population.

Podo et al. (2002)

Study sample

This prospective cohort study recruited patients, both men and women, from nine genetics centres within Italy. At the time of publication, 105 women had been recruited and no men. Women were included if they were aged 25 years or over and men if they were 50 years of age or older. The mean age at recruitment was 46 years with a range of 25-77 years. Risk stratification was performed by criteria specific to this study. Participants had to be known BRCA 1 or BRCA2 mutation carriers or have a one in two probability of being a carrier, i.e. have a first-degree relative who was a proven mutation carrier. Two women were also included whose families had a very high incidence of breast cancer that was likely associated to a mutation other than BRCA 1 or BRCA2. Women with a personal history of breast cancer were included if it was unilateral (40 in total). They received unilateral screening if they had undergone mastectomy and bilateral if they had received breast conserving surgery. If women were on HRT, they were included after stopping treatment for three months. Exclusion criteria were: pregnancy, breast feeding, current chemotherapy, terminal illness and specific contraindications to MRI.

Interventions and comparators

Surveillance consisted of CBE, XRM, US and MRI at yearly intervals. The BIRADS system was used to classify the XRM, but the cut-off for an abnormal surveillance screen was not documented. It was not reported whether the radiologists interpreting the images were blinded to the results of the other imaging modalities. US was performed with a probe set at a frequency of $\geq$7.5MHz. The MRI results will be discussed in a subsequent chapter. The study reported on the preliminary phase of this research and therefore the follow-up was incomplete. Only 21 months of the study had been completed at the time of publication.

Outcomes

Cancer detection rate

Eight tumours were detected in total, seven in the prevalent round and one in the incident rounds. This gives a detection rate of 76 per 1,000 women under surveillance. Only one tumour was detected by XRM (9 per 1,000 women under surveillance) and similarly US only detected one tumour (9 per 1,000 women under surveillance). Five of the tumours were detected in patients with a previous personal history of breast cancer.

Accuracy measures

Accuracy measures were not calculated due to the incomplete follow-up.

Tumour characteristics

There were five invasive tumours, two DCIS and one combined DCIS and LCIS. The tumour size ranged from three to 27 mm and none had lymph node involvement. The tumour characteristics were not stratified by mode of detection.

Interval tumours

Interval tumours were not reported due to the incomplete follow-up.

In summary, this study shows a similar performance for XRM and US in the surveillance of this very high risk group, including mostly mutation carriers and a high proportion having a personal history of breast cancer. If MRI had not also been used in this study, the majority of tumours would have remained undetected. This suggests that in such a high risk group, surveillance by XRM and US is not adequate. There needs to be further results from this study to comment on the measures of accuracy, interval tumours and thereby the surveillance interval. No further reports from this group were identified in the literature search.

Crystal et al. (2003)

Study sample

This prospective cohort study recruited 1,517 women from dedicated mammographic units in Israel, of which 318 women had a high risk of breast cancer. There was no age restriction and the mean age at recruitment was 52.1 years, with a range of 31-84 years. Risk stratification was specific to the study. Women were considered to be at high risk if they had a first degree family history of breast cancer. Women with a personal history of breast cancer were included. Like the study by Kolb et al. (1998), Crystal et al. included women only if they had dense breasts, defined as BIRADS category 2-4, and had normal CBE and XRM. Therefore US was being examined as an adjunct test in this study also. It was also stipulated that all women be asymptomatic.

Interventions and comparators

Surveillance consisted of US in women with normal XRM and CBE. The frequency of the probe used for US was 5-12MHz. Comparisons of cancer detection were made between the high risk and average risk women. The classification system for images was not documented. The comparison of US and CBE and

XRM is implicit as those receiving US were known to have normal CBE and XRM. The results will over-estimate the performance of US as there are no results for tumours detected by XRM that would not be detected by US. The mean follow-up was not reported but the range was 8-30 months.

Outcomes

Cancer detection rate

Four tumours were detected in the high risk group, giving a rate of additional cancers detected of 12.6 per 1,000 women under surveillance. There was a significant difference between the cancer detection rates in the average risk and high-risk women ($p<0.04$).

Accuracy measures

Accuracy measures were not calculated in this study.

Tumour characteristics

There were three high-grade DCIS lesions and one intermediate grade ductal carcinoma. The tumours ranged in size from 4-12mm. One tumour had lymph-node spread, but it was not clear that this was in the cohort of high-risk women. The tumour characteristics were not stratified by the modality of surveillance.

Interval tumours

There had been no interval tumours detected at the time of publication.

In summary, this study suggested that in women with dense breasts there may be tumours identified by US, when used as an adjunct test, that were not identified by CBE or XRM. There appears to be significantly more tumours detected by the adjunct use of US in women at high risk of breast cancer than in women at normal risk. The results of this study are not generalisable to populations of women without normal XRM and CBE or with less dense breasts.

Trecate et al. (2003)

Study sample

This prospective cohort study recruited 23 women at high risk of breast cancer from the National Cancer Institute of Milan, Italy. There was no age restriction and no average age of the cohort was given. The age range was 30-61 years. Risk stratification was specific to this study. The women included were either BRCA1 or BRCA2 mutation carriers, had a one in two probability of being a carrier or over a 50 per cent risk of carrying a susceptibility gene for familial breast cancer based on family history. Women with a personal history of breast cancer were included (six women).

Interventions and comparators

Surveillance depended on the age group of the women. All ages had CBE every six months. Mammography was annual and commenced at 25 years with bilateral one view, and then increased to bilateral double view from 30 years and above. Annual US was performed alone from 20-25 years, then with XRM from 25-35 years, then six months after XRM from 35-40 years and above 40 years only if requested by the radiologist. The US was performed with either 7.5MHz or 10-12MHz probes (ATL HDI 3500, Philips). Annual MRI was performed for all ages for two years during the study. The results of MRI will be discussed in a subsequent chapter. The method of classifying the images was not documented. It was not reported whether the radiologists interpreting the images were blinded to the results of other modalities of surveillance. Follow-up was not documented. The study was conducted over a seven-month period but the dates were not given. It is unclear if this work may have been related to the study by Podo et al. (2002). Comparisons were made between the different modalities of surveillance.

Outcomes

Cancer detection rate

Four breast cancers were detected overall. This gives an overall detection rate (including MRI) of 170 per 1,000 women under surveillance). Three tumours were detectable by CBE (130 per 1,000 women under surveillance) but none of the tumours were detected by XRM or US.

Measures of accuracy

No measures of accuracy were calculated in this study.

Tumour characteristics

All four tumours were invasive. Only two tumours had the size recorded and these were 10mm and 30mm. No record of the lymph-node status was documented. There was no stratification of tumour characteristics by modality of surveillance.

Interval tumours

No interval tumours were documented.

In summary, this study suggests that US and XRM are not effective additions to surveillance for breast cancer in women at very high risk of breast cancer, i.e. mostly mutation carriers with a high proportion having a personal history of breast cancer. The results are extremely limited by the very small sample size and the lack of detail documented in the publication. The study focuses on very high risk women and may not be generalisable to all women at high risk of breast cancer.

Sim et al. (2004)

Study sample

This retrospective cohort study reanalysed data from a study by Stoutjesdijk et al. (2001). The original retrospective study examined the accuracy and efficacy of surveillance of women at high risk of breast cancer with XRM and MRI. The results of the original study are discussed in a subsequent chapter. Data were collected from a single university medical centre in the Netherlands. However, 84 women in the original study also underwent US examination. Sim et al. analysed the data on these women. There were no age restrictions specified. The mean age of the women who had US was not documented, but the mean age of the 42 women who underwent biopsy was 42.4 years, with a range of 25-58 years. Risk stratification was performed according to family history and utilising the Claus tables. The participants all had a 15 per cent or greater estimated lifetime risk of developing breast cancer. In the original study they were divided in to three groups; mutation carriers (50% to 85% risk), very high risk (30% to 50% lifetime risk) and high risk (15% to 30% lifetime risk). The other inclusion criteria were: having no personal history of breast cancer and having at least two years of histopathological details available post-screening. There were 66 women excluded in the original study cohort for not having sufficient information available.

Interventions and comparators

Surveillance in the original study was biannual CBE and annual XRM and MRI. The frequency of US scanning is not documented. BIRADS classification was used to classify the screening with scores of 3, 4 and 5 being referred for biopsy. The radiologist interpreting the US images was not blinded to the results of other modalities of screening and it is acknowledged that findings could be influenced by this. The follow-up was at least two years in every woman. Comparisons were made between surveillance by US and the results of the original study for XRM and MRI.

Outcomes

Cancer detection rate

A cancer detection rate of 0.24 was reported in the paper for US. There were not enough raw figures documented to verify this calculation. This compared to 0.17 with XRM and 0.28 with XRM and US combined.

Sensitivity

The sensitivity for US was 83.3 per cent, compared with 53.9 per cent for XRM and 92.9 per cent for XRM and US combined.

Specificity

The specificity for US was 65.5 per cent, compared with 85.7 per cent for XRM and 62.5 per cent for XRM and US combined.

PPV

The PPV for US was 50 per cent, compared with 63.6 per cent for XRM and 52 per cent for XRM and US combined.

NPV

The NPV for US was 90.5 per cent, compared with 80 per cent for XRM and 95.2 per cent for XRM and US combined.

AUC

The AUC for US was 0.712 (95% CI, 0.55 to 0.87), compared with 0.586 (95% CI, 0.40 to 0.77) for XRM and 0.761 (95% CI, 0.61 to 0.91) for XRM and US combined.

All of the above measures of accuracy are limited by the fact that there was not enough raw data documented to verify them or to calculate confidence intervals. Confidence intervals were only documented for the AUC.

Tumour characteristics

There were no documentation of tumour characteristics.

Interval tumours

There was no interval tumours documented.

In summary, this study suggests that there is a role for US in addition to mammography in the surveillance of women at high risk of breast cancer. The authors suggest that this combination, when compared to results of the original study (Stoutjesdijk et al. 2001), may be as efficacious as utilising MRI, but with a lower cost. It is suggested that there is also likely to be more widespread experience in US usage than MRI, which improves performance. However, the study is limited by the very small sample size and its retrospective design. It is also limited as it is extrapolating conclusions from the results of another study that was not employing US as a surveillance tool, but an additive examination used only selected women. This means that there will be selection bias in the women who underwent US examination and this is likely to have biased the results in favour of US. Consequently the results of this reanalysis of the work by Stoutjesdijk et al. must be interpreted with great caution.

Warner et al. (2004)

Study sample

This prospective cohort study recruited 236 female BRCA1 and BRCA2 mutation carriers from familial cancer centres in southern Ontario and Montreal in Canada. There were no age restrictions and the mean age at first surveillance was 46.6 years, with a range of 25-65 years). Risk stratification was performed by

all participants being BRCA1 or BRCA2 mutation carriers. This was therefore a very high risk group, 31 per cent of whom were of Ashkenazi Jewish descent. In addition, 30 per cent had a personal history of breast cancer. Exclusion criteria were: a past history of unilateral breast cancer if the contralateral breast was not intact, pregnant or lactating women, history of bilateral breast cancer currently undergoing chemotherapy or known to have metastatic disease and women weighing over 91kg (technical reasons). Thirty-one women left the study before completing three rounds, 16 underwent bilateral mastectomy, three were too large for MRI machine, three stopped due to pregnancy, four developed metastatic cancers, four were lost to follow-up and one did not wish to continue participating.

Interventions and comparators

Surveillance consisted of biannual CBE and annual XRM, US and MRI, all performed on the same day. Surveillance commenced at least one year after the woman's last mammogram. CBE was coded as normal, suggestive of benign disease, indeterminate, or suspicious of malignancy. Indeterminate CBE exams were repeated after three months. The MRI results will be discussed in a subsequent chapter. US was performed with a 7.5MHz transducer (the first seven patients did not receive US). All participants underwent the first screen, but only 58 per cent had the second and 36 per cent the third. BIRADS was used to classify the images and scores of 4 or 5 were biopsied. Each imaging study was read and scored independently by a radiologist experienced in breast imaging and radiologists were blinded to the results of CBE. All patients were followed up for a minimum of one year after their last screening examination. Comparisons were drawn between different modalities of surveillance.

Outcomes

Cancer detection rate

Twenty-two cancers were detected overall in 21 women. Seven of these women had a past history of breast cancer. This gives a cancer detection rate of 93 per 1,000 women under surveillance. Two were detected by CBE (8 per 1,000 women under surveillance), eight by XRM (34 per 1,000 women under surveillance) and seven by US (30 per 1,000 women under surveillance). Two tumours were detected by XRM alone and two by US alone.

Sensitivity

All the measures of accuracy in the paper are presented individually for each year of surveillance. These results have been combined to give overall results for the three rounds of surveillance. There was not enough raw data to calculate measures of accuracy for CBE.

The sensitivity of CBE, XRM and US respectively were 9 per cent (95% CI, 1% to 29%) 36% (95% CI, 17.1 to 59.3%) and 33 per cent (95% CI, 14.6 to 56.9%)

Specificity

The specificity of XRM and US respectively were 99 per cent (95% CI, 98.7 to 99.9%) and 96 per cent (95% CI, 93.7 to 97.7%).

PPV

The PPV of XRM and US respectively were 88 per cent (95% CI, 51.7 to 99.7%) and 29 per cent (95% CI, 12.6 to 51.1%).

NPV

The NPV of XRM and US respectively were 97 per cent (95% CI, 94.8 to 98.35) and 97 per cent (95% CI, 94.5 to 98.2%).

AUC

The AUCs for XRM and US respectively were 0.77 and 0.65. The AUC for CBE is also given at 0.48 and the combination strategy of CBE and XRM and US was 0.81. There were no confidence intervals documented for the AUCs.

Tumour characteristics

Sixteen tumours were invasive and six were DCIS. The mean size of the invasive tumours was 11mm at the first surveillance round and 13mm at the second round. Fifteen cases had lymph-node sampling and two were node-positive. The tumour characteristics are not documented stratified by modality of surveillance.

Interval tumours

There was one interval tumour, detected seven months after a third screen. Retrospectively this was visible on XRM at the last surveillance visit.

Mortality

All 22 patients with tumours were still alive and disease-free at the time of publication of the article.

In summary, this study suggests a similar efficacy and accuracy of XRM and US in the surveillance of high-risk women, although the PPV of US is much lower than that of XRM. The combined strategy of CBE with XRM and US gives the highest AUC. The results of this study are limited to the very high risk population of women who are proven BRCA1 or BRCA2 mutation carriers, including those with a personal history of breast cancer. It is therefore not generalisable to all women with an increased risk of breast cancer due to a family history. Further studies with larger numbers and longer follow-up, and including women of other risk groups are required.

Kuhl et al. (2005b)

Study sample

This prospective cohort study recruited 529 women from high-risk clinics in a single hospital in Germany. There was no age restriction and the mean age of the whole cohort was 41.7 years with a range 27-59 years. Risk stratification was performed according to the Consortium on Familial Breast and Ovarian Cancer of the German Cancer Aid. All participants had greater than a 20 per cent lifetime risk of breast cancer. In women that did not have a personal history of breast cancer, the Claus tables were also used to stratify risk. Women with a personal history of breast cancer were included provided the women had not had bilateral mastectomy, had not had chemotherapy within the last 12 months and had no metastases (139 women had a personal history). Another inclusion criterion was being asymptomatic.

Interventions and comparators

Surveillance consisted of biannual CBE and US and annual XRM and MRI. If abnormalities found on CBE or US at the round without XRM or MRI, these additional modalities were used to further investigate this. Surveillance commenced at 30 years, or five years before the youngest family member affected with the disease. (NB: in the first two years, women under 30, or 30-39years with dense breasts did not receive XRM, but this was subsequently abandoned and all women received XRM (these data were not included in the calculation of accuracy measures). MRI of both entire breasts was performed on a 1.5T system (NT/INTERA; Philips, Best, the Netherlands). US was performed with 7.5-13MHz probes. Each imaging study was read and scored independently by a different radiologist who had substantial experience with the respective imaging technique. The radiologists were informed about the clinical findings from CBE and the risk status of the patient but were blinded to the results of the respective other imaging modalities. BIRADS was used to classify the images and scores of 4 or 5 went for biopsy. The mean follow-up time was 5.3 years, with a range of 2-7 years. The number of total annual surveillance rounds for which data on all three imaging modalities was available was 1,452, and this was used in the calculation of accuracy measures. Comparisons are made between the three risk groups and the different modalities of surveillance.

Outcomes

Cancer detection rate

A total of 43 tumours arose in 41 patients during the study period. It is documented that 40 of these were detected by imaging. That gives a cancer detection rate for the overall surveillance strategy of 76 per 1,000

women under surveillance. Eleven (25%) of these patients had a prior history of breast cancer. CBE identified only one tumour (2 per 1,000 women under surveillance) which was also detected on imaging. Fourteen tumours were detected by XRM (26 per 1,000 women under surveillance) and only 1 was diagnosed by XRM that was not diagnosed by MRI. Seventeen tumours were detected by US (32 per 1,000 women under surveillance), two of these were at the half-yearly CBE and US screen and were not palpable. Twenty-one tumours were detected by US and XRM combined (40 per 1,000 women under surveillance).

Sensitivity

The overall sensitivity for CBE was 2.3 per cent (0.1 to 12.3%), for XRM was 32.6 per cent (95% CI, 19.0 to 48.5%), for US was 39.5 per cent (95% CI, 25.0 to 55.6%) and for XRM and US combined was 48.8 per cent (95% CI, 33.3 to 64.5%).

Overall, there was no significant difference in sensitivity between XRM and US ($p< 0.05$).

When stratified by risk groups XRM, US and the combination of XRM+US all became less sensitive as the lifetime risk of breast cancer increased. The sensitivities for XRM+US were 83.3 per cent for women with a 20 per cent lifetime risk, 45 per cent for women with a 21-40 per cent lifetime risk and 37.5 per cent for the mutation carrier group.

Specificity

The overall specificity for XRM was 96.8 per cent (95% CI, 95.7 to 97.7%), for US was 90.5 per cent (95% CI, 88.8 to 92.0%) and for XRM and US combined was 89.0 per cent (95% CI, 87.2 to 90.6%).

Stratification by risk group or by a past history of breast cancer does not appear to affect the specificity

PPV

The overall PPV for XRM was 23.7 per cent (95% CI, 1 to 29%), for US was 11.3 per cent (95% CI, 6.7 to 17.4%) and for XRM and US combined was 11.9 per cent (95% CI, 7.5 to 17.6%).

Overall, the PPV was significantly higher for XRM when compared to US or US and XRM combined ($p=0.02$).

Stratification by risk group or by a past history of breast cancer does not appear to affect the PPV either

Tumour characteristics

Thirty-four tumours were invasive and nine were DCIS. Of the 21 cancers detected by XRM and US, 16 were invasive and the rest were DCIS. The invasive cancers had a mean size of 13.9mm and five were node- positive. The tumour characteristics were stratified by modality of screening. There were no significant differences in the characteristics of the tumours detected by XRM or US (*p* values all > 0.05).

Interval tumours

The interval tumour rate is given as 2 per cent in this cohort. It is unclear if this is a percentage of the women under surveillance or of the tumours that arose. It was also documented that there was one interval cancer that arose between surveillance rounds. However, it was reported that 40 of the 43 cancers were detectable by imaging, which would suggest three interval cancers. These figures were reported in an unclear manner.

In summary, this study suggests that the addition of US to XRM does not significantly improve the sensitivity of surveillance of women at high risk of breast cancer and does significantly reduce the specificity and PPV. However, it is difficult to compare US with XRM in this study as they were carried out at different intervals, US biannually and XRM annually. Therefore it is not a direct comparison. The data on interval tumours is somewhat unclear in its documentation. This study included women at high risk who had a personal history of breast cancer, but the majority of the results were not significantly different if stratified by personal history.

Summary

There were nine studies identified of relevance to the accuracy and efficacy of surveillance of women at high risk of breast cancer with XRM and US. Eight of the studies were prospective cohorts and one was a retrospective cohort study (Sim et al. 2004). The retrospective study was limited by its design and by the fact that it re-analysed data from a previous study in which the use of US may have been more diagnostic than surveillant. A total of 2,944 women received surveillance in the nine studies. There was heterogeneity between the studies in terms of risk status of the surveillance strategies, surveillance intervals, the participants' risk status and age. Three studies recruited women that were either known BRCA1 or BRCA2 mutation carriers or had a 50 per cent chance of being a mutation carrier (Podo et al. 2002; Trecate et al. 2003; Warner et al. 2004). Kolb et al. (1998) and Crystal et al. (2003) recruited participants with a first degree family history of breast cancer and Hou et al. (2002) included those with a first or second degree family history. O'Driscoll et al. (2001) recruited women with a moderate risk of breast cancer, three times higher than the age-matched general population but excluding those at highest risk who were involved in other trials. Kuhl et al. (2005b) required at least a 20 per cent lifetime risk of breast cancer and the retrospective study (Sim et al. 2004) stipulated at least a 15 per cent lifetime risk of breast cancer.

The surveillance included XRM and US in all the studies. Four studies also had MRI in their strategies and the results of MRI will be discussed in subsequent chapters (Kuhl et al. 2005b; Podo et al. 2002; Trecate et al. 2003; Warner et al. 2004). Two studies examined US as an adjunct test in women with normal CBE and XRM results (Crystal et al. 2003; Kolb et al. 1998). This would overestimate the effect of US in comparison to XRM, as it does not include the tumours that would have been detectable by XRM but not picked up on US if the tests had been performed at the same time. These studies also only recruited women with dense breasts. This would also overestimate the efficacy of US compared to XRM as there would be more false negatives on XRM in women with dense breasts. Seven studies used US probes with frequencies of 7.5 MHz or over. One study (Sim et al. 2004) did not document the frequency of the probe used. Crystal et al. (2003) documented the probe frequency as between five to 12 MHZ, so it was unclear exactly what frequency was used.

The outcomes in these studies were the intermediate outcome measures of cancer detection rate and measures of accuracy. The tumour characteristics were documented in eight of the nine studies. Only one study (Kuhl et al. 2005b) examined the tumour characteristics by modality of surveillance and found no significant differences. None of the studies compared tumour characteristics in the surveillance population to a population not receiving surveillance. Interval tumours were documented, but only provide information for the surveillance strategies as a whole. Survival was calculated by Hou et al. (2002) but not adjusted for lead time. Mortality was documented by Warner et al. (2004) but there was not sufficient follow-up for this to be meaningful.

The results of the cancer detection rates and measures of accuracy are summarised in **Tables 15** and **16** for surveillance with US, XRM and the combination of XRM and US.

The cancer detection rates are similar between XRM and US surveillance and the combination of XRM and US appears to offer a slightly higher cancer detection rate. The study by Hou et al. (2002) did not have measures of statistical significance calculated in the study. However, the participants in this study were likely to have a reduced sensitivity for XRM as they generally had smaller denser breasts. The results of this study are not generalisable to other populations. The two studies which give measures of accuracy and have the raw data to calculate statistical significance (Warner et al. 2004 and Kuhl et al. 2005b) demonstrate that there is no significant difference between the sensitivity and NPV of XRM and US surveillance in women at high risk of breast cancer. As discussed in previous chapters, the sensitivity of XRM and US decrease as the risk of breast cancer increases, being especially low in mutation carriers (Kuhl et al. 2005b). This decreasing sensitivity is also demonstrated for the combination of XRM and US in the surveillance of women at high risk of breast cancer (Kuhl et al. 2005b). There is a significant difference in the specificity and PPV, with lower values for US than XRM due to the number of false positive examinations created by US surveillance.

Table 15. Cancer detection rates in surveillance of women at high risk of breast cancer with XRM and US

Study	Cancer Detection Rate Overall	Cancer detection rate by modality		
		US	XRM	US + XRM
Kolb et al. (1998)	N/A as US used as adjunct test	1.8 per 1,000 w/s	All had normal XRM	N/R
O'Driscoll et al. (2001)	6.7 per 1,000 w/s	6.7 per 1,000 w/s	0	N/R
Hou et al. (2002)	22,0 per 1,000 w/s	20.0 per 1,000 w/s	12.0 per 1,000w/s	N/R
Podo et al. (2002)	76.0 per 1,000 w/s (includes MRI)	9.0 per 1,000 w/s	9.0 per 1,000 w/s	N/R
Crystal et al. (2003)	N/A as US used as adjunct test	12.6 per 1,000 w/s	All had normal XRM	N/R
Trecate et al. (2003)	170 per 1,000 w/s (includes MRI)	0	0	N/R
Sim et al. (2004)	N/A as only studied US	Not reported per women under surveillance		
Warner et al. (2004)	93 per 1,000 w/s (includes MRI)	30.0 per 1,000 w/s	34.0 per 1,000 w/s	N/R
Kuhl et al. (2005b)	76 per 1,000 w/s (includes MRI)	32.0 per 1,000 w/s	26.0 per 1,000 w/s	40.0 per 1,000 w/s

w/s = women under surveillance N/A = not applicable N/R = not reported
The cancer detection rates reported by women under surveillance cannot be compared across studies due to the differing surveillance intervals and differing length of the studies.

Table 16. Measures of accuracy in surveillance of women at high risk of breast cancer with XRM and US

Study	Accuracy	US (95% CI)	XRM	US +XRM	P values
Kolb et al. (1998)	No accuracy measures				
O'Driscoll et al. (2001)	No accuracy measures				
Hou et al. (2002)	Sensitivity Specificity	86% (65.1- 97.1%) 99.4% (98.7 - 99.8%)	50% (28.2-71.8%) 99.5% (98.9-99.95)	N/R	No measures of accuracy were documented in this paper.
Podo et al. (2002)	No accuracy measures				
Crystal et al. (2003)	No accuracy measures				
Trecate et al. (2003)	No accuracy measures				
Sim et al. (2004)	Sensitivity Specificity PPV NPV AUC	83.3% 65.5% 50% 90.5% 0.712 (0.55-0.87)	53.3% 85.7% 63.6% 80.0% 0.586 (0.4-0.77)	92.9% 62.5% 52.0% 95.2% 0.761 (0.61-0.91)	No raw data to enable calculation of CIs or *p* values.
Warner et al. (2004)	Sensitivity Specificity PPV NPV AUC	33% (14.6-66.9%) 96% (93.7-97.7%) 29% (12.6-51.15) 97% (94.5-98.2%) 0.65	36.3% (17.1-59.3%) 99.8% (98.7-99.9%) 88.9% (51.7-99.7%) 96.9% (94.8-98.3%) 0.77	64% (no CI reported) N/R N/R N/R 0.81	= 0.91 < 0.01 < 0.01 =0.89
Kuhl et al. (2005b)	Sensitivity Specificity PPV NPV	39.5% (25.0-55.6%) 90.5% (88.8-92.0%) 11.3% (6.7-17.4%) 98.0% (97.1-98.7%)	32.6% (19.0-48.5%) 96.8% (95.7-97.7%) 23.7% (14% to 37%) 97.9% (97.0-98.6%)	48.8% (33.3-64.5%) 89.0% (87.2-90.6%) 11.9% (6.7-17.4%) 98.0% (97.1-98.7%)	US versus XRM = 0.6 < 0.001 = 0.03 = 0.98

In conclusion, surveillance in women at high risk of breast cancer with XRM or US has equivalent sensitivity and NPV, but US has a lower specificity and PPV. The combination of XRM and US has a better sensitivity than either modality alone, yet retains the poorer specificity and PPV of US. This is due to the number of false positives generated by US. US has the advantage of not using ionising radiation for surveillance and being a functional tool for biopsy. However, the number of false positives generated is a disadvantage as it would lead to anxiety and a higher rate of invasive investigations. Due to this, US may remain more diagnostic and other modalities of surveillance may be required in women at high risk of breast cancer. The following chapters examine the role of MRI in this surveillance.

Table 17. Primary research studies appraised investigating the accuracy and efficacy of ultrasound surveillance compared to mammography on outcomes from breast cancer

Source	Study design Evidence grading	Comparison interventions and dates of testing	Sample	Outcomes and verification	Results	Comments
Kolb et al. (1998) USA	Prospective cohort study III-2 (CX P1 Q2)	Ultrasound examination (after normal mammography and CBE) US bandwidth was 5-10 MHz, set at 7.5MHz. Dates of Surveillance were Jan 1995 to April1997. Comparison was made of cancer detection between women at high risk and women at average risk in the study. There is an implicit comparison of XRM and US as all participants had already had normal XRM.	Sample = 3,626 patients, but only 565 were high-risk women due to a family history of breast cancer. Recruited from a single medical centre. Mean age in the overall study was 59.3 years but the mean age is not given specifically for the high-risk group. Inclusion criteria • normal mammographic and CBE exams; • dense breasts (BIRADS 2-4). Risk stratification strategy not given but says that all women in the high-risk group had a primary family member (mother, daughter sister or brother) with breast cancer.	Relevant outcomes • cancer detection. Verification of positive result was through pathology. Verification of negative results was through follow-up, although this is not explicit.	Cancer detection: 1 tumour was detected among these high-risk women. Comparison was made between the cancer detection rate in women at high risk and those at average risk. There was a second group of high-risk women whose risk was related to previous breast cancer or breast abnormalities (n=487). Even with these 2 groups combined there was not a statistically significant difference between the cancer detection rate and that of women at average risk (p=0.9).	Limitations include: Verification bias is likely Selection bias may be present as there is little information given specifically about the women at high risk, including little information about the exact risk stratification strategy used. Authors' conclusions: The general findings were that US can depict small, early-stage otherwise occult cancers similar in size and stage to mammographically identified non-palpable cancers and smaller and lower in stage than palpable cancers in dense breasts. In regards to the high-risk group, there was no statistically significant difference found between the cancer detection rate in women at high risk and women at average risk. Therefore, they concluded that there was no reason why normal risk women with dense breasts should be excluded from such surveillance. It was emphasised that these findings did not translate to a reduced mortality in women who received surveillance and that this remained to be proved. Reviewers' conclusions: This study's main focus is on women with dense breasts and the component looking at women at high risk is small.

Table 17. Primary research studies appraised investigating the accuracy and efficacy of ultrasound surveillance compared to mammography on outcomes from breast cancer (continued)

Source	Study design Evidence grading	Comparison interventions and dates of testing	Sample	Outcomes and verification	Results	Comments
Kolb et al. (1998) USA *Continued*						There was no significant difference found in the cancer detection rate between the high-risk and average risk groups. However, the findings do suggest an increased efficacy overall in using US (in addition to CBE and mammography) in women with dense breasts. This finding may be pertinent to women at a high risk as they require surveillance from an earlier age, when breasts tend to be denser.

Table 17. Primary research studies appraised investigating the accuracy and efficacy of ultrasound surveillance compared to mammography on outcomes from breast cancer (continued)

Source	Study design Evidence grading	Comparison interventions and dates of testing	Sample	Outcomes and verification	Results	Comments
O'Driscoll et al. (2001) UK	Prospective cohort study III-2 (C1 P1 Q2)	Surveillance protocol involved mammography and bilateral breast ultrasound. All women also instructed on BSE. The radiologist performing the ultrasound was blinded to the mammography report. A second radiologist read the mammography and was blinded to the ultrasound findings (61% of the patients had previous mammographic films available for comparison). First 29 USs were done with a 7.5 MHz probe, but the rest (120) were done with an 8-12 MHz probe.	Sample no = 149 women at moderate risk of breast cancer. Mean age on entering surveillance = 42.15 yrs (range 30-69 yrs) Recruited from a clinical genetics department Inclusion criteria • Women with a family history that placed them at a moderately increased risk of breast cancer. Exclusion criteria • Women at high risk were not approached as they were already involved in several studies. Risk stratified by criteria reported in full in the paper. Moderate risk was defined as a risk 3 times that of the aged matched general population.	Relevant outcomes Cancer detection Interval cancers Verification of positive tests was with biopsy and pathology results Verification of negative tests was with follow-up Mean follow-up time = 13.7 months	Cancer detection. Thirteen patients were identified to have one focal solid lesion on US that warranted biopsy. After considering the mammography and US results together, plus the fact that some lesions had been previously examined, there were 10 lesions recommended for biopsy. Nine of these were recommended on US criteria alone and one on both mammographic and US criteria. Out of the 10 biopsies, 1 adenoid cystic carcinoma was identified. Interval cancer: There was one interval cancer – 10.5 months after a negative mammogram and ultrasound examination. Pathology was an invasive lobular carcinoma with extensive lobular carcinoma *in situ*. Lymph node negative.	Limitations include: Small pilot study and small number of tumours. Other than risk stratification was unclear on characteristics of women in cohort, also of how they were selected and if there were differences between them and women selected who chose not to participate. No information on potential confounders, such as HRT or OCP use. One radiologist performed most of US examinations; results may vary with the skill of the radiologist and this may affect the reproducibility of the study results Could not differentiate in the results which were prevalent and which were incident rounds. Could not identify from results which results used the lower frequency probe and which used the higher frequency probe. Likely verification bias. Authors' conclusions: We found that surveillance for breast cancer with US in a cohort of women at moderately increased risk of breast cancer, does not lead to an unacceptably high biopsy rate. These findings indicate that surveillance with US and mammography in patients at increased risk of breast cancer may be beneficial and a larger randomised study to examine issues of acceptability, reproducibility and cost effectiveness is timely.

Table 17. Primary research studies appraised investigating the accuracy and efficacy of ultrasound surveillance compared to mammography on outcomes from breast cancer (continued)

Source	Study design Evidence grading	Comparison interventions and dates of testing	Sample	Outcomes and Verification	Results	Comments
O'Driscoll et al. (2001) UK *Continued*						Reviewers' conclusions This was a fairly small pilot study and must be interpreted with the limitations above. The design was good in terms of blinded interpretation of the results. The results do suggest that US may detect tumours not detected by mammography in high risk women. The results may be underestimated due to the use of a low frequency probe in a proportion of the examinations. As acknowledged by the authors, a larger study would be needed to confirm these results and to assess the degree of false positive results created by adding US to a breast cancer surveillance programme for women at high risk of breast cancer.

Table 17. Primary research studies appraised investigating the accuracy and efficacy of ultrasound surveillance compared to mammography on outcomes from breast cancer (continued)

Source	Study design Evidence grading	Comparison interventions and dates of testing	Sample	Outcomes and verification	Results	Comments
Hou et al. 2002 Taiwan	Prospective cohort Study III-2 (C1 P1 Q2)	Surveillance protocol was annual CBE, mammography and US. Pre-menopausal women received surveillance during the 2nd week of the menstrual cycle to minimise the occurrence of breast densities or enhancing masses related to the menstrual cycle. 4-view film mammograms were conducted and reviewed by one radiologist. US performed with a 7.5MHz frequency transducer probe. Dates: May 1994 to August 2001. No comparisons were made in this study other than between modalities of surveillance. BIRADS was used to classify screens and a cut-off of 4 and above was employed.	Sample no = 935 women. Mean age at surveillance = 48.6 years (range 35-75). Recruited Inclusion criteria: • >35 years old. • female relatives of breast cancer patients (mothers, daughters, grandmothers, sisters). Exclusion criteria: • pregnant or lactating. • past history of breast cancer. • known metastatic diseases No specific risk stratification process carried out. Just all relatives of breast cancer patients.	Relevant outcomes: • cancer detection rate; • mode of detection; • tumour stage; • node status; • interval tumours; • 5-year overall survival and event-free survival (free from cancer related death and tumour spread); • Sensitivity; • Specificity. Verification of positive result, by any of 3 modalities of surveillance was through biopsy and pathology results. Verification of a negative result was through follow-up. Median follow-up was 41.8 months (range 12-82 months). Verification of interval cancers.	Cancer detection rate: 21 cancers were detected, giving an overall cancer detection rate of 22 per 1,000 women under surveillance. Of the women with tumours, 1 was a BRCA1 mutation carrier, 2 were BRCA2 mutation carriers and the other 18 were mutation status unknown. Mode of detection: CBE detected 7 tumours. Mammography detected 11 tumours. US detected 19 tumours. Tumour size, stage and node status: 16 were invasive cancers, 2 were DCIS, 2 were mucinous carcinomas and 1 was a medullary carcinoma. Mean tumour size was 12mm. 7 were node-positive and 14 were node-negative. 1 interval cancer was reported Five-year overall survival was 90.4% and the disease-free survival rate was 80.9%.	Limitations included: Verification bias is likely. Lead time bias and length bias are likely in terms of the survival data. This population was not explicitly risk stratified and it is difficult to assess their overall risk of breast cancer. There are no characteristics of the overall group of women screened, other than being relatives of breast cancer patients and the mean age. It is unclear if they have any additional risks for breast cancer. Only a prevalent round was examined and it is likely that the cancer detection rate would be higher in this round than in subsequent rounds, There is no mention of how interval cancers are verified as being true interval cancers. Authors' conclusions: Based on a higher sensitivity of sonography for detecting breast cancer in the high-risk group in our study, sonography is superior to mammography and physical examination of the breasts in the surveillance of women at high risk for breast cancer in Taiwan. If sonography will replace mammography as a surveillance tool, needs further research.

Table 17. Primary research studies appraised investigating the accuracy and efficacy of ultrasound surveillance compared to mammography on outcomes from breast cancer (continued)

Source	Study design Evidence grading	Comparison interventions and dates of testing	Sample	Outcomes and verification	Results	Comments
Hou et al. 2002 Taiwan *Continued*					The documented sensitivities were: CBE 33.3% Mammography 52.4% US 90.4% The documented specificities were: CBE 83.5% Mammography 82.2% US 86.3% The measures of accuracy have been recalculated with the interval tumour in the denominator and are: Sensitivity: CBE 31.8% (95% CI, 13.9-54.9%) XRM 50% (95% CI, 28.2 - 71.8%) US 86.4% (95% CI, 65.1-97.1%) Specificity: CBE 99.4% (95% CI, 98.7-99.8%) XRM 99.6% (95% CI, 98.9-99.9%) US 99.4% (95% CI, 98.7-99.8%) The figures for specificity do not agree with those in the paper and it is not evident how they were calculated. There were no measures of statistical significance documented in this study.	Otherwise the low cost of US and convenience for women who live in rural areas suggests that sonography will be a useful tool for breast cancer surveillance in Taiwanese women in the high-risk group and in countries with a low incidence of breast cancer. Reviewers' conclusions: This study suggests that sonography is much more accurate for the surveillance of women at high risk of breast cancer, than mammography or CBE. However, these findings are specific to this population and are not generalisable. As discussed by the authors, the sensitivity of mammography is likely reduced by the higher proportion of Asian women with smaller denser breasts, which are less fatty, and also the overall lower incidence if breast cancer in this Taiwanese population. Sonography may be a useful modality of surveillance in these women, and especially in rural areas or areas without access to MRI, however it is unlikely to achieve such good results in a Western population.

Table 17. Primary research studies appraised investigating the accuracy and efficacy of ultrasound surveillance compared to mammography on outcomes from breast cancer (continued)

Source	Study design Evidence grading	Comparison interventions and dates of testing	Sample	Outcomes and Verification	Results	Comments
Podo et al. (2002) Italian multi-centre study	Prospective cohort Study III-2 (C1 P2 Q2)	Surveillance protocol: CBE, mammography, US and MRI at yearly intervals. Mammography: standard mediolateral oblique and cranio-caudal views were obtained of each breast. Further views taken when necessary. Findings reported using the BIRADS system (1, negative; 2, benign; 3, probably benign; 4, suspicious abnormality and 5, highly suggestive of malignancy). US performed at a frequency of >7.5MHz. MRI was performed on coronal and axial planes. One pre-contrast and 5 post-contrast images were taken. Gd-chelate (0.1 mmol/kg) was injected as contrast. MRI was reported using a system that is based on a combination of morphological and enhancement parameters. (0-2 = benign, 3=uncertain, 4-8=malignancy).	Sample no = 105 patients were enrolled in the first annual round (14 of these women also underwent a second round). Forty (38%) had a previous personal history of breast cancer. Mean age at recruitment 46 years, median age 51years (age range 25-77years). Mean age at diagnosis was 55.3 years, median 52.5 years (range 35-70 years). Recruited from 9 cancer genetics centres within Italy. Inclusion criteria: • very high risk of breast cancer; • women >25 years age • men >50 years age	Relevant outcomes: • Cancer detection rate • Mode of detection • Tumour size, stage and node status Verification of positive findings is by biopsy (either MRI or US guided) and pathology. Verification of negative findings is through follow-up – it is acknowledged that these are preliminary findings and the follow-up is incomplete.	Cancer detection rate: 8 tumours were detected in total, 7 in the prevalent screen and 1 in the incident screen. 5 of these patients had a previous personal history of breast cancer, 3 were BRCA1 mutation carriers, 3 were BRCA2 mutation carriers and 2 with unknown mutation status. Mode of detection: Both mammography and US detected only 1 tumour. Tumour size, stage and node status: 2 invasive ductal carcinomas 2 invasive lobular carcinomas 1 invasive ductal and lobular carcinoma 2 DCIS 1 DCIS and LCIS Tumour size ranged from 3 to 27mm. There were no node-positive tumours. The follow-up is incomplete and therefore sensitivity and specificity cannot be calculated.	Limitations included: Only the preliminary report of this study. Verification bias particularly in this study (as acknowledged by the authors) as it is just a preliminary report and sufficient follow-up of negative results has not yet been achieved. This cohort varies from other studies as it is a very high risk group and includes a high proportion of women with a personal history of breast cancer. No comment on women undertaking risk reducing strategies such as on Tamoxifen or having had a bilateral salpingo-oophorectomy. Authors' conclusions: The authors' conclusions relate to the overall surveillance strategy, including MRI. The findings of this study substantiate those of existing studies, that MRI is a more sensitive and more accurate imaging modality than conventional imaging for detecting breast cancer in women at a high risk of this disease (both pre- and post-menopausal women). A previous personal history of breast cancer was associated with higher probability of breast cancer detection during surveillance. .

Table 17. Primary research studies appraised investigating the accuracy and efficacy of ultrasound surveillance compared to mammography on outcomes from breast cancer (continued)

Source	Study design Evidence grading	Comparison interventions and dates of testing	Sample	Outcomes and Verification	Results	Comments
Podo et al. (2002) Italian multi-centre study. *Continued*		In the case of non-benign scores (3-8) which were detected only by MRI, the MRI was repeated after 1-2 months. If the lesion was confirmed then a biopsy was undertaken. Pre-menopausal women had MRI within the 2nd week of the menstrual cycle. Dates of surveillance: June 2000 to March 2002 (preliminary report of first phase, 21 months, of the study).	• women who had personal history of breast cancer were allowed if unilateral. Unilateral mammography done if had had a mastectomy and bilateral if had had breast conservation. • if on HRT, were included but this was stopped and surveillance not started until been off it for 3 month Exclusion criteria: • pregnancy; • breast feeding; • current chemotherapy; • terminal illness; • specific contra-indications to MRI. Risk stratification: Only recruited subjects who were known BRCA1 or BRCA2 mutation carriers, or had a 1 in 2 probability of being a carrier (first-degree relative who was a proven mutation carrier).			The authors conducted a review of other existing literature and perform a meta-analysis of the results of the studies to date. They note that there are considerable differences in the design of these studies, but state that there are some consistent conclusions. The overarching finding is that MRI is more sensitive and significantly more accurate than conventional imaging in the surveillance of women at a high risk of breast cancer They point to the need for more extensive, multi-centre and multi-national trials on the evaluation of benefits and costs associated with the introduction of MRI into appropriate surveillance programmes specifically addressed to subjects at high genetic risk of Breast cancer. Reviewers' conclusions This study appears to show that XRM and US perform equivalently in the surveillance of women at high risk of breast cancer. However, these are only preliminary results of this study and measures of accuracy could not be calculated without further follow-up data. Unfortunately a further report of this work cannot be found and it is perhaps ongoing. These results are also limited in their external validity by being from a very high risk cohort, especially as a high proportion of women with a personal history of breast cancer were included.

Table 17. Primary research studies appraised investigating the accuracy and efficacy of ultrasound surveillance compared to mammography on outcomes from breast cancer (continued)

Source	Study design Evidence grading	Comparison interventions and dates of testing	Sample	Outcomes and Verification	Results	Comments
Podo et al. (2002) Italian multi-centre study. *Continued*			2 women also included whose families had a very high risk or incidence of breast cancer that was likely associated to a non BRCA1 or BRCA2 mutation. 40 of the 105 women also had a personal history of breast cancer.			The pulling together of results from other studies was hampered by variation in the design of the studies and also the outcomes measured.

Table 17. Primary research studies appraised investigating the accuracy and efficacy of ultrasound surveillance compared to mammography on outcomes from breast cancer (continued)

Source	Study design Evidence grading	Comparison interventions and dates of testing	Sample	Outcomes and verification	Results	Comments
Crystal et al. (2003) Israel	Prospective cohort study III-2 (CX P2 Q2)	Surveillance protocol: Mammography and CBE, with US as an adjunct test. All mammography and sonography was performed by radiologists experienced in these fields. US was performed with a bandwidth of 5-12MHz In 86.6% of the 1,517 sonography examinations the radiologist performing the examination had also reviewed the mammogram. In the 204 other cases, the mammogram was reviewed by the radiologist prior to sonography (so none were blinded to Mammography results). Women were examined clinically prior to sonography by the examining radiologist	Sample no = 1,517 women altogether with 318 women at high risk of breast cancer Mean age = 52.1years (range 31-84 years) Recruited from dedicated mammographic units in Israel. Inclusion criteria • Asymptomatic; • dense breasts (BIRADS density categories 2, 3 or 4); • normal mammography and CBE findings; • BIRADS (for density). In the usual risk women, 7% were BIRADS 2. 78.9% were BIRADS 3 and 14% were BIRADS 4. Of the high-risk women 22.6%, 63.5% and 13.8% were BIRADS 2, 3 and 4 respectively.	Relevant outcomes • Cancer detection rate • Tumour stage, size and node status • Interval cancers Verification of positive results was via biopsy Verification of negative results was via follow-up. Mean follow-up was not reported but the range was 8 – 30 months.	Cancer detection rate: 90 women (5.9%) overall were found to have complex cysts or solid lesions on US. Biopsies were done for 38 of these and the other lesions were to be followed up with further US examinations. Of the latter, 55 patients had stable appearances at one year follow-up and 7 women had only had 6 month follow-up, but this was also stable in them all. 15 of the biopsies were done in baseline risk women and 5 were done in high risk women. There were 3 carcinomas detected out of the 15 biopsies in average risk women, and 4 carcinomas detected out of the 5 biopsies done in high-risk women. Therefore the cancer detection rates were 0.245 for average risk women and 1.26 for high risk women. This was a significant difference (p = <0.04) The overall cancer detection rate of 0.46% was reportedly only slightly lower than screening mammographic results in centres of excellence and those cited in peer-reviewed literature. .	Limitations included: Few characteristics reported of women involved, except risk stratification, and no details of those who did not participate. High-risk women were both those with a family history and those with a personal history of breast cancer. Although the latter women only received sonography of the unaffected breast, this is still a different risk categorisation from the majority of studies in this field. The bandwidth of the probe is wide and therefore it is unclear how this fits with our inclusion criteria. The exclusion of women with lower BIRADS categories means that this is a subset of women of high risk and results are not applicable to this population as a whole. In addition, US was only performed as an adjunct test if CBE and mammography was normal; this is again a different subgroup of the population as a whole. Verification bias likely. Authors' conclusions: We show that screening sonography in cases of mammographically dense breast tissue permits the effective detection of otherwise occult small breast cancers. Our results particularly point to a potential benefit in high-risk women.

Table 17. Primary research studies appraised investigating the accuracy and efficacy of ultrasound surveillance compared to mammography on outcomes from breast cancer (continued)

Source	Study design Evidence grading	Comparison interventions and dates of testing	Sample	Outcomes and verification	Results	Comments
Crystal et al. (2003) Israel *Continued*		. If a palpable lesion was detected they were excluded from the study. In addition, in lesions that were detected by US alone, the case was reviewed by a breast surgeon who re-examined the woman and if the lesion was felt to be palpable by the surgeon then the women was excluded (1 case in this study). Dates of study from Jan 2000 to Jan 2002.	Risk stratification was done simply, with women considered to be at high risk if they had a first-degree family history of breast cancer. However, it also included women with a personal history of breast cancer. Comparisons were made between women at baseline risk and women at high risk. Cancer detection rates were also compared with rates considered to be adequate in mammographic screening programmes in centres of excellence.		Tumour stage, size and node status: The tumours in the baseline risk women were sized from 10-12mm and were a lobular carcinoma, a low grade ductal carcinoma and a high grade ductal carcinoma. The tumours in the high-risk women were sized from 4-12mm. There were 3 high-grade ductal carcinomas and 1 intermediate grade ductal carcinoma. Only 1 tumour was lymph-node positive, Unfortunately it is not explicitly reported whether this was among the baseline or high risk women, although it appears to have been in the high risk group. Of the 7 cancers detected, 2 were in BIRADS category 4, 5 were in BIRADS 3 and no none were detected in the BIRADS 2 category. There was no significant difference found between biopsy rate and cancer detection rate between different radiologists. Interval cancers: No interval cancers had been detected at the date of publication; the follow-up time to that point ranged from 8-30 months.	In this group, the 1.3% detection rate was significantly higher than that in women at baseline risk, and it was also higher than the acceptable detection rate for screening mammography. Additional studies to examine issues of reproducibility and cost effectiveness are needed. We therefore recommend the implementation of sonography for breast cancer surveillance in high-risk women with mammographically dense breast tissue. Reviewers' conclusions: This study is, on the whole, well designed; however it does not translate directly to the population in question in this review. The population here all have negative mammography and CBE, are at high risk, not just from family history but from personal history, and have high breast density. This will all affect the accuracy and effectiveness of the test and the results are not therefore applicable to populations other than one selected in a similar manner. The suggestion of these results would be that ultrasonography is an efficacious addition to surveillance in women with normal mammograms who are at high risk of breast cancer and who have dense breasts (BIRADS 2 to 4).

Table 17. Primary research studies appraised investigating the efficacy and accuracy of ultrasound surveillance compared to mammography on outcomes from breast cancer (continued)

Source	Study design Evidence grading	Comparison interventions and dates of testing	Sample	Outcomes and verification	Results	Comments
Trecate et al. (2003) Italy (NB: Podo is an author on this one as well but we cannot find any further reports from the Podo et al. trial.)	Prospective cohort study III-2 (C1 P2 Q3)	Surveillance protocol: Outlined in full in the paper and was dependent on age group, CBE was performed every 6 months for all ages. Mammography was annual and commenced at 25 years with bilateral one-view, and then increased to bilateral double-view from 30 years and above. Double-view was performed in craniocaudal and mediolateral oblique projections. One-view was performed in the mediolateral oblique projection for younger women. Annual US was performed alone from 20-25 years, then with mammography from 25-35 years, then 6 months after mammography from 35-40 years and above 40 years only if requested by the radiologist. US was performed with either 7.5MHz or 10-12 MHz probes (ATL HDI 3500, Philips).	Sample no = 23 women at high risk of breast cancer (2 cases did not get US). No average age of women given, range was 30-61 years. Inclusion criteria: • BRCA1 or BRAC2 mutation carrier, or 1 in 2 probability to be a mutation carrier (on the basis of positive mutational analysis in close relatives). With a negative or positive personal history for breast or ovarian cancer. OR • High risk for breast cancer according to criteria specified in paper. Risk stratification: As above, either BRCA1 or BRCA2 carrier, 1 in 2 probability of being a carrier or >50% risk of carrying a susceptibility gene for familial breast cancer on basis of family history.	Relevant outcomes: • cancer detection rate; • mode of detection; • tumour size and stage. Verification of positive results was with pathology and verification of negative results was with follow-up. There is no mention of the mean length of follow-up.	Cancer detection: 4 breast cancers were detected overall. Mode of detection: 3 tumours were detectable by CBE but none of the tumours were detected by mammography or US examination (although 1 woman did not receive an US). Tumour size and stage: All 4 tumours were invasive: 2 ductal invasive carcinomas, 1 lobular invasive carcinoma and 1 which was mixed ductal and lobular. 2 occurred in mutation carriers and 2 in women at high risk through family history. Only 2 tumours had the size recorded and these were 10mm and 30mm. No record of nodal status was given. There was no mention of interval tumours.	Limitations include: Small sample size. There are few characteristics given of the women selected other than their risk assessment. There is no information on how they were selected and the characteristics of any women who did not agree to participate. There is no mention of mean age, reproductive history, exogenous hormone use or preventative strategies (i.e. Tamoxifen use or BSO). There is also no indication of which women were having prevalent or incident surveillance screens and for how long they were followed up in the study. There is likely verification bias and this is more likely, the shorter the follow-up period. Authors' conclusions: The authors' conclusions relate to the overall surveillance strategy with MRI included. Breast MRI demonstrated to be a very useful technique for investigating breast disease. It is not influenced by breast density and does not use ionising radiation. For these reasons, it has been proposed to support mammography in the surveillance of BRCA-mutated patients. Moreover, according to the reported results, breast MRI seems very helpful in the high risk patients group. We believe the breast MRI can be very useful within this kind of surveillance, with a less invasive approach to the disease.

Table 17. Primary research studies appraised investigating the accuracy and efficacy of ultrasound surveillance compared to mammography on outcomes from breast cancer (continued)

Source	Study design Evidence grading	Comparison interventions and dates of testing	Sample	Outcomes and verification	Results	Comments
Trecate et al. (2003) Italy *Continued*		MRI was performed annually for all ages for 2 years during the study. A Siemens Vision 1.5 was used with a dedicated double coil. One pre-contrast image and 5 post-contrast images were taken. The contrast agent was Gd-DTPA at 0.1mmol/kg. The method of interpreting the MRI or mammography is not presented. The study was conducted over a 7-month period; however the exact dates are not given.	The latter refers to at least 3 cases of breast cancer before 60 years of age, at least 3 cases of breast cancer before 60 years of age and ovarian cancer at any age, or at least 3 cases of breast cancer before 60 years of age and male breast carcinoma at any age. 5 of the women had a personal history of breast cancer, 1 for ovarian cancer and 1 for ovarian and breast cancer. (1 had had a mastectomy, but the others had conservative surgery combined with radiation therapy). Recruited from the National Cancer institute in Milan, Italy			In the case of confirmed good diagnostic results, it could be proposed to be used every other year as an alternative to mammography. Reviewers' conclusions: This study suggests that XRM and US perform equivalently in the surveillance of women at high risk of breast cancer and that they are both less effective than CBE. However, the sample is very small, as is the number of tumours detected. It is difficult to know how long the women were followed up for and this would affect the reliability of the results. There could be false negatives that had not yet come to light. There is also a specific method of risk stratification in this study, which includes women with a personal history of breast cancer (although only if they are BRCA1 or BRCA2 mutation carriers), and this will affect the generalisability of the study. In addition the results are not presented in a very clear manner and it is not possible to determine the overall sensitivity and specificity for all the modalities of surveillance utilised, which would have been valuable information.

Table 17. Primary research studies appraised investigating the accuracy and efficacy of ultrasound surveillance compared to mammography on outcomes from breast cancer (continued)

Source	Study design Evidence grading	Comparison interventions and dates of testing	Sample	Outcomes and verification	Results	Comments
Sim et al. (2004) The Netherlands	Retrospective cohort study III-2 (C1 P1 Q3)	Mammography, MRI and breast US were performed in the original study. Results of MRI and mammography are reported in the original study and this paper focuses on US. 10MHz probe was used for the US scanning. Retrospective data from Nov 1994 to Feb 2001 (study by Stoutjesdijk et al., 2001).	Sample no = 84 women in the original study underwent US and MRI examination. Of these women, 48 biopsies were performed in 42 women. Mean age of the women biopsied was 42.4 years (range 25-58 years) Inclusion criteria: • >15% lifetime risk of breast cancer from family history of breast or ovarian cancer. • Adequate follow-up of at least 2 years or histolo-pathological details. 66 women were excluded in the study this data came from, due to inadequate follow-up or lack of histopathological correlation. Risk stratification had been done by the Claus model (family history factors).	Relevant outcomes • cancer detection rate; • sensitivity; • specificity; • PPV; • NPV; • Accuracy. Verification of positive results was by histopathological results. Verification of negative results was by at least 2 years follow-up. In the case of interval tumours, the results were verified as being true negatives by reviewing the last surveillance films to confirm they were negative.	Cancer detection: There were 48 biopsies done in 42 women and cancer was diagnosed in 15 cases (in 13 women). 2 women had bilateral but asynchronous lesions occurring at least 1 year apart. The cancer detection rate was 0.24 for US alone. Mode of detection: Of these 42 women, 7 did not have US examination. Sensitivity: US 83.3% XRM 53.9% Specificity: US 65.5% XRM 85.7% PPV: US 50% XRM 63.6% NPV: US90-.5% XRM 80% Diagnostic accuracy: US70.7% XRM75.6% Results of MRI can be found in Stoutjesdijk et al. (2001)	Limitations include: Re-analysis of second hand data; the primary study was not looking specifically at US examinations. Only a small cohort of women had undergone US and it is likely that this was for a specific reason i.e. that the other examinations were equivocal. This would then enhance the accuracy of US compared with the other modalities. Verification bias likely. Results presented without enough raw data to check calculations. Radiologist not blinded to other investigations so results not purely a result of the US examination, will be biased by findings of other modalities. US examinations are always heavily operator dependent. This affects internal validity if different sonographers and external validity as may not be reproducible by another operator. Authors' conclusions: The authors' conclusions relate to surveillance by all modalities used in the Stoutjesdijk study, including MRI. Underscores need to provide surveillance for these women earlier. From the overall results of the original study MRI had the highest cancer detection rate, followed by mammography and US in that order. In other studies that examined surveillance in women at high risk of breast cancer, the biggest discrepancies between results for different modalities are seen in US. This may be explained by US being so highly operator dependent.

Table 17. Primary research studies appraised investigating the accuracy and efficacy of ultrasound surveillance compared to mammography on outcomes from breast cancer (continued)

Source	Study design Evidence grading	Comparison interventions and dates of testing	Sample	Outcomes and verification	Results	Comments
Sim et al. (2004) The Netherlands *Continued*						In many centres, US is the only alternative to mammography and should not be neglected, particularly for surveillance high-risk women for breast cancer. It would be reasonable to extend US surveillance to high risk individuals following mammography. When the results of the original study are also looked at, the combined mammography and US test would match MRI in sensitivity, specificity, PPV, NPV and accuracy at a fraction of the costs and should not be ignored in centres that do not have breast MRI expertise. Reviewers' conclusions This study suggests that there is a role for US in addition to mammography in the surveillance of women at high risk of breast cancer and that this may be as efficacious as utilizing MRI, but with a lower cost. It is suggested that there is also likely to be more wide spread experience in US usage then MRI, which improves performance. However, the study is limited by the small sample size and also as it is extrapolating conclusions from the results of another study that was not employing US as a screening tool, but an additive examination used only selected women. This means that there will be selection bias in the women who underwent US examination and this is likely to have biased the results in favour of US. Consequently the results of this reanalysis of the work by Stoutjesdijk et al. must be interpreted with great caution.

Table 17. Primary research studies appraised investigating the accuracy and efficacy of ultrasound surveillance compared to mammography on outcomes from breast cancer (continued)

Source	Study design Evidence grading	Comparison interventions and dates of testing	Sample	Outcomes and verification	Results	Comments
Warner et al. (2004) Ontario and Montreal, Canada	Prospective cohort Study III-2 (C1 P2 Q2)	Study protocol: CBE biannually and mammography, US and MRI all performed annually 4 modalities all performed the same day. (commencing at least 1 year after the woman's last mammogram.) CBE coded as normal, suggestive of benign disease, indeterminate, or suspicious of malignancy. Indeterminate CBE exams were repeated after 3 months. Mammography was conventional 4-view film. Further views done when necessary. MRI was performed with 1.5 T magnet (Signa, General Electrical Medical Systems). The first 38 patients in the first year were done in a single-turn elliptical coil after a bolus injection of 0.1mmol/kg of Gd-DTPA. Images were taken in the coronal plane. For the remaining patients, a phased-array coil arrangement was used.	Sample no = 236 female BRCA1 and BRCA 2 mutation carriers. Mean age at first surveillance 46.6 years (range 25-65years). Mean age of diagnosis was 47.4 years (33.4-63 years). Recruited from familial cancer clinics Inclusions: • BRCA1 or BRCA2 mutation carrier. Exclusions: • past history of unilateral breast cancer if the contralateral breast not intact; • pregnant or lactating women (participation deferred); • history of bilateral breast cancer, currently undergoing chemotherapy or known to have metastatic disease;	Relevant outcomes • cancer detection rate; • mode of Detection • tumour stage, size and node status; • interval cancers; • mortality; • sensitivity; • specificity; • PPV; • NPV; • ROC curves. N.B. the PPV and specificity do not include in the denominator women that had additional diagnostic studies that did not result in biopsy. Verification of positive results was by pathology, biopsy was undertaken if there was suspicion from any of the four modalities of surveillance. Verification of a negative result was through follow-up.	Cancer detection: 22 cancers were detected in 21 women (1 bilateral). 93 per 1,000 under surveillance (7 of these women had previous breast cancer). Mode of detection. 2 were detected by CBE (9.1%). 8 by mammography (36%). 7 by US (33%). 2 cancers (9.1%) were detected by mammography alone, 2 were detected by US alone (9.5%, not all women had undergone US testing). Tumour stage, size and node status: 6 tumours were DCIS and 16 were invasive (15 infiltrating ductal and 1 invasive lobular). The mean size of the invasive tumours was 11mm at the first round and 13mm at the second round(overall range 5-60mm) 15 cases were node sampled and 2 were node-positive. Interval cancers: There was only 1 interval cancer, detected, in a 40-year-old BRCA1 mutation carrier 7 months after her 3rd surveillance screen (retrospectively this tumour was visible on MRI and on mammography at last surveillance.	Limitations included: Likely verification bias. Selected participants are very high risk, being proven mutation carriers and also including those with a prior history of breast cancer. It is not clear which were incident and which were prevalent rounds and which tumours were detected at which round. (A large number of women had had prior mammography). No mention of whether women had had risk reducing measures such as bilateral salpingo oophorectomy or Tamoxifen. Was quite a high level of attrition in the study and the characteristics of those women are not outlined. This may have introduced bias. Authors' conclusions: The authors' conclusions relate to the surveillance strategy overall, including MRI. This study of BRCA mutation carriers demonstrates that the addition of annual MRI and US to mammography and CBE significantly improves the surveillance for detecting early breast cancers. The use of US did detect additional tumours, but had a high false-positive rate and in light of this its benefit remains to be seen. There was no observed benefit from CBE over and above the 3 imaging modalities.

Table 17. Primary research studies appraised investigating the accuracy and efficacy of ultrasound surveillance compared to mammography on outcomes from breast cancer (continued)

Source	Study design Evidence grading	Comparison interventions and dates of testing	Sample	Outcomes and verification	Results	Comments
Warner et al. (2004) Ontario and Montreal, Canada *Continued*		This provided sagital images. US used a 7.5MHz transducer (the first 7 patients did not receive US). Each imaging modality was read independently by a radiologist and scored on the 5 point BIRADS scale. All lesions with a score of 4 or 5 were biopsied. Pre-menopausal women had surveillance performed mid menstrual cycle to avoid changes due to cyclical hormonal variation. Radiologists were blinded to the results of CBE	• women weighing >91kg (technical reasons). Risk stratification not really performed as only BRCA mutation carriers included (all very high risk group). There were 137 (58%) BRCA1 mutation carriers and 99 (42%) BRCA2 mutation carriers. 31% were Ashkenazi Jews. 30% had a history of breast cancer, 9% a history of ovarian cancer and 60% had no history of cancer or a history of another type of cancer. 85% of the women (*n*=205) had had mammography within the last 15 months and therefore this was an incident rather than a prevalent round for them.	All patients were followed up for a minimum of 1 yr from the date of the last surveillance examination.	visit). Another woman, who elected to have a bilateral mastectomy after breast cancer was found, had a 2mm focus of DCIS in the contra lateral breast which had not shown up at surveillance 2 months earlier. Mortality All 22 patients who had tumours diagnosed were still alive and disease-free at the time the article was written. It was felt that the cancers detected on the second round were of an earlier stage. The 2 node-positive tumours were detected in the first round. However, it was not exactly clear that the first round was really a prevalent round as a high percentage of women had had prior mammography. It was found that false-positives and false-negatives decreased from the first to the second and then to the third round of surveillance. The measures of accuracy are therefore presented by the surveillance modality and by the year of surveillance. These can be seen in the paper, but overall values for the 3 years are reported here.	MRI-based surveillance is likely to become the cornerstone of breast cancer surveillance for BRCA1 and BRCA2 mutation carriers, but it is necessary to demonstrate that this surveillance tool lowers breast cancer mortality before it can be recommended for general use. Reviewers' conclusions: This study suggests that US has an equivalent cancer detection rate and sensitivity to XRM but a lower specificity and PPV. This is due to the higher number of false-positives generated by US. The combination of XRM and US has a higher sensitivity than either alone but retains the poorer specificity and PPV. As the authors suggest, this study does not answer whether this translates into reduced mortality. However, the tumours detected did seem to be of an earlier stage and smaller size, with only 2 tumours node-positive. The results of this study are limited to the very high risk population of women who are proven BRCA1 or BRCA2 mutation carriers, and including those with a personal history of breast cancer. It may therefore not be generalisable to all women with an increased risk of breast cancer due to a family history. Further studies with larger numbers and longer follow-up, and including women of other risk groups are required.

Table 17. Primary research studies appraised investigating the accuracy and efficacy of ultrasound surveillance compared to mammography on outcomes from breast cancer (continued)

Source	Study design Evidence grading	Comparison interventions and dates of testing	Sample	Outcomes and verification	Results	Comments
Warner et al. (2004) Ontario and Montreal, Canada *Continued*		31 women left the study before completing 3 rounds, 16 underwent bilateral mastectomy, 3 were too large for MRI machine, 3 stopped due to pregnancy, 4 developed metastatic cancers, 4 were lost to follow-up and 1 did not wish to continue participating. All participants underwent the first round, but only 58% the second and 36% the third (a total of 120 women were still undergoing surveillance when the paper was written). No direct comparisons were made in this study other than between modalities. Dates of surveillance were between Nov 1997 and March 2003.	45% were pre-menopausal and 55% were post-menopausal.		Sensitivities of combinations of modalities: XRM + CBE = 45% CBE + XRM + US = 64% Measures of accuracy of individual modalities: Sensitivity (95% CI) XRM = 36% (17.1 to 59.3%) US = 33% (14.6 to 56.9%) Specificity (95% CI) XRM = 99.8% (98.7 to 99.9%) US = 96% 93.7 to 97.7%) PPV (95% CI) XRM = 89% (51.7 to 99.7%) US = 29% (12.6 to 51.1%) NPV (95% CI) XRM = 97% (94.8 to 98.3%) US = 97% (94.5 to 98.25) AUC: XRM= 0.77 US = 0.65 CBE + XRM + US = 0.81	

Table 17. Primary research studies appraised investigating the accuracy and efficacy of ultrasound surveillance compared to mammography on outcomes from breast cancer (continued)

Source	Study design Evidence grading	Comparison interventions and dates of testing	Sample	Outcomes and Verification	Results	Comments
Kuhl et al. (2005b) Germany	Prospective Cohort study III-2 (C1 P2 Q2)	Surveillance protocol Biannual CBE and US and annual XRM and MRI. If abnormalities found on CBE or US at round without XRM or MRI, these additional modalities were used to further investigate this. Surveillance commenced at 30 years or 5 years before the youngest family member affected with the disease. (N.B. in first 2 years, women under 30, or 30-39yrs with dense breasts did not receive XRM, but this was subsequently abandoned and all women received XRM) Mammography (XRM): Annual conventional film screen XRM performed with at least 2 views per breast (medio-lateral oblique and caudal-cranial), obtained and interpreted in accordance with German radiological practice guidelines.	Sample no = 529 (out of 590 eligible women – 49 were lost to follow-up after 1 surveillance round and 12 were also excluded as they had a clinical abnormality at initial examination) Inclusion criteria: • asymptomatic women • Personal history of breast cancer included provided that the patient had not undergone bilateral mastectomy, had not received chemotherapy within the previous 12 months and had no metastases. (139 women were included with a personal history of breast cancer) Exclusion criteria: • Clinical signs of breast cancer • chemotherapy within the previous 12 months	Relevant outcomes: cancer detection; mode of detection; tumour size; tumour stage; node status; interval tumours; sensitivity; specificity; PPV; NPV. Verification of a positive result was achieved by histology (for positive imaging studies). Verification of a negative result was achieved by follow-up (for negative imaging studies. If a breast cancer was identified clinically (by palpation) between surveillance rounds or at the 6-month clinical visit, the imaging studies of the previous round were considered false-negative. Mean follow-up was 5.3 years (range 2-7 years) (a total of 1,542 annual surveillance rounds were completed).	Cancer detection: A total of 43 breast cancers were identified in 41 patients (11 of these women had a prior history of breast cancer), 40 of these were said to be detectable by imaging. 81 per 1,000 women under surveillance Mode of detection: CBE identified only one tumour (also detected on imaging). XRM identified 14 tumours (only 1 was diagnosed by XRM that wasn't diagnosed by MRI). US identified 17 tumours (2 of these were at the half-yearly CBE and US and they were not palpable). US + XRM detected 21 tumours Tumour size, stage and node status: Of the 21 cancers detected by XRM and US, 16 were invasive and the rest were DCIS. The invasive cancers had a mean size of 13.9mm and 5 were node positive. Interval tumours: The paper states that 40 out of 43 tumours in this cohort were detected by imaging.	Limitations included: CBE and the imaging studies were performed within a time frame of 8 weeks. Few sample characteristics presented, such as OCP or HRT use, or the use of preventative strategies such as tamoxifen or BSO. Verification bias is likely. Unclear documentation of interval tumours. Lack of blinding to the results of the CBE Author's conclusions: The authors' conclusions relate to the surveillance strategy as a whole including MRI. If US is used in combination with XRM, it can help compensate for some but by far not for all of the shortcomings of XRM, and it causes a substantial number of false positive diagnoses. If MRI is used for surveillance, XRM proved to be of limited and ultrasound of no additional value. US may however be useful to bridge the relatively long time interval between annual surveillance rounds. Propose that in view of the insufficient diagnostic accuracy of XRM and USS, that breast MRI should be considered an integral part of surveillance programmes for women at high familial risk in particular in documented carriers of pathogenic BRCA mutations .

Table 17. Primary research studies appraised investigating the efficacy of ultrasound surveillance compared to mammography on outcomes from breast cancer (continued)

Source	Study design Evidence grading	Comparison interventions and dates of testing	Sample	Outcomes and verification	Results	Comments
Kuhl et al. (2005b) Germany *Continued*		Diagnoses coded according to the BI-RADStm diagnostic categories on a 5-point scale (1, negative; 2, benign; 3, probably benign; 4, suspicious abnormality; 5, highly suggestive of malignancy). Breast MRI: Standard dynamic axial contrast-enhanced breast MRI of both entire breasts was performed on a 1.5T system (NT/INTERA; Philips, Best, the Netherlands) after injection of 0.1mmol/kg body weight gado-pentetate dimeglumine (Magnevist, Schering, Berlin, Germany) Ultrasound (US): performed with 7.5MHz-13MHz probes (Siemens Elegra, GE logic 500 and ATL HDI 5000; Siemens, Erlangen, Germany). The entire breast was systematically examined by the physician who interpreted the study. Diagnoses were scored on a 5-point scale identical to the XRM BIRADS categories.	• women having undergone bilateral mastectomy. Recruited from high-risk clinics in a single gynaecology department. Risk stratification: According to definition of the Consortium on Familial Breast and Ovarian Cancer of the German Cancer Aid, corresponding to a lifetime risk of breast cancer of at least 20% (two or more cases of breast cancer on the same side of the family, including at least two cases with onset before age 50 years, or with breast or ovarian cancer, irrespective of age, families with at least one case of breast cancer diagnosed before 35 years, families with three or more cases of breast cancer on the same side of the family, and women who met the criteria for high familial risk, irrespective of the result of mutational analysis)	Verification of last surveillance round was by continued surveillance in 428 women, telephone interview in 52 women and for 6 women who had prophylactic mastectomy it was by pathology of the specimen. XRM: BIRADS of 4 or 5, biopsy was recommended irrespective of finding in US or MRI. BIRADS 3 was managed by 6-months follow-up until receiving a BIRADS 2 or biopsy clarification. US categorised as BI-RADS 3 managed by short-term (6 months) US follow-up. BIRADS 4 or 5 managed by US-guided biopsy (14G, semi-automatic or automatic biopsy gun) except for the following constellation: if an US finding that was suspicious was clearly benign on XRM or MRI no biopsy was performed.	However, a sentence in the discussion states that the rate of interval cancers was 2% in this cohort. This translates to 10 tumours that arose in the interval between screens if it is 2% of the participants and 1 tumour if it is 2% of the tumours overall. The latter is most likely but this is unclear. Comparisons: When stratified by risk groups, the detection rates at both the prevalent and incident rounds were much higher in the mutation carriers than the other 2 risk groups, but these differences are not statistically significant. Sensitivity (95% CI): XRM 32.6% (19 to 48.55) n = 14/43 US 39.5% (25.0 to 55.6%) n = 17/43 XRM+US 48.8% n = 21/43 (33 to 64.5%) When stratified by risk groups XRM, US and the combination of XRM+US all become less sensitive as the lifetime risk of breast cancer increases, with sensitivities of 25%, 25% and 37.5% respectively for the mutation carrier group. Sensitivity of XRM and US by risk group : Risk 20% = 83.3% (36 to 100%)	Reviewer's conclusions: This study suggests that US and XRM have similar sensitivity but US has a lower specificity and PPV. This is due to the number of false-positives generated by US. The combination of US and XRM has a higher sensitivity than either but retains the lower specificity and PPV of US When stratified by risk group the sensitivity of XRM, US and XRM all decreased. These modalities may not be so accurate or effective in women at highest risk and these women (mutation carriers) may require additional surveillance modalities or more intensive surveillance. The limitations of this study must be taken into account in the interpretation.

Table 17. Primary research studies appraised investigating the accuracy and efficacy of ultrasound surveillance compared to mammography on outcomes from breast cancer (continued)

Source	Study design Evidence grading	Comparison interventions and dates of testing	Sample	Outcomes and verification	Results	Comments
Kuhl et al. (2005b) Germany *Continued*		Each imaging study was read and scored independently by a different radiologist who had substantial experience with the respective imaging technique. The readers were informed about the clinical findings from CBE and the risk status of the patient but were blinded to the results of the respective other imaging modalities. Comparisons are made between the 3 risk groups and the different modalities of surveillance. Dates of study were February 1996 to February 2002. BIRADS 4 or above led to a biopsy.	In women without a personal history of breast cancer the Claus tables were also used to quantify risk. Women were then stratified into 3 risk groups for analysis: • mutation carriers; • high lifetime risk (20-40%); • moderate lifetime risk (20%).	MRI: Suspicious scores (4 or 5) were managed by magnetic resonance-guided biopsy. Findings categorised as BIRADS 3 short-term follow-up after 6 months was recommended with further management corresponding to that of XRM BIRADS 3 lesions BIRADS 3 categories in all imaging that received short-term follow-up were not considered positive for the calculation of outcomes. Invasive cancer and DCIS were considered a malignant diagnosis but LCIS and atypical ductal hyperplasia were considered to be benign.	Risk 21-40% = 45.0% (23 to 68%) Mutation carriers = 37.5% (9 to 76%) Specificity (95% CI): XRM 96.8% (95.7 to 97.7%) n = 1364/1409 US 90.5% (88.8 to 92.0%) n = 1275/1409 XRM+US 89.0% (87.2 to90.6%) n = 1254/1409 Stratification by risk group does not appear to affect the specificity. PPV (95% CI): XRM 23.7% (14 to 37%) n = 14/59 US 11.3% (6.7 to 17.4%) n = 17/151 XRM+US 11.9% (7.5 to 17.6%) n = 21/176 The PPV increases with the increasing risk of breast cancer, this will be affected by the higher incidence in women at higher risk. NPV (95% CI): XRM 97.9% (97.0 to 98.6%) n = 1364/1393 US 98% (97.1 to 98.7%) n = 1275/1301 Not enough data to calculate NPV for the combination strategies.	

Chapter 6: Accuracy and efficacy of MRI

SECONDARY RESEARCH

The search strategy did not identify any systematic reviews that compared the accuracy and/or effectiveness of MRI surveillance in women at high risk of breast cancer to no surveillance at all, or to surveillance with clinical breast examination.

PRIMARY RESEARCH: STUDY DESIGNS AND QUALITY

Study design

The search identified 4 eligible primary research studies. All four studies were prospective cohort studies (Kriege et al. 2004; Kuhl et al. 2005b; Trecate et al. 2003; Warner et al. 2004). The studies by Trecate et al. (2003) and Kuhl et al. (2005b) were conducted in single centres in Italy and Germany respectively. The other two were multi-centre studies undertaken in Canada (Warner et al. 2004) and the Netherlands (Kriege et al. 2004). No studies were identified that compared MRI surveillance to no surveillance. Therefore, it is unknown at this stage whether there is any benefit in terms of survival and response to cancer treatment of surveillance with MRI in women at high risk of breast cancer over no surveillance in this population. Each of the included studies compared surveillance with MRI to CBE in women at high risk of breast cancer. All of these studies also included surveillance with XRM, and three included surveillance with US (Kuhl et al. 2005b; Trecate et al. 2003; Warner et al. 2004). The results of surveillance with these modalities will be discussed in subsequent chapters.

Study setting

The study by Trecate et al. (2003) took place in a clinic in Milan, Italy. The study by Kriege et al. (2004) was a conducted in six familial-cancer clinics in the Netherlands. Warner et al. (2004) recruited women from familial cancer clinics in southern Ontario and Montreal, Canada, although all surveillance tests were undertaken in one centre in Ontario. The study by Kuhl et al. (2005b) was undertaken at a medical school in Bonn, Germany.

Below is an overview of study designs and aspects of quality represented by these studies. Full details of the papers appraised including methods, key results, limitations and conclusions are provided in Evidence **Table 19**. One study (Warner et al. 2004) examined mortality. None of the studies examined response to treatment of cancers diagnosed by the surveillance tests examined. However the studies examined the size and node status of the tumours identified, which may act as a surrogate outcome in the evaluation of effectiveness of the surveillance test for the early detection of breast cancer in women at high risk of breast cancer because of genetic or family history.

Trecate et al. (2003)

Study sample

In a small study, Trecate et al. (2003) enrolled 23 women at high risk of breast cancer on the basis of being a BRCA1 or BRCA2 mutation carrier or a one in two probability of being a carrier, on the basis of a positive genetic test in a close relative, or being at a high risk for breast cancer according to criteria specified relating to family history.

Interventions and comparators

MRI was performed annually for all ages for two years during the study. One pre-contrast image and five post-contrast images were taken. The contrast agent was Gd-DTPA at 0.1mmol/kg. In addition, CBE was performed every six months for all ages. The women also underwent surveillance using annual mammography and ultrasound scanning, the results of which will be presented in subsequent chapters. The methods for interpreting the findings of each surveillance test were not presented; nor was it clear when the tests took place in relation to each other. It was not reported whether the radiologists interpreting the images was blinded to the results of the other imaging modalities.

Outcomes

The principal outcomes in this study were breast cancer detection rate, the mode of tumour detection, and tumour size, stage and node status.

Four breast cancers were detected overall, all of which were detected by MRI, and three of which were detected by CBE. None of the tumours were detected by mammography or ultrasound examination, although one of the women who had breast cancer did not receive an ultrasound scan.

Kriege et al. (2004)

Study sample

The study by Kriege et al. (2004) was a multi-centre prospective cohort study, in which MRI was compared to mammography for the surveillance of women with a genetic or familial pre-disposition to breast cancer to determine whether surveillance with MRI facilitated early diagnosis of breast cancer. The study was conducted in six familial-cancer clinics in the Netherlands. Surveillance consisted of a clinical breast examination performed by an experienced physician every six months and imaging studies (mammography and MRI) performed annually by experienced radiologists. The XRM results will be reported in a subsequent chapter. A total of 1909 women, who had a cumulative lifetime risk of breast cancer of 15 per cent or more, received surveillance. The mean age of these women was 40 years, with a range of 19-72. Within the group of 358 carriers of pathogenic mutations, 276 women had BRCA1 mutation, 77 women had a BRCA2 mutation, one woman had BRCA1 and BRCA2 mutations, two women had a PTEN mutation and two women had a TP53 mutation. Women with symptoms of breast cancer or a personal history of breast cancer were excluded.

Interventions and comparators

Both imaging investigations were performed on the same day or in the same time period, between days 5-15 of the menstrual cycle. The MRI screening test was undertaken according to a standard protocol with the use of gadolinium-containing contrast medium. Results of imaging examinations were scored in a standardised way according to the BIRADS 5-point scale (1, negative; 2, benign; 3, probably benign; 4, suspicious abnormality; 5, highly suggestive of malignancy).

Outcomes

The primary outcomes were measures of test accuracy including sensitivity, specificity and positive predictive values of each surveillance test. The results of each exam were blinded so that the two examinations were not linked. If one of the imaging exams was a BI-RADS 3 or 0 ('need additional imaging evaluation') further investigation by ultrasound with or without fine-needle aspiration was advised, or the MRI or mammogram was repeated. When one of the two exams was BIRADS 4 or 5 a cytologic or histologic evaluation of a biopsy specimen was performed.

The mean follow-up period was 2.7 years. During this time, 51 malignant tumours were detected (44 invasive breast cancers, six DCIS and one non-Hodgkin's lymphoma). A total of 32 breast cancers were detected by MRI. Conversely, 13 cancers were missed by MRI (including five DCIS, four interval cancers and one tumour which was detected only by CBE). The sensitivities of clinical breast examination for detecting invasive cancer were 6.7 per cent (with a cut-off of 'suspicious') and 17.8 per cent (with a cut-off of 'probably benign'). The sensitivity of MRI for detecting invasive cancer was 71.1per cent (with a BIRADS cut off of 3). The specificities were 98.1 per cent and 89.8 per cent for CBE and MRI respectively. To evaluate the discriminating capacity of the imaging methods, receiver

operating curves were generated. The area under the curve was 0.827 for MRI. The AUC was not calculated for CBE.

Warner et al. (2004)

Study sample

Warner et al. (2004) recruited women from familial cancer clinics in southern Ontario and Montreal, Canada, although all surveillance tests were undertaken in one centre in Ontario. Women were included if they were a BRCA1 or BRCA2 mutation carrier. Exclusion criteria were a past history of unilateral breast cancer if the contralateral breast was not intact, history of bilateral breast cancer, currently undergoing chemotherapy or known to have metastatic disease. For technical reasons, women weighing more than 91kg were also excluded. The participation of pregnant or lactating women was also deferred. There were 236 participants in this study.

Interventions and comparators

CBE was performed biannually while the MRI was performed annually. Both the CBE and the MRI scan were performed on the same day, commencing at least one year after the woman's last mammogram. The CBE results were coded as normal, suggestive of benign disease, indeterminate, or suspicious of malignancy. Indeterminate CBE exams were repeated after 3 months. The MRI was performed with a 1.5 Tesla magnet. The first 38 patients in the first year were done in a single-turn elliptical coil after a bolus injection of 0.1mmol/kg of Gd-DTPA. Images were taken in the coronal plane. For the remaining patients, a phased-array coil arrangement was used which provided sagittal images. The results of the MRI were read and scored independently from the other modalities of surveillance by a radiologist and scored on the 5-point BIRADS scale. All lesions with a score of 4 or 5 were biopsied.

Outcomes

Sensitivity and specificity of MRI and clinical breast examination were the primary outcomes. Other relevant outcomes included the cancer detection rate, tumour stage, size and node status, interval cancers and mortality. The verification of positive screens was by pathology, biopsy was undertaken if there was suspicion from any surveillance modalities, while the verification of a negative screen was through follow-up. All patients were followed up for a minimum of one year from the date of the last surveillance examination.

Warner et al. (2004) found 22 cancers in 21 women (one woman had bilateral cancer). Seven of these women had previous breast cancer). Two cancers were detected by CBE (9.1%) and 17 by MRI (77%). Seven cancers (32%) were detected by MRI alone. Six of the detected tumours were DCIS and 16 were invasive (15 infiltrating ductal and 1 invasive lobular). The mean size of the invasive tumours was 11mm at the first surveillance round and 13mm at the second round with an overall range of 5-60mm. Fifteen cases were node sampled and two were node-positive. There was only one interval cancer, detected in a 40 year old BRCA1 mutation carrier seven months after her 3rd screen (retrospectively this tumour was visible on MRI and on mammography at last surveillance visit). Another woman, who elected to have a bilateral mastectomy after breast cancer was found, had a 2mm focus of DCIS in the contralateral breast which had not shown up at surveillance two months earlier. All 22 patients who had tumours diagnosed were still alive and disease-free at the time the article was written. It appeared that the cancers detected on the second screening round were of an earlier stage. The two node-positive tumours were detected in the first surveillance round. However, it was not clear whether the first surveillance round was really a prevalent round as a high percentage of women had had prior mammography.

After the first round of screening, 16.5 per cent of participants underwent a diagnostic MRI scan to clarify the status of an indeterminate or possibly suspicious lesion. This rate of referral for a second MRI decreased at the second and third rounds of surveillance to 9.6 per cent and 7.1 per cent respectively. For an additional 7.6 per cent of patients, a 6-month follow-up MRI was recommended for lesions that remained indeterminate and this rate decreased at the second and third rounds of surveillance to 2.9 per cent and 2.4 per cent respectively. A total of 2.1 per cent of CBEs were thought to be suspicious at the first round of surveillance and CBE was repeated three months later. The corresponding rate for the second and third years of surveillance was 0.4 per cent for CBE. This is

potentially suggestive of a learning effect whereby those reading the MRI scans become more skilled as a result of increased experience.

Kuhl et al. (2005b)

Study sample

The 529 study participants in Kuhl et al. (2005b) were recruited by the 'high-risk' departments of gynaecology between February 1992 and February 1996. Women were recruited if they were clinically asymptomatic and met the criteria of high familial risk corresponding to a lifetime risk of at least 20 per cent. Inclusion criteria were: two or more cases of breast cancer on the same side of the family, including at least two cases with onset before age 50 years, or with breast or ovarian cancer, irrespective of age, families with at least one case of breast cancer diagnosed before 35 years, families with three or more cases of breast cancer on the same side of the family, and women who met the criteria for high familial risk, irrespective of the result of mutational analysis. Personal history of breast cancer was not an exclusion, provided that the patient had not undergone bilateral mastectomy, had not received chemotherapy within the previous 12 months, and had no metastases. Women were excluded from the study if they had clinical signs of breast cancer, chemotherapy within the previous 12 months, and those having undergone bilateral mastectomy.

Interventions and comparators

The protocol was biannual CBE and US, and annual XRM and MRI. Surveillance consisted of each of the imaging tests being performed within a time frame of eight weeks. Each imaging study was read and scored independently by a different radiologist who had substantial expertise with the respective breast imaging technique. Although the radiologists were informed about the clinical findings from CBE and the risk status of the patient, they were blinded to the results of the respective other imaging modalities. The diagnoses were coded according to the BIRADS diagnostic categories on a 5-point scale (1, negative; 2, benign; 3, probably benign; 4, suspicious abnormality; 5, highly suggestive of malignancy). For the breast MRI, standard dynamic axial contrast-enhanced breast MRI of both entire breasts was performed on a 1.5T system after injection of 0.1mmol/kg body weight (Gd-DTPA).

Outcomes

The primary outcomes were the test accuracy measures of sensitivity, specificity and positive predictive values for each of the imaging modalities used alone or in various combinations. Validation of the imaging results was achieved either by histology for positive imaging studies or by follow-up for negative imaging studies. If a breast cancer was identified clinically by palpation between surveillance rounds or at the 6-month clinical visit the imaging studies of the previous round were considered false-negative. For any of the surveillance tests classified as a BIRADS score of 4 or 5, biopsy was recommended irrespective of findings of the other surveillance tests. Each biopsy was guided by whichever surveillance modality detected the lesion. Where a BIRADS score of 3 was recorded on any of the imaging tests, women were managed by 6-months follow-up until receiving a BIRADS 2 score or biopsy clarification.

Women were observed for a mean observation period of 5.3 years with a range of 2-7 years. In the entire cohort of 529 patients, a total of 43 breast cancers were identified in 41 patients. There were 34 invasive cancers and nine DCIS. Clinical breast examination identified only one tumour, which was also detected by imaging. Of the 40 cancers diagnosed by imaging studies, 39 cancers were detected by MRI.

The sensitivity scores for CBE and MRI were 2.3 per cent and 91 per cent respectively. The specificity for MRI was 97.2 per cent. There was not sufficient data to calculate the specificity of CBE.

The positive predictive value for MRI for all women regardless of risk status was 50 per cent (95% CI, 38% to 62%). Of the 19 cancers diagnosed only by MRI, 5 were high-grade intra-ductal cancers and 14 were invasive cancers with a median size of 7.5mm. Thirty-one cancers were identified in women without a previous history of cancer, and 12 cancers in 11 women with breast cancer history. Of the latter 12 cancers, three were classified as local recurrences and nine cancers occurred in the contra-lateral breast and/or were histologically categorised as second primary cancers. Two cancers were

palpable at the time of diagnosis (one in the regular surveillance interval, and one was an interval cancer diagnosed between surveillance rounds). The remaining cancers were asymptomatic (not palpable).

Summary:

Four primary studies were identified that examined MRI and CBE surveillance of women at high risk of breast cancer owing to genetic predisposition or family history. No studies were found that compared MRI surveillance in women at high risk of breast cancer with no surveillance at all. All of the four primary studies were prospective cohort studies. A total of 2,697 women received surveillance in the four studies. There was some heterogeneity of risk level among the women included. In Warner et al. (2004) only women with BRCA genetic mutations were included. Warner et al. (2004) and Kriege et al. (2004) also excluded women with a personal history of breast cancer while women with a personal history of breast cancer were included in Kuhl et al. (2005b), conditional in the latter study on the contralateral breast being imaged.

MRI surveillance was used in all the included studies and images were taken before and after the bolus injection of contrast enhancement. The MRI surveillance tests were performed annually. In one study (Kriege et al. 2004) clinical breast examination was performed at six-month intervals. The MRI screening results can only be compared with clinical breast examinations undertaken at approximately the same time as the MRI tests. In Kuhl et al. (2005b) tests could take place eight weeks apart, an interval which may preclude comparisons being made between these tests. In addition, the length of the optimal surveillance interval is unknown at this time.

Measures of diagnostic test accuracy (sensitivity, specificity and positive predictive values) were outcomes examined in the more recent studies (Kriege et al. 2004; Kuhl et al. 2005b; Warner et al. 2004). Data were also reported for particular subgroups, such as women with BRCA mutations and also at varying levels of risk for breast cancer based on family history by Kuhl et al. (2005b). There were few measures of effectiveness studied in the included trials. None of the studies examined survival outcomes, aside from Warner et al. (2004), who stated that all the women under surveillance were still alive at the time of writing. In addition, there were no data presented on response to treatment as a result of the possible earlier diagnosis of cancerous tumours by MRI surveillance.

The principal results from the four included studies examining the comparison between MRI and CBE are summarised below. In the two studies that provided the most detailed information (Kriege et al. 2004; Warner et al. 2004), MRI appears to be more sensitive than CBE in detecting cancerous lesions in women at high risk of breast cancer. The specificities for both tests in both Kriege et al. (2004) and Warner et al. (2004) are relatively high. However, in the study by Kriege et al, a slightly lower specificity score for MRI (89.8% compared to 98.1% for CBE) is suggestive of a higher rate of false positive results for MRI surveillance. This is an important consideration, given the higher levels of anxiety faced by women who are referred for biopsy because of a suspicious surveillance test result, in particular because their family history makes it highly likely they have witnessed family members be diagnosed and treated for breast cancer. In the study by Warner et al. (2004), the specificity for MRI was 95.4 per cent. The investigators in Warner et al. (2004) reported that on the second and third years/rounds of surveillance, the referral rates for a diagnostic MRI test decreased by nearly 50 per cent. This is potentially suggestive of a learning effect whereby those reading the MRI scans become more skilled as a result of increased experience and the availability of previous films for comparison, and this may result in fewer false-positive results over the course of the study.

Table 18. Summary of results for MRI surveillance compared to CBE

	Screening modality	No of cancers detected/total cancers	Sensitivity (95% CI)	Statistical significance	Specificity (95% CI)
Trecate et al. (2003)	MRI CBE	4/4 3/4	Not calculated	Not tested	Not calculated
Kriege et al. (2004)	MRI CBE	32/45 8/45	71.1%* (56 to 84%) 17.8%* (8 to 32%)	Not tested	89.8%-99.9%† 98.1%-99.9%†
Warner et al. (2004)	MRI CBE	17/22 2/22	77% (55 to 92%) 9.1% (1 to 29%)	Not tested	95.4% (92.9 to 97.2%) N/R
Kuhl et al. (2005b)	MRI CBE	39/43 1/43	90.7% (78 to 97%) 2.3% (0 to 12%)	Not tested	97.2% (96.2 to 98%) Not calculated

* calculated for test results with BIRADS cut-off of 3 (and a cut-off of 'probably benign' for CBE)
† Specificity calculated for a range of BIRADS cut-offs, lower range BIRADS 3 and higher range BIRADS 5 (and cut-off for CBE, lower range 'probably benign' and higher range 'suspicious')

In conclusion, MRI surveillance appears to be superior to clinical breast examination for the detection of breast cancer in women at high risk of breast cancer owing to family history or genetic predisposition. The specificity scores for MRI were also high, indicating that the testing in these two studies is associated with a relatively small number of false-positive test results. However, there were no data on whether surveillance with MRI confers any benefit at all compared to no surveillance in this population. At this point, there is also no data on whether the early detection of cancerous lesions by MRI surveillance confers any benefits in terms of survival of women at high risk of breast cancer who have been diagnosed with breast cancer.

Table 19. Primary research studies appraised investigating the accuracy and efficacy of MRI screening compared to usual care on outcomes from breast cancer

Source	Study design Evidence grading	Comparison interventions and dates of testing	Sample	Outcomes and verification	Results	Comments
Trecate et al. (2003) Italy (NB: Podo is an author on this one as well but we cannot find any further reports from the Podo et al trial.)	Prospective cohort study III-2 (C1 P2 Q3)	Surveillance protocol: outlined in full in the paper and was dependent on age group. CBE was performed every 6 months for all ages. Mammography was annual and commenced at 25 years with bilateral one-view, and then increased to bilateral double-view from 30 years and above. Double-view was performed in craniocaudal and mediolateral oblique projections. One-view was performed in the mediolateral oblique projection for younger women. Annual US was performed alone from 20-25 years, then with mammography from 25-35 years, then 6 months after mammography from 35-40 years and above 40 years only if requested by the radiologist. US was performed with either 7.5MHz or 10-12 MHZ probes (ATL HDI 3500, Philips). MRI was performed annually for all ages for 2 years during the study. A Siemens Vision 1.5 was used with a dedicated double coil.	Sample no = 23 women at high risk of breast cancer (2 cases did not get US). No average age of women given, range was 30-61 years. Inclusion criteria: • BRCA1 or BRCA2 mutation carrier or 1 in 2 probability to be a mutation carrier (on the basis of positive mutational analysis in close relatives) With a negative or positive personal history for breast or ovarian cancer OR • High risk for breast cancer according to criteria specified in paper Risk stratification: As above, either BRCA1 or BRCA2 carrier, 1 in 2 probability of being a carrier or >50% risk of carrying a susceptibility gene for familial breast cancer on basis of family history.	Relevant outcomes: Cancer detection rate. Mode of detection. Tumour size and stage. Verification of positive results was with pathology and verification of negative results was with follow-up. There is no mention of the mean length of follow-up.	Cancer detection: 4 breast cancers were detected overall. Mode of detection: All 4 tumours were detected by MRI and 3 were detectable by CBE. It is stated that there were no false positives or false negatives for MRI. Tumour size and stage: All 4 tumours were invasive: 2 ductal invasive carcinomas, 1 lobular invasive carcinoma and 1 which was mixed ductal and lobular. 2 occurred in mutation carriers and 2 in women at high risk through family history. Only 2 tumours had the size recorded and these were 10mm and 30mm. No record of nodal status was given. There was no mention of interval tumours.	Limitations include: Small sample size. There are few characteristics given of the women selected other then their risk assessment. There is no information on how they were selected and the characteristics of any women who did not agree to participate. There is no mention of mean age, reproductive history, exogenous hormone use or preventative strategies (i.e. Tamoxifen use or BSO). There is also no indication of which women were having prevalent or incident surveillance screens and for how long they were followed up in the study. There is likely verification bias and this is more likely, the shorter the follow-up period. Authors' conclusions: The authors' conclusions relate to the entire surveillance strategy including XRM and US. Breast MRI demonstrated to be a very useful technique for investigating breast disease. It is not influenced by breast density and does not use ionising radiation. For these reasons, it has been proposed to support mammography in the surveillance of BRCA mutated patients. Moreover, according to the reported results, breast MRI seems very helpful in the high-risk patients group. We believe the breast MRI can be very useful within this kind of surveillance, with a less invasive approach to the disease. In the case of confirmed good diagnostic results, it could be proposed to be used every other year as an alternative to mammography.

Table 19. Primary research studies appraised investigating the accuracy and efficacy of MRI surveillance compared to usual care on outcomes from breast cancer (continued)

Source	Study design Evidence grading	Comparison interventions and dates of testing	Sample	Outcomes and verification	Results	Comments
Trecate et al. (2003) Italy *Continued*		One pre-contrast image and 5 post-contrast images were taken. The contrast agent was Gd-DTPA at 0.1mmol/kg. The method of interpreting the MRI or mammography is not presented. The study was conducted over a 7-month period; however the exact dates are not given.	The latter refers to at least 3 cases of breast cancer before 60 years of age, at least 3 cases of breast cancer before 60 years of age and ovarian cancer at any age, or at least 3 cases of breast cancer before 60 years of age and male breast carcinoma at any age. 5 of the women had a personal history of breast cancer, 1 for ovarian cancer and 1 for ovarian and breast cancer (1 had had a mastectomy, but the others had conservative surgery combined with radiation therapy).			Reviewers' conclusions: This study suggests that MRI is a very effective tool for the surveillance of women at high risk of breast cancer. However, the sample is very small and it is difficult to know how long the women were followed up for and this would affect the reliability of the results. There could be false negatives that had not yet come to light. There is also a specific method of risk stratification in this study, which includes women with a personal history of breast cancer (although only if they are BRCA1 or BRCA2 mutation carriers), and this will affect the generalisability of the study. In addition the results are not presented in a very clear manner and it is difficult to determine the overall sensitivity and specificity for all the modalities of screening utilised, which would have been valuable information.

Table 19. Primary research studies appraised investigating the accuracy and efficacy of MRI surveillance compared to usual care on outcomes from breast cancer (continued)

Source	Study design Evidence grading	Comparison interventions and dates of testing	Sample	Outcomes and verification	Results	Comments
Warner et al. (2004) Ontario and Montreal, Canada	Prospective cohort Study III-2 (C1 P2 Q2)	Study protocol: CBE biannually and mammography, US and MRI all performed annually 4 modalities all performed the same day. (commencing at least 1 year after the woman's last mammogram) CBE coded as normal, suggestive of benign disease, indeterminate, or suspicious of malignancy. Indeterminate CBE exams were repeated after 3 months. Mammography was conventional 4-view film. Further views done when necessary. MRI was performed with 1.5 T magnet (Signa, General Electrical Medical Systems). The first 38 patients in the first year were done in a single-turn elliptical coil after a bolus injection of 0.1mmol/kg of Gd-DTPA. Images were taken in the coronal plane. For the remaining patients, a phased-array coil arrangement was used. This provided sagital images. US used a 7.5MHz transducer (the first 7 patients did not receive US).	Sample no = 236 female BRCA1 and BRCA2 mutation carriers. Mean age at first surveillance 46.6 years (range 25-65 years) Mean age of diagnosis was 47.4 years (33.4-63 years) Recruited from familial cancer clinics. Inclusions: • BRCA1 or BRCA2 mutation carrier. Exclusions: • past history of unilateral breast cancer if the contra lateral breast not intact; pregnant or lactating women (participation deferred); • History of bilateral breast cancer, currently undergoing chemotherapy or known to have metastatic disease; • Women weighing >91kg (technical reasons).	Relevant outcomes: Cancer detection rate; Mode of detection; Tumour stage, size and node status; Interval cancers; Mortality; Sensitivity; Specificity; PPV; NPV; ROC curves. NB: the PPV and specificity do not include in the denominator women that had additional diagnostic studies that did not result in biopsy. Verification of positive results was by pathology, biopsy was undertaken if there was suspicion from any of the four modalities of surveillance. Verification of a negative result was through follow - up. All patients were followed up for a minimum of 1 yr from the date of the last surveillance examination.	Cancer detection: 22 cancers were detected in 21 women (1 bilateral)(7 of these women had previous breast cancer). Mode of detection: 2 were detected by CBE (9.1%), 17 by MRI (77%). 7 cancers (32%) were detected by MRI alone. MRI detected 9 of the 12 cancers missed by conventional surveillance (Mamm + CBE). Tumour stage, size and node status: 6 tumours were DCIS and 16 were invasive (15 infiltrating ductal and 1 invasive lobular). The mean size of the invasive tumours was 11mm at the first screening round and 13mm at the second round (overall range 5-60mm) 15 cases were node sampled and 2 were node positive. Interval cancers: There was only 1 interval cancer, detected in a 40 year-old BRCA1 mutation carrier 7 months after her 3rd surveillance screen. (retrospectively this tumour was visible on MRI and on mammography at last surveillance).	Limitations include: Likely verification bias. Selected participants are very high risk, being proven mutation carriers and also including those with a prior history of breast cancer. It is not clear which rounds were incident and which were prevalent, and which tumours were detected at which round (a large number of women had had prior mammography). No mention of whether women had had risk-reducing measures such as nilateral salpingo oophorectomy or Tamoxifen. Was quite a high level of attrition in the study and the characteristics of those women are not outlined. This may have introduced bias. Authors' conclusions: The authors' conclusions relate to the overall surveillance strategy including XRM and US. This study of BRCA mutation carriers demonstrates that the addition of annual MRI and US to mammography and CBE significantly improves the surveillance for detecting early breast cancers. The use of US did detect additional tumours, but had a high false positive rate and in light of this its benefit remains to be seen. There was no observed benefit from CBE over and above the 3 imaging modalities. MRI-based surveillance is likely to become the cornerstone of breast cancer surveillance for BRCA1 and BRCA2 mutation carriers, but it is necessary to demonstrate that this surveillance tool lowers breast cancer mortality before it can be recommended for general use.

Table 19. Primary research studies appraised investigating the accuracy and efficacy of MRI surveillance compared to usual care on outcomes from breast cancer (continued)

Source	Study design Evidence grading	Comparison interventions and dates of testing	Sample	Outcomes and verification	Results	Comments
Warner et al. (2004) Ontario and Montreal, Canada *Continued*		Each imaging modality was read independently by a radiologist and scored on the 5-point BIRADS scale. All lesions with a score of 4 or 5 were biopsied. Pre-menopausal women had surveillance performed mid menstrual cycle to avoid changes due to cyclical hormonal variation. Radiologists were blinded to the results of CBE. 31 women left the study before completing 3 rounds, 16 underwent bilateral mastectomy, 3 were too large for MRI machine, 3 stopped due to pregnancy, 4 developed metastatic cancers, 4 were lost to follow-up and 1 did not wish to continue participating. All participants underwent the first round, but only 58% the second and 36% the third (a total of 120 women were still undergoing surveillance when the paper was written). No direct comparisons were made in this study.	Risk stratification not really performed as only BRCA mutation carriers included (all very high risk group). There were 137 (58%) BRCA1 mutation carriers and 99 (42%) BRCA 2 mutation carriers. 31% were Ashkenazi Jews. 30% had a history of breast cancer, 9% a history of ovarian cancer and 60% had no history of cancer or a history of another type of cancer. 85% of the women (*n*=205) had had mammography within the last 15 months and therefore this was an incident rather than a prevalent round for them. 45% were pre-menopausal and 55% were post-menopausal.		Another woman, who elected to have a bilateral mastectomy after breast cancer was found, had a 2mm focus of DCIS in the contra lateral breast which had not shown up at surveillance 2 months earlier. Mortality: All 22 patients who had tumours diagnosed were still alive and disease-free at the time the article was written. It was felt that the cancers detected on the second round were of an earlier stage. The 2 node-positive tumours were detected in the first round. However, it was not exactly clear that the first round was really a prevalent round as a high percentage of women had had prior mammography. It was found that false-positives and false-negatives decreased from the first to the second and then to the third round of surveillance. This is especially seen for the false-positives in MRI, which decreased from 15 to 4 to 1. This may have been due to increasing experience in the radiologists in interpreting these scans. The measures of accuracy are therefore presented by the modality of surveillance and by the year of the surveillance. These can be seen in the paper, but overall values for the 3 years are reported here.	Reviewers' conclusions: This study demonstrates that MRI is superior to CBE in the surveillance of BRCA1 and BRCA2 mutation carriers for breast cancer. As the authors suggest, this does not answer whether this translates into reduced mortality. However, the tumours detected did seem to be of an earlier stage and smaller size, with only 2 tumours node positive. The results of this study are limited to the very high risk population of women who are proven BRCA1 or BRCA2 mutation carriers and including those with a personal history of breast cancer. It may therefore not be generalisable to all women with an increased risk of breast cancer due to a family history. Further studies with larger numbers and longer follow-up, and including women of other risk groups are required.

Table 19. Primary research studies appraised investigating the accuracy and efficacy of MRI surveillance compared to usual care on outcomes from breast cancer (continued)

Source	Study design Evidence grading	Comparison interventions and dates of testing	Sample	Outcomes and verification	Results	Comments
Warner et al. (2004) Ontario and Montreal, Canada *Continued*		Dates of surveillance were between Nov 1997 and March 2003.			Measures of accuracy of individual modalities: Sensitivity (95% CI): CBE = 9% (1% to 29%) MRI = 77% (54.6 to 92.2%) Specificity (95% CI): MRI = 95% (92.9 to 97.2%) (was 99% in 3rd year) PPV (95% CI): MRI = 46% (29.5 to 63.1%) NPV (95% CI): MRI = 99% (97.2 to 99.6%) AUC: MRI = 0.89 CBE = 0.48	

Table 19. Primary research studies appraised investigating the accuracy and efficacy of MRI surveillance compared to usual care on outcomes from breast cancer (continued)

Source	Study design Evidence grading	Comparison interventions and dates of testing	Sample	Outcomes and Verification	Results	Comments
Kriege et al. 2004 NEJM	Prospective multicentre cohort study III-2 (C1 P1 Q2)	Clinical breast examination: performed by an experienced physician every six months. Imaging studies performed annually by radiologists. XRM: oblique and cranio-caudal views and if necessary, compression views or magnifications. MRI: Dynamic breast MRI with gadolinium-containing contrast medium according to a standard protocol. Whenever possible, both imaging investigations were performed on the same day or in the same time period, between days 5-15 of the menstrual cycle.	1,909 women with a genetic risk for breast cancer. Mean age 40 years (range 19-72). Within the group of 358 carriers of pathogenic mutations, 276 had BRCA1 mutation, 77 had a BRCA2 mutation, 1 woman had BRCA1 and BRCA2 mutations, 2 had a PTEN mutation and 2 had a TP53 mutation. Inclusion criteria: Cumulative lifetime risk of breast cancer 15% or more owing to a familial or genetic predisposition and age 25-70 years. Women could be tested at an age younger than 25 if they had a family history of breast cancer being diagnosed before the age of 30 years, since testing began at an age 5 years younger than that at which the youngest family member was found to have cancer. Women with symptoms of breast cancer or a personal history of breast cancer were excluded.	Results of both imaging examinations scored in a standardised way according to the BIRADS on a 5-point scale (1, negative; 2, benign; 3, probably benign; 4, suspicious abnormality; 5, highly suggestive of malignancy). The results of each exam were blinded so that the two examinations were not linked. When on of the imaging exams was a BIRADS 3 or 0 ('need additional imaging evaluation') further investigation by US with or without fine-needle aspiration was advised, or MRI or XRM was repeated. When one of the two exams was BIRADS 4 or 5 a cytologic or histologic evaluation of a biopsy specimen was performed. When the results of XRM and MRI were negative but the findings on CBE were rated as uncertain or suspicious additional investigations were also performed. The diagnosis of malignant tumours was based on the results of a histologic examination.	Cancer detection: 51 malignant tumours (44 invasive breast cancers, 6 DCIS and 1 non-Hodgkin's lymphoma) arose. 45 of the breast tumours were screen-detected and 5 were interval tumours. The figures were all calculated including the 5 interval tumours but excluding 5 tumours that did not have sufficient data. It is not possible to recalculate these without the interval tumours as it is not clear, once stratified what groups they would be in. Mode of detection: 3 tumours were detected by CBE 21 tumours were detected by MRI at BIRADS cut off of 4 and 32 at a BIRADS cut off of 3. Sensitivity (95% CI): CBE 6.7% (1.4 to 18.3%) 'suspicious' CBE 17.8% (8 to 32%) 'probably benign' MRI 46.7% (31.7 to 62.15) BIRADS 4 MRI 71.1% (55.7 to 83.6%)BIRADS 3 Specificity (95% CI): CBE 99.9% (99.8 to 99.9%) 'suspicious' MRI 98.9% (98.6 to 99.2%) BIRADS 4 MRI 89.8% (88.8 to 90.7%) BIRADS 3 PPV: CBE 50% (11.8 to 88.2%) 'suspicious' MRI 32.3% (21.2 to 45.0%)BIRADS 4 MRI 7.15 (4.9 to 9.8) BIRADS 3	Of the 1,952 women included, 8 withdrew from the study before the first visit and another 35 were excluded because they ultimately proved not to be carriers in a family with a proven mutation, and therefore had a less than 15% lifetime risk of developing breast cancer. Of the 1,909 remaining women, 88 (4.6%) left the study or were lost to surveillance before October 2003. 65 of these 88 women underwent prophylactic mastectomy. Another 89 women (4.7%) remained under surveillance but later refused surveillance by MRI because of claustrophobia or other reasons. Area under ROC curve was significantly higher for MRI than for XRM, indicating that MRI surveillance could better discriminate between malignant and benign cases. Inclusion of only invasive cancer: the difference between sensitivity of MRI and mammography was even greater than the difference overall. Of the 20 cancers not detected by XRM or CBE, 11 of the 19 invasive tumours were smaller than 10mm and only 1 was associated with a positive node Larger tumours (>2cm diameter) were found more often in women with BRCA1, BRCA2, PTEN, and TP 53 mutations than in the other 2 risk groups in the study, suggesting that more frequent surveillance is needed in these two groups.

Table 19. Primary research studies appraised investigating the accuracy and efficacy of MRI surveillance compared to usual care on outcomes from breast cancer (continued)

Source	Study design Evidence grading	Comparison interventions and dates of testing	Sample	Outcomes and verification	Results	Comments
Kriege et al. 2004 NEJM *Continued*		.		One of the investigators, an expert pathologist, reviewed all the biopsy specimens that formed the basis for the diagnosis of breast cancer	NPV: CBE 98.9% (98.6 to 99.2%) 'suspicious' MRI 99.4% (99.1 to 99.6%) BIRADS 4 MRI 99.6% (99.4 to 99.8%) BIRADS 3 Area under ROC: MRI 0.827 Tumour characteristics: There were 44 invasive tumours and 6 DCIS. The number of tumours less than 10mm in size was a significantly higher in the study cohort than in symptomatic women (not receiving surveillance) in both the National Cancer Registry control group ($p<0.001$) and the genetic study control group ($p=0.04$). Lymph nodes were negative in 66.7% (28/42) of the study cohort. This was also significantly higher in the study cohort than the number of node negative tumours in the National Cancer Registry control group ($p<0.001$) and the genetic study control group ($p=0.001$).	Authors' conclusions: The surveillance programme used in this study, especially MRI, can detect breast cancer at an early stage in women at risk for breast cancer. However a drawback of MRI is that it has a lower specificity than XRM and as a result, MRI will generate more findings judged as uncertain, which require short-term follow-up or additional investigations. Reviewers' conclusions: A generally well conducted study with conclusions drawn from the data presented above, and the respective surveillance tests performed either on the same day or within a short period of the first surveillance test undertaken. The results for CBE and MRI suggest that MRI is more sensitive than CBE in the surveillance of women at high risk of breast cancer, but has equivalent specificity and NPV at BIRADS 4. Lowering the BIRADS cut-off to 3 increases the sensitivity of MRI but decreases the specificity and the PPV. This is due to a higher number of false- positive examinations that arise at a lower cut-off.

Table 19. Primary research studies appraised investigating the accuracy and efficacy of MRI surveillance compared to usual care on outcomes from breast cancer (continued)

Source	Study design Evidence grading	Comparison interventions and dates of testing	Sample	Outcomes and verification	Results	Comments
Kuhl et al. (2005b) Germany	Prospective Cohort study III-2 (C1 P2 Q2)	Surveillance protocol: Biannual CBE and US and annual XRM and MRI. If abnormalities found on CBE or US at round without XRM or MRI, these additional modalities were used to further investigate this. Surveillance commenced at 30 years, or 5 years before the youngest family member affected with the disease. (NB: in first 2 years, women under 30, or 30-39 years with dense breasts did not receive XRM, but this was subsequently abandoned and all women received XRM). Mammography (XRM): Annual conventional film screen XRM performed with at least 2 views per breast (mediolateral oblique and caudalcranial), obtained and interpreted in accordance with German radiological practice guidelines. Diagnoses coded according to the BIRADS diagnostic categories on a 5-point scale (1, negative; 2, benign; 3, probably benign; 4, suspicious abnormality; 5, highly suggestive of malignancy).	Sample no = 529 (out of 590 eligible women; 49 were lost to follow-up after 1 surveillance round and 12 were also excluded as they had a clinical abnormality at initial examination). Inclusion criteria: • asymptomatic women • personal history of breast cancer included provided that the patient had not undergone bilateral mastectomy, had not received chemotherapy within the previous 12 months and had no metastases (139 women were included with a personal history of breast cancer). Exclusion criteria: • clinical signs of breast cancer; • chemotherapy within the previous 12 months; • women having undergone bilateral mastectomy.	Relevant outcomes: cancer detection; mode of detection; tumour size; tumour stage; node status; interval tumours; sensitivity; specificity; PPV; NPV. Verification of a positive result was achieved by histology (for positive imaging studies). Verification of a negative result was achieved by follow-up (for negative imaging studies. If a breast cancer was identified clinically (by palpation) between surveillance rounds or at the 6-month clinical visit the imaging studies of the previous round were considered false-negative. Mean follow-up was 5.3 years (range 2-7 years). A total of 1,542 annual surveillance rounds were completed.	Cancer detection: A total of 43 breast cancers were identified in 41 patients (11 of these women had a prior history of breast cancer), 40 were said to be detectable by imaging. However, the figures may be misleading as they do not correlate with the interval cancer rate and in some cases may refer to imaging after an interval cancer arose. Mode of detection: CBE identified only one tumour, also detected by imaging. MRI identified 39 tumours. Tumour size, stage and node status: Of the 39 tumours detected by MRI, 31 were invasive and 8 were *in situ.* The invasive tumours had a mean size of 12.4mm and five of them were node-positive.14 invasive cancers were detected by MRI that were not detected by XRM or US; these had a mean size of 9mm and none were node-positive.	Limitations included: CBE and the imaging studies were performed within a time frame of 8 weeks. Few sample characteristics presented, such as OCP or HRT use, or the use of preventative strategies such as tamoxifen or BSO. Verification bias is likely. Unclear documentation of interval tumours. Lack of blinding to the results of the CBE. Author's conclusions The authors' conclusions relate to the surveillance strategy as a whole, including CBE and US. If US is used in combination with XRM, it can help compensate for some but by far not for all of the shortcomings of XRM, and it causes a substantial number of false positive diagnoses. If MRI is used for surveillance, XRM proved to be of limited and ultrasound of no additional value. US may however be useful to bridge the relatively long time interval between annual surveillance rounds. Propose that in view of the insufficient diagnostic accuracy of XRM and USS, that breast MRI should be considered an integral part of surveillance programmes for women at high familial risk in particular in documented carriers of pathogenic BRCA mutations Reviewer's conclusions Similar to those of the authors above. MRI is the most effective surveillance modality, especially in women in the highest risk group. MRI maintains good sensitivity in all risk groups. The limitations of this study must be taken into account in the interpretation.

Table 19. Primary research studies appraised investigating the accuracy and efficacy of MRI surveillance compared to usual care on outcomes from breast cancer (continued)

Source	Study design Evidence grading	Comparison interventions and dates of testing	Sample	Outcomes and verification	Results	Comments
Kuhl et al. (2005b) Germany *Continued*		Breast MRI: Standard dynamic axial contrast-enhanced breast MRI of both entire breasts was performed on a 1.5T system (NT/INTERA; Philips, Best, the Netherlands) after injection of 0.1mmol/kg body weight gadopentetate dimeglumine (Magnevist, Schering, Berlin, Germany) Ultrasound (US): performed with 7.5-13MHz probes (Siemens Elegra, GE logic 500 and ATL HDI 5000; Siemens, Erlangen, Germany). The entire breast was systematically examined by the physician who interpreted the study. Diagnoses were scored on a 5-point scale identical to the XRM BIRADS categories. Each imaging study was read and scored independently by a different radiologist who had substantial experience with the respective imaging technique. The readers were informed about the clinical findings from CBE and the risk	• Recruited from high risk clinics in a single gynaecology department Risk Stratification: According to definition of the Consortium on Familial Breast and Ovarian Cancer of the German Cancer Aid, corresponding to a lifetime risk of breast cancer of at least 20% (two or more cases of breast cancer on the same side of the family, including at least two cases with onset before age 50 years, or with breast or ovarian cancer, irrespective of age, families with at least one case of breast cancer diagnosed before 35 years, families with three or more cases of breast cancer on the same side of the family, and women who met the criteria for high familial risk, irrespective of the result of mutational analysis). In women without a personal history of breast cancer the Claus tables were also used to quantify risk.	Verification of last surveillance round was by continued surveillance in 428 women, telephone interview in 52 women and for 6 women who had prophylactic mastectomy it was by pathology of the specimen. XRM: BIRADS of 4 or 5, biopsy was recommended irrespective of finding in US or MRI. BIRADS 3 was managed by 6-months follow-up until receiving a BIRADS 2 or biopsy clarification. US categorised as BIRADS 3 managed by short-term (6 months) US follow–up. BI-RADS 4 or 5 managed by US-guided biopsy (14G, semi-automatic or automatic biopsy gun) except for the following constellation: if an US finding that was suspicious was clearly benign on XRM or MRI no biopsy was performed. MRI: Suspicious scores (4 or 5) were managed by magnetic resonance-guided biopsy.	Interval tumours: The paper states that 40 out of 43 tumours in this cohort were detected by imaging. However, a sentence in the discussion states that the rate of interval cancers was 2% in this cohort. This translates to 10 tumours if it is 2% of the population or 1 tumour if it is 2% of the number of tumours detected overall. The latter is more likely but it is unclear. There is no indication of which risk group these interval tumours were detected in. Comparisons: When stratified by risk groups, the detection rates at both the prevalent and incident rounds were much higher in the mutation carriers than the other 2 risk groups, but these differences are not statistically significant. Sensitivity (95% CI): CBE 2.3% (0.1 to 12%) MRI 90.7% (77.9 to 97.4%) n = 39/43 When stratified by risk groups , MRI maintains good sensitivity across all risk groups. A sensitivity of 100% is documented for each risk group (but the denominator for calculating this is smaller than the overall number of women, 34 instead of 43 – how this figure is arrived at is unclear).	

Table 19. Primary research studies appraised investigating the accuracy and efficacy of MRI surveillance compared to usual care on outcomes from breast cancer (continued)

Source	Study design Evidence grading	Comparison interventions and dates of testing	Sample	Outcomes and verification	Results	Comments
Kuhl et al. (2005b) Germany *Continued*		status of the patient but were blinded to the results of the respective other imaging modalities. Comparisons are made between the 3 risk groups and the different modalities of surveillance. Dates of study were February 1996 to February 2002.	Women were then stratified into 3 risk groups for analysis: mutation carriers; high lifetime risk (20-40%); moderate lifetime risk (20%);	Findings categorized as BI-RADS 3 short-term follow-up after 6 months was recommended with further management corresponding to that of XRM BIRADS 3 lesions. BIRADS 3 categories in all imaging that received short-term follow-up were not considered positive for the calculation of outcomes. Invasive cancer and DCIS were considered a malignant diagnosis but LCIS and atypical ductal hyperplasia were considered to be benign.	Specificity (95% CI): XRM 96.8% (95.7 to 97.7%) *n* = 1364/1409 MRI 97.2% (96.2 to 98.0%) *n* = 1370/1409 Stratification by risk group does not appear to affect the specificity. PPV (95% CI): XRM 23.7% (14 to 37%) *n* = 14/59 MRI 50% (38.4 to 61.5%) *n* = 39/78 The PPV increases with the increasing risk of breast cancer, this will be affected by the higher incidence in women at higher risk. NPV (95% CI) XRM 97.9% (97.0 to 98.6%) *n* = 1364/1393 MRI 99.7% (99.2 to 99.9%) *n* = 1370/1374	

Chapter 7: Accuracy and efficacy of MRI and mammography

SECONDARY RESEARCH: STUDY DESIGNS AND QUALITY

The search strategy identified two relevant reviews that examined the effectiveness of MRI surveillance for breast cancer. The methods and conclusions are described in **Table 20 (pages 194-195)**. As discussed in the methodology chapter, the papers may not have employed the same inclusion and exclusion criteria as have been applied in this review and therefore the results must be interpreted with caution.

The report by Mark et al. (2003) was considered by the United States-based Technology Evaluation Center as part of its assessment programme. The objective of the review was to evaluate the effectiveness of MRI of the breast for the surveillance of asymptomatic women thought to be at high genetic risk of breast cancer because of the confirmed presence, or high risk of (due to presence in relatives), BRCA1 or BRCA2 mutations or a pattern of breast cancer history in multiple first-degree relatives, often occurring at a young age and with bilaterality. The review was based on a search to November 2003. Although five studies were included, only two studies on a total of 301 women provided data on sensitivity and specificity of MRI surveillance (Kuhl et al. 2000; Warner et al. 2001). In these studies, the reference standard included an eventual positive or negative histological diagnosis in patients who had positive MRI or mammography, and an interval of clinical follow-up to detect false-negative readings. Both studies showed that the sensitivity of MRI was 100 per cent and that of mammography was 33 per cent. Specificities of MRI were 95 per cent and 91 per cent in Kuhl et al. (2000) and Warner et al. (2001) respectively. Kuhl et al. (2000) reported specificity of 93 per cent for mammography while the corresponding specificity of mammography was found in Warner et al. (2001) to be 99.5 per cent. However, as stated by Mark et al. (2003) the comparisons are based on a very small number of cancer cases detected in each relatively small study. Another limitation of the data was that the results of the studies only represented the yield of usually one annual round of surveillance or two rounds at the maximum. There were no data available on the effectiveness of MRI surveillance for patient-related outcomes such as tumour size, stage or mortality. Mark et al. (2003) concluded that the findings of two reasonably performed comparative studies demonstrated probable superiority and definite non-inferiority of MRI in sensitivity for detecting breast cancer in high genetic risk women when compared with mammography.

The purpose of the systematic review by Irwig et al. (2004) was to examine the accuracy of proposed new technologies for breast cancer surveillance, including ultrasound, MRI, full-field digital mammography and computer-aided detection. The review was based on a search of MEDLINE to December 2002, and four studies were identified in which MRI was evaluated. Studies were included that reported data on both sensitivity and specificity. The use of MRI surveillance for breast cancer was evaluated in four studies, which evaluated the test in women at high risk of cancer on the basis of genetic mutations or a family history. In all four studies, the technology was contrast-enhanced MRI and a dedicated breast coil was used in all studies. There were less than 40 cancers diagnosed across all four studies which included a total of 576 women. The results suggested that MRI was more sensitive than mammography but may have a lower specificity with false positive rates of 5-9 per cent. These results should be interpreted with some caution as the numbers of cancers found were very small and the data yield was in most cases from a short time of surveillance i.e. usually one annual round.

Table 20. Secondary research appraised relevant to accuracy and efficacy of MRI surveillance

Source	Search method	Criteria for inclusion/exclusion	Results	Comments
Mark et al. 2003	Search: MEDLINE (via PubMed) 1966-November 2003 and restricted to English language publications, and studies using human subjects. Manual searches of recent issues of pertinent journals, reviewing reference lists of well-known papers, and by contacting known experts in the field. Databases searched: MEDLINE, Current Contents, Key words: magnetic resonance imaging, high risk, screening, breast neoplasm, genetic.	Population: Women considered at high risk of breast cancer because of confirmed presence of BRCA1 or BRCA2 mutation, high risk of aforesaid mutation due to known presence of the mutation in relatives, pattern of breast cancer in multiple first-degree relatives often occurring at a young age and with bilateralism, consistent with a high probability of harbouring BRCA mutations or other hereditary breast cancer. Family history would include multiple individuals with breast or ovarian cancer with breast cancer occurring at a young age. Intervention: MRI of the breast. Outcomes: Sensitivity, specificity, or lacking confirmation of false-negatives, detection rates and referral rates, patient-related outcomes. Inclusion criteria: Study using MRI of the breast to screen women, selection of patients at high risk for breast cancer as outlined above, criteria for selection to have MRI were to be determined prospectively so as to produce unbiased estimates of test performance among patients meeting specified criteria of that particular study, use of an appropriate reference standard such as histological confirmation and/or clinical follow-up, reporting of sensitivity and specificity, or lacking conformation of false negatives, reported detection rate and referral rate. Exclusion criteria: Not explicitly stated.	Five studies were identified. The inclusion criteria varied between studies. The small numbers of cancers detected in the studies precluded any way of determining whether the results of MRI surveillance are dependent or vary depending on the definition of high risk. Calculations of sensitivity and specificity were made in two studies (Kuhl et al. 2000 and Warner et al. 2001), and the reference standard included eventual positive or negative histological findings diagnosis in patients who have positive tests and interval of clinical follow-up to detect false-negative readings. These two studies had a total of 388 women participating. The sensitivity of MRI was 100% in both Kuhl et al. (2000) and Warner et al. (2001) and 33% for mammography in both studies. Specificity of MRI was 95% (91/96) in Kuhl et al. (2000) and 91% in Warner et al. (2001) In three studies in which detection rates between MRI and mammography are presented, the detection rate is higher for MRI (2.8-17%) than for XRM (0 or were not reported). The recall rates were higher for MRI (8.3%-10%) than for XRM (4%)	The results of all the included studies represent the yield of only a limited period of surveillance, usually one or at the most, two rounds of annual surveillance. In one study (Tilanus-Linthorst et al. 2000) patients had mammograms 6 months before MRI. Patients were selected based on mammogram results and the delay between each surveillance procedure makes it uncertain whether the cancers were missed cancers or interval cancers. In only two studies (Kuhl et al. 2000 and Warner et al. 2001) were the methods and results fully reported. In the two studies reporting sensitivity/specificity results, only Kuhl et al. (2000) reported blinding the test assessors. The results for sensitivity and specificity are from a very small number of cancers detected in each study (9 in one study and 6 in the other). Author's conclusions: The findings of reasonably performed comparative studies demonstrate probable superiority and definite non-inferiority of MRI in terms of sensitivity for detecting breast cancer in high risk women. The specificity was equal to specificity in the study by Kuhl et al. (2000) but worse in the other studies. Reviewer's conclusions: Generally a well conducted systematic review and with conclusions that reflect the data presented in the report.

Table 20. Secondary research appraised relevant to accuracy and efficacy of MRI surveillance (continued)

Source	Search method	Criteria for inclusion/exclusion	Results	Comments
Irwig et al. 2004	Search: MEDLINE 1966-December 2002, and extended by examining the references given in relevant primary studies and review articles contact with content experts and targeted further MEDLINE searches , e.g. on authors of earlier studies. 649 papers were identified Databases searched: MEDLINE. Key words: breast neoplasms, sensitivity and specificity, mass screening, CAD, CT scan, MRI, MRS, PET, electrical impedance, digital mammography, FFDM, scintimammography, optical mammography.	Population: Women asymptomatic for breast cancer, including populations at higher risk because of genetic predisposition or those in whom mammography may be less accurate because they are younger or have radiologically dense breast tissue. Intervention: 'New' tests for screening asymptomatic women for breast cancer. Outcomes: Sensitivity and specificity. Inclusion criteria: Papers were included only if they included 'new' tests for the detection of breast cancer in asymptomatic women. Exclusion criteria: Studies concerning the development of the test, use of the test in individual cases or as a diagnostic tool in women with a clinically or mammographically detected breast abnormality.	MRI: Examined in four studies which evaluated the test in women at high risk of cancer, usually on the basis of genetic mutations or a family history of breast cancer. In all studies the technology was contrast-enhanced MRI and all studies used a dedicated breast coil. There were less than 40 cancers in all the studies (in one study the number of cancers was not described but thought to be between 3-10 cancers). The review authors state that the results suggest that MRI is more sensitive (100%) than mammography (33-46%) in selected populations but may have higher rates of false-positive results requiring biopsy (5-9%) compared to mammography (1-7%).	Narrow range of databases searched. Not clear that data extraction and review of internal validity of the included studies was undertaken. It appears that the conduct and reporting of the studies included in the review was limited and the populations were too small to allow adequate precision of in measures of specificity and sensitivity. Data was not provided on interval cancers. The test was evaluated in consecutive participants in only one of the four included trials. Author's conclusions: Although some of proposed new tests appear to be promising, there was a need for larger and better quality of studies of new technology starting soon after introduction to allow concurrent evaluation and implementation. Reviewer's conclusions: Although the studies included in the review were of questionable quality, the review is under-reported and therefore it is difficult to draw any conclusions about the effectiveness or accuracy of MRI surveillance for breast cancer.

PRIMARY RESEARCH

Study design

The search identified 10 eligible primary research studies, including eight prospective cohort studies and two retrospective cohort studies. Four studies were conducted in single centres (Kuhl et al. 2005b; Morris et al. 2003; Stoutjesdijk et al. 2001; Trecate et al. 2003). The remaining included studies were multi-centre studies undertaken in the United Kingdom (Leach et al. 2005), Canada (Warner et al. 2004) the United States (Hartman et al. 2004; Lehman et al. 2005), the Netherlands (Kriege et al. 2004), Germany (Kuhl et al. 2005b) and Italy (Podo et al. 2002). No studies were identified that compared MRI surveillance to no surveillance. Therefore, it is unknown at this stage whether there is any benefit in terms of survival and response to cancer treatment of surveillance with MRI in women at high risk of breast cancer over no surveillance of this population. Each of the included studies attempted to compare the diagnostic accuracy of MRI with mammography surveillance in women at high risk of breast cancer owing to family or genetic history. The studies by Kuhl et al. (2005b), MARIBS (Leach et al. 2005) and Warner et al. (2004) also examined the use of imaging techniques in combination.

Study setting

Stoutjesdijk et al. (2001) was undertaken in the Netherlands. The study by Trecate et al. (2003) took place in a clinic in Milan, Italy, and Morris et al. (2003) was undertaken at the Sloan-Kettering Memorial Cancer Centre in the United States. The prospective multi-centre cohort study by Podo et al. (2002) was undertaken in 12 centres in Italy, five of which were institutes of cancer and research treatment and the remaining seven were university general hospitals. The prospective cohort study by Hartman et al. (2004) was a single-centre study in which MRI was compared to mammography and ductal lavage at a cancer genetics clinic in Stanford, California. Clinical breast examinations were also performed but no data were reported for the results of the CBEs. The study by Kriege et al. (2004) was a multicentre prospective cohort study, in which MRI was compared to mammography for surveillance of women with a genetic or familial predisposition to breast cancer to determine whether surveillance with MRI facilitated early diagnosis of breast cancer. The study was conducted in six familial-cancer clinics in the Netherlands. Warner et al. (2004) recruited women from familial cancer clinics in southern Ontario and Montreal, Canada, although all surveillance tests were undertaken in one centre in Ontario. The MARIBS (Leach et al. 2005) study enrolled asymptomatic women at high risk for breast cancer in 22 familial breast cancer centres in the UK, and the study by Kuhl et al. (2005b) was undertaken at a medical school in Bonn, Germany. Lehman et al. (2005) was set in 13 research institutions, hospitals and clinics in the United States, Canada and Europe.

Below is an overview of study designs and aspects of quality represented by these studies. Full details of the papers appraised including methods, key results, limitations and conclusions are provided in Evidence **Table 25**. One study (Warner et al. 2004) examined mortality. None of the studies examined response to treatment of cancers diagnosed by the surveillance tests examined. However, the studies examined the size and node status of the tumours identified, which may act as a surrogate outcome in the evaluation of effectiveness of the surveillance test for the early detection of breast cancer in women at high risk of breast cancer because of genetic or family history.

Stoutjesdijk et al. (2001)

Study sample

The data in the retrospective cohort study by Stoutjesdijk et al. (2001) was collected from the radiology reports and pathology databases of the University Medical Centre St Radboud, Nijmegen, where annual breast MRI for women at risk of early-onset familial cancer has been practised since 1994. All reports of breast cancer surveillance with MRI and mammography between November 1994 and February 2001 were initially selected. The group whose results were to be analysed were refined on the following inclusion criteria: lifetime risk greater than 15 per cent based on family history or the presence of BRCA1 or BRCA2 germ-line mutation, no personal history of breast cancer, and adequate follow-up data had to be available for the confirmation of the radiological findings, either by follow-up

MRI or mammography at least 2 years later or confirmation of positive radiological findings by histology. A total of 179 women with an age range of 21-71 years received one or more surveillance screens with MRI or mammography during the study.

Interventions and comparators

The surveillance protocol was for CBE to be performed biannually and annual imaging by mammography and MRI. The MRI examination was carried out with a 1.5T system (Magnetom Vision; Siemens, Erlanger, Germany) with a standard bilateral dedicated breast coil. Examinations were done in all pre-menopausal women in the second week of the menstrual cycle to minimise glandular tissue enhancement.

After a pre-contrast series of images, contrast agent was given by injection. The scanning orientation was axial but was changed to coronal in 1999 to reduce artifacts in the axillary region due to motion of the heart. Mediolateral oblique and craniocaudal projections were obtained, plus magnification views in both projections if required. Both the MRI images and mammograms were classified with BIRADS (1, negative; 2, benign; 3, probably benign; 4, suspicious abnormality and 5, highly suggestive of malignancy). The radiologists interpreting the images were blinded to the results of the other imaging studies.

Outcomes

Outcomes examined by the investigators in Stoutjesdijk et al. (2001) were the numbers of cancers detected, mode of detection, tumour stage and node status, and the occurrence of interval cancers. Verification of a positive result was by histology within two months and verification of a negative result was by a minimum of two years of follow-up of MRI or mammography.

Thirteen malignant tumours were detected in 179 women. All the 13 tumours were imaged by MRI, and 6 out of the 13 were detected by mammography. Three tumours were DCIS (all risk categories 2, 3 or 2 and 3) and the remainder of the detected tumours were invasive (1 was a non-Hodgkins lymphoma). Two of the tumours were found in mutation carriers. Four tumours were node positive (all women in risk category 3) and the rest were negative. There were no interval cancers found in this study.

Podo et al. (2002)

Study sample

The prospective multi-centre cohort study by Podo et al. (2002) was undertaken in 12 centres in Italy, five of which were institutes of cancer and research treatment and the remaining seven were university general hospitals. Nine centres participated in the first phase of the trial, which is reported here. One hundred and five patients were enrolled in the first annual round (14 of these women also underwent a second round of surveillance). Forty women (38%) had a previous personal history of breast cancer. The mean age at recruitment was 46 years, with a median age of 51 years (age range 25-77 years). The mean age at diagnosis was 55.3, median 52.5 (range 35-70 years). Women were included who were known BRCA1 or BRCA2 mutation carriers, or had a 1 in 2 probability of being a carrier i.e. a first-degree relative who was a proven mutation carrier. Two women were also included whose families had a very high risk or incidence of breast cancer that was likely associated to non-BRCA1 and BRCA2 mutations.

Interventions and comparators

The surveillance protocol consisted of a clinical breast examination (CBE), mammography, US and MRI at yearly intervals. The MRI was performed using coronal and axial planes. Contrast enhancement was also used, with Gd-chelate (0.1 mmol/kg) injected. One pre-contrast and five post-contrast images were taken. Pre-menopausal women had MRI within the 2nd week of the menstrual cycle. MRI was reported using a system that is based on a combination of morphological and enhancement parameters (0-2 = benign, 3 = uncertain, 4-8 = malignancy). In the case of non-benign scores (3-8) which were detected only by MRI, the MRI was repeated after 1-2 months. If the lesion was confirmed, a biopsy was undertaken. For the mammograms, standard mediolateral oblique and craniocaudal views were obtained of each breast, and the findings reported using the BIRADS system.

It was not reported whether the radiologists interpreting the images were blinded to the results of the other imaging modalities.

Outcomes

The principal outcomes in Podo et al. (2002) related to the breast cancer detection rate, the mode of tumour detection, and tumour size, stage and node status. The verification of positive findings was by biopsy (either MRI or ultrasound-guided), and pathology. Verification of negative findings was through follow-up. The authors acknowledged that as this report was of a preliminary study, the follow-up was incomplete.

A total of eight tumours were detected in this study. Of these, five women had a previous personal history of breast cancer, three women were BRCA1 mutation carriers, three women were BRCA2 mutation carriers and the mutation status of two women was unknown. Of the eight tumours detected, seven were detected by MRI and one by XRM. Two each of the detected tumours were invasive ductal carcinomas, invasive lobular carcinomas and DCIS. In addition, two mixed tumours (one invasive ductal and lobular carcinoma and one DCIS and LCIS) were also found. No node-positive tumours were detected during surveillance. The tumour sizes ranged from 3-27mm. As follow-up was incomplete, there were no data reported on false negatives, therefore sensitivity and specificity data could not be calculated. There were also no data reported for the results of CBE.

This study by Podo et al. (2002) had some limitations. The sample size of 105 women was small and the patients enrolled were at very high risk of breast cancer. In addition, as the number of breast cancers was so small, it was difficult to come to any conclusions about the efficacy and accuracy of MRI in the detection of breast cancer. There was little information on the women selected and no information provided on selection, and the characteristics of women who declined to participate. The length of follow-up was also unknown. There was also no information provided on false-negatives as follow-up appears to be short-term.

Morris et al. (2003)

Study sample

Morris et al. (2003) was a retrospective review of the records of 367 asymptomatic women with normal findings on mammography who were at high risk of developing breast cancer due to personal history, lobular carcinoma *in situ*, atypia, or family history of breast cancer. Women received MRI scans at a median interval from XRM to MRI of 14 days (range 0-131 days). All the women had their first mammogram at the screening centre between January 2000 and December 2001.

Interventions and comparators

The participants had an MRI performed with the patient prone in a 1.5T commercially available system using a dedicated surface breast coil. Images were taken before and three times after a rapid bolus injection of 0.1mmol/L of gadopentetate dimeglumine delivered through an in-dwelling IV catheter. Image acquisition started after contrast material injection and saline bolus. Images were obtained sagitally for an acquisition time per volumetric acquisition of less than three minutes each. Total imaging time per breast, including three contrast-enhanced acquisitions, was approximately 20 minutes. Details of how the mammograms were taken are not reported in this study, which was a retrospective review of women at high risk of breast cancer who had negative results on mammography.

Outcomes

In Morris et al. (2003) the principal outcome was the frequency of recommending biopsy at the first MRI screening. Breast MRI examinations were interpreted by breast imaging specialists in conjunction with clinical history and other breast imaging studies such as mammograms and sonograms when available. The findings on MRI were reported using the BIRADS scale. MRI-detected lesions referred for biopsy included masses with spiculated or irregular margins, irregular shape or heterogeneous or rim enhancement and non-mass lesions showing linear or segmental enhancement. Other lesions were referred for biopsy at the discretion of the interpreting radiologist in conjunction with clinical history and other imaging studies.

A total of 64 women were referred for a non-palpable lesion detected on MRI in the 367 women who underwent surveillance (17%, 95% CI 14-22%). Biopsy was performed in 59 women, which revealed cancers that were occult on mammography and clinical breast examination in 14 women (24%, 95% CI 14-37%) comprising 4 per cent (95% CI 2-6%) of the 367 women who had breast MRI surveillance. The average size of the lesions that underwent biopsy was 1.0cm (range 0.4-5.9cm). Cancer was identified in 16 lesions in 14 women including 20 per cent (95% CI 12-31%) of the 79 lesions that were studied under biopsy. Among these 16 cancers, 10 (63%) were DCIS and six cancers (38%) were invasive cancers. However, the length of follow-up post-surveillance is unclear.

Trecate et al. (2003)

Study sample

In a small study, Trecate et al. (2003) enrolled 23 women at high risk of breast cancer on the basis of being a BRCA1 BRCA2 mutation carrier, or a 1 in 2 probability of being a carrier, on the basis of a positive genetic test in a close relative, or being at a high risk for breast cancer according to criteria specified relating to family history.

Interventions and comparators

MRI was performed annually for all ages for two years during the study. One pre-contrast image and five post-contrast images were taken. The contrast agent was Gd-DTPA at 0.1mmol/kg. In addition, CBE was performed every six months for all ages. The women received surveillance using mammography annually and commenced when the women were 25 years old with bilateral one-view, and then increased to bilateral double-view for women aged 30 years and above. Double-view was performed in craniocaudal and mediolateral oblique projections. One view was performed in the mediolateral oblique projection for younger women. Annual ultrasound scanning was performed alone from 20-25 years, then with mammography from 25-35 years, then six months after mammography from 35-40 years and above 40 years only if requested by the radiologist. The methods for interpreting the findings of each surveillance test were not presented; nor is it clear when the tests took place in relation to each other. It was not reported whether the radiologists interpreting the images were blinded to the results of the other imaging modalities.

Outcomes

The principal outcomes in this study were breast cancer detection rate, the mode of tumour detection, and tumour size, stage and node status.

Four breast cancers were detected overall, all of which were detected by MRI. None of the tumours were detected by mammography.

Hartman et al. (2004)

Study sample

In Hartman et al. (2004), 41 women all had an initial surveillance screen and an additional 15 had had a second surveillance screen and were included in the 2005 paper. The median age of the enrolled women was 42.5 years (range 27-72 years). The inclusion criteria were: that the women had to have either documented BRCA1 or BRCA2 mutations, or a greater than 10 per cent risk of developing breast carcinoma at 10 years. In addition, the women had to be older than 25 years, or 5 years younger than the earliest age at which a relative was diagnosed with breast cancer. Of the enrolled women, 22 women had a known BRCA1 mutation and six had a known BRCA2 mutation. The remaining 18 women were all classified as being of high risk owing to family history. Twelve women had a previous history of breast cancer and three women had a previous history of ovarian malignancy. Eleven of the 41 women (28.6%) had had previous bilateral oophorectomy and/or were on Tamoxifen at the time of the initial screen. Women were included who had a previous history of breast cancer, however they had to be one year post-completion of adjuvant therapy.

Interventions and comparators

CBE was performed biannually and mammography, breast MRI and ductal lavage were performed on an annual basis. Enrolment began in September 2001 and reported accrual ended in May 2003 in the 2004 paper. All examinations initially had to be carried out within an eight-week period, although this was subsequently shortened to two weeks after 2002. All mammograms were centralised for the last year of the trial to aim for consistency and quality control. The report does not comment on the views taken or the system of interpretation for the mammograms. MRI examinations were done unilaterally, one for each breast, 1-3 days apart. They were timed according to the menstrual cycle. The contrast agent was 0.1mmol/kg of gadolinium, and the interpretation criteria were said to be tailored to each patient's history and imaging findings. The radiologist performing the assessments of the MRI results was not blinded to the genetic status of the patient

Outcomes

The principal outcomes in Hartman et al. (2004) were the numbers of cancers detected and the mode of detection. Positive test results were verified through pathology results and verification of negative results was through follow-up. However, the mean follow-up time was not reported.

There were no invasive tumours detected during this study, but there was one high grade DCIS (identified in a BRCA1 mutation carrier) which was detected by MRI but not by mammography. However, there were 25 MRI results classified as 'abnormal' in the first surveillance round, of which 11 had a biopsy (MRI- guided). Fourteen lesions were followed up at six months. High-risk lesions that were surveillance detected by MRI in three women included radial scars and atypical lobular hyperplasia. There was no detailed information on the results of mammography and CBE presented in Hartman et al. (2004).

Kriege et al.(2004)

Study sample

The study by Kriege et al. (2004) was a multi-centre prospective cohort study, in which MRI was compared to mammography for the surveillance of women with a genetic or familial pre-disposition to breast cancer to determine whether surveillance with MRI facilitated early diagnosis of breast cancer. The study was conducted in six familial-cancer clinics in the Netherlands. Surveillance consisted of a clinical breast examination performed by an experienced physician every six months and imaging studies (mammography and MRI) performed annually by experienced radiologists. A total of 1,909 women, who had a cumulative lifetime risk of breast cancer of 15 per cent or more, received surveillance. The mean age of these women was 40 years (range 19-72). Within the group of 358 carriers of pathogenic mutations, 276 women had BRCA1 mutation, 77 women had a BRCA2 mutation, one woman had BRCA1 and BRCA2 mutations, two women had a PTEN mutation and two women had a TP53 mutation. Women with symptoms of breast cancer or a personal history of breast cancer were excluded.

Interventions and comparators

Both imaging investigations were performed on the same day or in the same time period, between days 5-15 of the menstrual cycle. For the mammography, oblique and craniocaudal views and if necessary, compression views or magnifications of the breast were taken. The MRI was undertaken according to a standard protocol with the use of gadolinium-containing contrast medium. Results of both imaging examinations were scored in a standardised way according to the BIRADS 5-point scale (1, negative; 2, benign; 3, probably benign; 4, suspicious abnormality; 5, highly suggestive of malignancy).

Outcomes

The primary outcomes in Kriege et al. (2004) were measures of test accuracy including sensitivity, specificity and positive predictive values of each surveillance test. The results of each exam were blinded so that the two examinations were not linked. If one of the imaging exams was a BIRADS 3 or 0 ('need additional imaging evaluation') further investigation by ultrasound with or without fine-needle aspiration was advised, or the MRI or mammogram was repeated. When one of the two exams was BIRADS 4 or 5 a cytologic or histologic evaluation of a biopsy specimen was performed.

The mean follow-up period was 2.7 years. During this time, 51 malignant tumours were detected (44 invasive breast cancers, six DCIS and one non-Hodgkin's lymphoma). A total of 32 breast cancers was detected by MRI. Conversely, 13 cancers were missed by MRI (including five DCIS, four interval cancers and one tumour which was detected only by CBE). The sensitivities of XRM and MRI for detecting invasive cancer were 40 per cent and 71.1 per cent respectively and the specificity was 95 per cent and 89.8 per cent respectively. To evaluate the discriminating capacity of the imaging methods, receiver operating curves were generated. The area under the curve was 0.686 for mammography and 0.827 for MRI; the difference between the areas was 0.141 (95% CI 0.02 to 0.262, $p<0.05$). However, mammography had a higher sensitivity than MRI for detecting DCIS (83% vs. 17% for MRI) ($p=0.02$).

Warner et al. (2004)

Study sample

Warner et al. (2004) recruited women from familial cancer clinics in southern Ontario and Montreal, Canada, although all surveillance tests were undertaken in one centre in Ontario. Women were included if they were a BRCA 1 or BRCA2 mutation carrier. Exclusion criteria were a past history of unilateral breast cancer if the contralateral breast was not intact, history of bilateral breast cancer, currently undergoing chemotherapy or known to have metastatic disease. For technical reasons, women weighing more than 91kg were also excluded. The participation of pregnant or lactating women was also deferred. There were 236 participants in this study.

Interventions and comparators

Surveillance consisted of biannual CBE and annual XRM, US and MRI, all performed on the same day. Surveillance commenced at least one year after the woman's last mammogram. The CBE results were coded as normal, suggestive of benign disease, indeterminate, or suspicious of malignancy. Indeterminate CBE exams were repeated after 3 months. The MRI was performed with a 1.5 Tesla magnet. The first 38 patients in the first year were done in a single-turn elliptical coil after a bolus injection of 0.1mmol/kg of Gd-DTPA. Images were taken in the coronal plane. For the remaining patients, a phased-array coil arrangement was used which provided sagittal images. The results of the MRI were read and scored independently by a radiologist and scored on the 5-point BIRADS scale. All lesions with a score of 4 or 5 were biopsied.

Outcomes

Sensitivity and specificity of CBE, XRM, MRI and US were the primary outcomes of interest. Other relevant outcomes included the cancer detection rate, tumour stage, size and node status, interval cancers and mortality. The verification of positive screens was by pathology, biopsy was undertaken if there was suspicion from any of the four screening modalities, while the verification of a negative screen was through follow-up. All patients were followed up for a minimum of one year from the date of the last screening examination.

Twenty-two cancers were found in 21 women (one woman had bilateral cancer). Seven of these women had previous breast cancer). Two cancers were detected by CBE (9.1%), eight by mammography (36%), seven by ultrasound (33%) and 17 by MRI (77%). Seven cancers (32%) were detected by MRI alone, two cancers (9.1%) were detected by mammography alone, and two cancers were detected by ultrasound alone. In Warner et al. (2004) six of the detected tumours were DCIS and 16 were invasive (15 infiltrating ductal and 1 invasive lobular). The mean size of the invasive tumours was 11mm at the first surveillance round and 13mm at the second round (overall range 5mm-60mm). Fifteen cases were node sampled and two were node-positive. There was only one interval cancer, detected in a 40 year old BRCA1 mutation carrier seven months after her 3rd surveillance round (retrospectively this tumour was visible on MRI and on mammography at last surveillance visit). Another woman, who elected to have a bilateral mastectomy after breast cancer was found, had a 2mm focus of DCIS in the contralateral breast which had not shown up at surveillance two months earlier. All 22 patients who had tumours diagnosed were still alive and disease-free at the time the article was written. It appeared that the cancers detected on the second surveillance round were of an earlier stage. The two node-positive tumours were detected in the first surveillance round. However, it was not clear whether the first surveillance round was really a prevalent round as a high percentage of women had had prior mammography.

After the first round of surveillance, 16.5 per cent of participants underwent a diagnostic MRI scan to clarify the status of an indeterminate or possibly suspicious lesion. This rate of referral for a second MRI decreased at the second and third rounds of surveillance to 9.6 per cent and 7.1 per cent respectively. For an additional 7.6 per cent of patients, a 6-month follow-up MRI was recommended for lesions that remained indeterminate and this rate decreased at the second and third rounds of surveillance to 2.9 per cent and 2.4 per cent respectively. A total of 2.1 per cent of CBEs were thought to be suspicious at the first round of surveillance and CBE was repeated three months later. The corresponding rate for the second and third years of surveillance was 0.4 per cent for CBE. This is potentially suggestive of a learning effect whereby those reading the MRI scans become more skilled as a result of increased experience.

Kuhl et al.(2005b)

Study sample

The 529 study participants in Kuhl et al. (2005b) were recruited by the 'high-risk' clinics in one department of gynaecology between February 1992 and February 1996. Women were recruited if they were clinically asymptomatic and met the criteria of high familial risk corresponding to a lifetime risk of at least 20 per cent. Inclusion criteria were: two or more cases of breast cancer on the same side of the family, including at least two cases with onset before age 50 years, or with breast or ovarian cancer, irrespective of age, families with at least one case of breast cancer diagnosed before 35 years, families with three or more cases of breast cancer on the same side of the family, and women who met the criteria for high familial risk, irrespective of the result of mutational analysis. Personal history of breast cancer was no exclusion provided that the patient had not undergone bilateral mastectomy, had not received chemotherapy within the previous 12 months and had no metastases. Women were excluded from the study if they had clinical signs of breast cancer, chemotherapy within the previous 12 months, and those having undergone bilateral mastectomy.

Interventions and comparators

The protocol was biannual CBE and US, and annual XRM and MRI. Surveillance consisted of each of the imaging tests being performed within a time frame of eight weeks. Each imaging study was read and scored independently by a different radiologist, who had substantial expertise with the respective breast imaging technique. Although the radiologists were informed about the clinical findings from a CBE and the risk status of the patient, they were blinded to the results of the respective other imaging modalities. Mammography was performed with at least two views per breast (mediolateral oblique and caudalcranial), obtained and interpreted in accordance with German radiological practice guidelines. The diagnoses were coded according to the BIRADS diagnostic categories on a 5-point scale (1, negative; 2, benign; 3, probably benign; 4, suspicious abnormality; 5, highly suggestive of malignancy). For the breast MRI, standard dynamic axial contrast-enhanced breast MRI of both entire breasts was performed on a 1.5T system after injection of 0.1mmol/kg body weight (Gd-DTPA).

Outcomes

The primary outcomes in Kuhl et al. (2005b) were the test accuracy measures of sensitivity, specificity and positive predictive values for each of the imaging modalities used alone or in various combinations. In between annual surveillance rounds, half-yearly CBE and breast ultrasound were performed. Validation of the imaging results was achieved either by histology (for positive imaging studies) or by follow-up (for negative imaging studies). The BIRADS scale was used to classify the results of the imaging examinations. If a breast cancer was identified clinically by palpation between surveillance rounds or at the 6-month clinical visit the imaging studies of the previous round were considered false-negative. For any of the surveillance tests classified as a BIRADS score of 4 or 5, biopsy was recommended irrespective of findings of the other surveillance tests. Each biopsy was guided by whichever surveillance modality detected the lesion. Where a BIRADS score of 3 was recorded on any of the imaging tests, women were managed by 6-months follow-up until receiving a BIRADS 2 score or biopsy clarification.

In Kuhl et al. (2005b), women were observed for a mean observation period of 5.3 years (range 2-7 years). In the entire cohort of 529 patients, a total of 43 breast cancers were identified in 41 patients. There were 34 invasive cancers and nine DCIS. One tumour was detected by CBE, but none of the 43 cancers were diagnosed by clinical breast examination alone. Of the 40 cancers diagnosed by imaging

studies, 14 cancers were identified by mammography, and 39 cancers were detected by MRI. Forty cancers were diagnosed by mammography and MRI.

The sensitivity scores for mammography, and MRI were 33 per cent and 91 per cent respectively. The sensitivity of mammography in higher risk groups, (mutation carriers and risk of breast cancer 21-40%) was 25 per cent compared to 100 per cent for both groups when screened by MRI. The sensitivity of MRI plus mammography was 93 per cent. The specificity for MRI and MRI plus mammography were also high at 97.2 per cent and 96 per cent respectively.

The positive predictive value for MRI for all women regardless of risk status was 50 per cent (95% CI, 38% to 62%). Of the 19 cancers diagnosed only by MRI, 5 were high-grade intra-ductal cancers and 14 were invasive cancers with a median size of 7.5mm. Thirty-one cancers were identified in women without a previous history of cancer and 12 cancers in 11 women with breast cancer history. Of the latter 12 cancers, three cancers were classified as local recurrences and nine cancers occurred in the contralateral breast and/or were histologically categorised as second primary cancers. Two cancers were palpable at the time of diagnosis, one in the regular surveillance interval, and one was an interval cancer diagnosed between surveillance rounds. The remaining cancers were asymptomatic (not palpable).

Lehman et al. (2005)

Study sample

Lehman et al. (2005) compared the performance of surveillance with mammography versus MRI in women at genetically high risk for breast cancer. Women were included if they were at least 25 years of age and their lifetime risk was greater than 25 per cent based on family history or genetic test confirmation and screening of the contralateral breast if they had a prior history of breast cancer. Women diagnosed with breast cancer more than five years prior to the study were eligible for bilateral screening if the probability of breast cancer was greater than 50 per cent based on the study risk algorithm or positive test for a mutation in BRCA1 or BRCA2. A total of 390 asymptomatic women at high risk of breast cancer were enrolled. The exclusion criteria were any contraindications to MRI examination i.e. claustrophobia, pregnancy, pacemaker, magnetic aneurysm clip or implanted magnetic device, and women who presented with palpable lesions or mammographic abnormalities prior to risk assessment.

Interventions and comparators

The study protocol specified that the mammography and the clinical breast examination (CBE) were to be performed within 90 days of the MRI scan. For the MRI, images were taken before and after the bolus injection of contrast.

Outcomes

Lehman et al. (2005) compared the diagnostic yield of cancer from each of mammography and MRI. Any suspicious MRI enhancing lesions were described based on lesion shape, borders, distribution and internal architecture. The overall assessment was classified on a 5-point scale (1, negative; 2, benign; 3, probably benign; 4, suspicious abnormality; 5, highly suggestive of malignancy). A lesion was identified as malignant if there was a focal mass with irregular or speculated margins, if enhancement was in a ductal distribution if a solid lesion showed rim enhancement, or if there was intense regional enhancement in less than one quadrant. Benign lesions were identified as those that had smooth or lobulated margins with internal septations or if the mass was cystic. All lesions given an assessment score of 4 or 5 were recommended for biopsy. A retrospective review was also performed that included all images from both surveillance modalities from patients with cancers diagnosed during the study.

There were 390 women enrolled by 13 sites from July 1999 and January 2002 and 367 women completed all the study examinations. Imaging evaluations recommended 38 biopsies and 27 biopsies were performed, resulting in four cancers diagnosed. This produced an overall cancer yield of 1.1 per cent (95%CI 0.3-2.8%). MRI detected all four cancers whereas mammography detected one cancer. The diagnostic yield of mammography was 0.3 per cent (95%CI 0.01-1.5%) compared to 0.8 per cent

(95%CI 0.3-2.0%) for MRI. As a result of a positive MRI examination, 27 biopsies were performed. Four of 27 lesions biopsied were diagnosed as malignant and 23 lesions were diagnosed as either benign, atypical ductal hyperplasia (by excisional biopsy) or lobular carcinoma *in situ.* There were seven true-positives and 20 false-positives on MRI and a positive predictive value of MRI of 12.9% (95%CI 3.6%, 30%). Of the remaining recommendations for biopsy, 11 women declined biopsies after positive findings including six women who had BIRADS 4 assessments on MRI (with negative, benign, or probably benign mammograms) and three women who had BIRADS 4 assessments on mammography (with negative, benign, or probably benign MRI assessments). Two women declined biopsies based on a 'probably benign' MRI assessment.

MARIBS (Leach et al. 2005)

Study sample

A total of 649 women underwent surveillance with both mammography and MRI in the MARIBS study (Leach et al. 2005). Criteria for inclusion were: known carriers of a deleterious BRCA1, BRCA2 or TP53 mutation, first-degree relative of someone with BRCA1, BRCA2 or TP53 mutation, strong family history of breast or ovarian cancer, family history consistent with classic Li-Fraumeni syndrome. Women were excluded if they had previous breast cancer or any other cancer such that expected prognosis was less than five years. Participants who underwent genetic testing with a negative result, and women who developed cancer were excluded from further participation.

Interventions and comparators

The investigators in the MARIBS (Leach et al. 2005) study endeavoured to perform each surveillance test on the same day. Mammography and MRI were performed annually. Mammography examinations took place either in an accredited screening centre or in a family history clinic working to NHSBSP standards, and were either 2-view or 1-view. The clinics participating in the study were linked to suitable MRI facilities and radiological skills at the inception of the study MRI screening. The protocol for MRI comprised high spatial resolution sequences before and after contrast medium injection of 0.2mmol per kg bodyweight of gadopentetate dimelglumine (Gd-DTPA). Radiologists unaware of the results of the other tests reported the findings of each imaging test. However, it is not clear if the reporting radiologist was a first or second radiologist to interpret the results of the screening test.

Outcomes

The primary outcomes were measures of sensitivity and specificity of each imaging technique, including the sensitivity and specificity when both techniques were used together. A scoring system for MRI reporting was used based on morphological and dynamic contrast uptake characteristics previously validated against histology. A worksheet was also developed to ensure consistency of method in the choice of regions of interest and their analysis. The findings for the mammography examinations were also double reported. Patients recalled because of an indeterminate or suspicious test had either a high temporal resolution study with 0.1mmol per kg Gd-DTPA or a repeat of the initial screening protocol, done at a different phase of the menstrual cycle to the initial test. The reporting radiologist and the attending doctor decided the diagnostic pathway.

There were 1,232 surveillance intervals of 6-54 months duration with the median surveillance interval at 12 months. Among women with more than one surveillance round, there were 1,232 surveillance intervals of 6-54 months in length with a median at 12 months. Eighty-five percent of surveillance intervals were 10-14 months. There were 35 breast cancers diagnosed in 649 women who received surveillance with both mammography and MRI. There were 14 cancers detected by mammography and 19 cancers were detected by MRI. The remaining two cancers diagnosed in this population were interval cancers.

Sensitivity was significantly higher for MRI (77%, 95% CI 60-90) than for mammography (40%, 95% CI 24-58) (p=0.01). The sensitivity was 94 per cent (95% CI 81, 99) when MRI and mammography was used in combination. Specificity was lower for MRI at 81 per cent (95% CI 80-83), compared to mammography (93%, 95% CI 92-95) and this difference was statistically significant (p<0.0001). However when MRI was combined with mammography, specificity was lower at 77 per cent (95% CI 75-79),

Subgroup analysis on women who were BRCA1 mutation carriers showed a pronounced difference in sensitivity between MRI and mammography (13 cancers: 92 per cent for MRI compared to 23 per cent for mammography). Although this difference was statistically significant, this finding must be interpreted with some caution as the study was not sufficiently powered to detect a difference in this particular subgroup of women.

The area under the receiver operator characteristic curve for MRI was 0.85 (95%CI 0.84, 0.87) compared to 0.70 (95%CI 0.68, 0.72) for mammography and this difference was statistically significant (p=0.035). Of the cancers detected by MRI alone, three were Grade 1 tumours, five were Grade 2 tumours and 11 lesions were Grade 3 tumours. Seven women had invasive ductal carcinoma and DCIS, nine women had invasive ductal cancer and two participants had invasive lobular cancer. Mammography detected six cancers that were not detected by MRI, and 19 cancers were detected by MRI that were not detected by mammography.

Summary

Two systematic reviews and 10 primary studies were identified. The systematic review by Mark et al. (2003) was generally well conducted. However, only two primary studies were identified that reported any data pertaining to accuracy. The sensitivity of MRI in both the included studies in this review was found to be 100 per cent and compared to 33 per cent for mammography. Specificities of MRI were 95 per cent and 91 per cent in Kuhl et al. (2000) and Warner et al. (2001). One study reported specificity of 93 per cent for mammography (Kuhl et al. 2000) while the corresponding specificity of mammography was found in the other included study (Warner et al. 2001) to be 99.5 per cent. However, as stated by Mark et al. (2003) the comparisons are based on a very small number of cancer cases detected in each relatively small study. Another limitation of the data was that the results of the studies only represented the yield of usually one annual round of surveillance or two rounds at the maximum. There was no data available on the effectiveness of MRI surveillance for patient-related outcomes such as tumour size, stage or mortality. Mark et al. (2003) concluded that the findings of two reasonably performed comparative studies demonstrated probable superiority and definite non-inferiority of MRI in sensitivity for detecting breast cancer in high genetic risk women.

No studies were found that compared MRI surveillance in women at high risk of breast cancer owing to genetic predisposition or family history, with no surveillance at all. It is posited that the early detection of cancerous lesions in this particular population, and early treatment, may confer benefits in terms of mortality reduction. However there is no data to suggest that this is true for cancerous lesions diagnosed by MRI.

Of the 10 primary studies, eight were prospective cohort studies and two were retrospective cohort studies. A total of 4,428 women received surveillance in the 10 studies. There was some heterogeneity of risk level among the women included. In Warner et al. (2004) only women with BRCA genetic mutations were included, in MARIBS (Leach et al. 2005), women were included of varying levels of risk. Warner et al. (2004), Kriege et al. (2004) and MARIBS (Leach et al. 2005) also excluded women with a personal history of breast cancer while women with a personal history of breast cancer were included in Kuhl et al. (2005b) and in Lehman et al. (2005), conditional in the latter study on the contralateral breast being imaged.

MRI was used in all the included studies and images were taken before and after the bolus injection of contrast enhancement. Data from the use of mammography as a surveillance test was also used in all the included studies. Although most of the included studies state that the women under study received a clinical breast examination, only four reported data relating to the accuracy of CBE as a surveillance test (Kriege et al. 2004; Kuhl et al. 2005b; Trecate et al. 2003; Warner et al. 2004). The MRI screening tests were performed annually; although in one study (Kriege et al. 2004) clinical breast examination was performed at six-month intervals. Therefore, the MRI screening results can only be compared with clinical breast examinations undertaken at approximately the same time as the MRI tests. Mammography was usually performed within stipulated timeframes from the MRI test in the studies that compared MRI to mammography. In MARIBS (Leach et al. 2005), the mammography test was performed on the same day as the MRI scan, while in Lehman et al. (2005), the comparator test was to take place within 90 days of the first test. In Kuhl et al. (2005b) tests could take place eight weeks apart, an interval which may preclude comparisons being made between these tests. In addition, the length of the optimal surveillance interval is unknown at this time.

Measures of diagnostic test accuracy (sensitivity, specificity and positive predictive values) were outcomes examined in the more recent studies (Kriege et al. 2004; Kuhl et al. 2005b; Leach et al. 2005; Warner et al. 2004). Data were also reported for particular subgroups such as women with BRCA mutations and also at varying levels of risk for breast cancer based on family history in Kuhl et al. (2005b) and MARIBS (Leach et al. 2005). There were few measures of effectiveness studied in the included trials. None of the studies examined survival outcomes, aside from Warner et al. (2004), who stated that all the women receiving surveillance were still alive at the time of writing. In addition, there were no data presented on response to treatment as a result of the possible earlier diagnosis of cancerous tumours by MRI surveillance.

MRI vs. mammography

The results in terms of sensitivity data for the included studies are summarised in **Table 21** below. The table includes only those studies for which enough data were provided for the calculation of sensitivity scores. This summary is for all women enrolled in the primary studies. Data relating to women who carried BRCA mutations is presented in **Table 22**. Surveillance with MRI appears to be associated with substantially higher sensitivity than for mammography in terms of detecting cancers in women at high risk of breast cancer owing to familial or genetic history. However the results are based on a relatively small number of cancers detected so should be interpreted with some caution.

Table 21. Summary of results for studies comparing the sensitivity of MRI surveillance to mammography surveillance

Study	Sample	Screening modality	Number of cancers detected/total cancers	Sensitivity (95% CI)	Statistical significance
Kriege et al. (2004)	1909	MRI XRM	32/45 18/45	71.1%* (56 to 84%) 40.0%*(95%CI 26-56)	Not tested
Warner et al. (2004)	236	MRI XRM	17/22 8/22	77% (95%CI 55-92) 36% (95%CI 17-59)	p=0.02
Kuhl et al. (2005b)	529	MRI XRM	39/43 14/43	90.7% (95% CI 78-97) 32.6% (95%CI 19-48)	Not tested
MARIBS (Leach et al. 2005)	649	MRI XRM	27/35 14/35	77% ((95%CI 60-90) 40% (95%CI 24-58)	p=0.01

* Sensitivity calculated for cut-off of BIRADS 3

Data relating to the sensitivity scores for carriers of BRCA mutations is outlined below in **Table 22**. The difference in sensitivity scores between MRI and mammography is particularly pronounced in BRCA carriers in the studies where this has been examined. The MARIBS study (Leach et al. 2005) included data here for women who had a first-degree relative with a confirmed positive genetic test for a mutation in the BRCA1 gene. Given that these women have a higher absolute risk in the age-range studied in MARIBS (Leach et al. 2005) than the other risk groups MRI screening might be particularly useful in this group. In women with BRCA2 mutations, the gain was smaller and not statistically significant. The data in the Kuhl et al. (2005b) study relating to mutation carriers is also useful but should be interpreted with some caution as the results are based on a very low number of cancers detected.

Table 22. Summary of results for studies comparing the sensitivity and specificity of MRI surveillance to mammography surveillance in women with BRCA mutations

Study	Sample	Screening modality	Number of cancers detected/total cancers	Sensitivity (95% CI)	Statistical significance
Kuhl et al. (2005b)	529	MRI XRM	8/8 2/8	100% (63-100) 25% (3-65)	Not tested
MARIBS (Leach et al. 2005)	120	MRI XRM	27/35 14/35	92% (64-100) 23% (5-54)	p=0.0001

A summary of results is also presented in **Table 23** for the specificity in the trials in which these calculations could be made. This table includes the results for all women enrolled in the trial. Data relating to carriers of BRCA mutations is outlined above in **Table 22**. In the MARIBS (Leach et al. 2005) study, the specificity score for MRI was lower than for mammography. From this it appears that surveillance with MRI is associated with a higher number of false-positive results compared to mammography, which has implication for resource use or, the increase in resource use in following up false-positive results with other examinations such as biopsy. In addition, there are also issues around psychological and emotional issues such as increased anxiety for the women with positive test results given that due to their family history, may have witnessed close family members develop and be treated for breast cancer.

However, in the other studies in which MRI was examined alongside mammography, the specificity of MRI was relatively high, although in all cases lower than that for mammography. It may be that the increase in MRI scans being examined by the investigators over time resulted in a decrease in false-positive results at later time periods within the studies. For example, in the study by Warner et al. (2004) after the first round of surveillance, 16.5 per cent of participants underwent a diagnostic MRI scan to clarify the status of an indeterminate or possibly suspicious lesion. This rate of referral for a second screen by MRI decreased at the second and third rounds of screening to 9.6 per cent and 7.1 per cent respectively. For an additional 7.6 per cent of patients, a 6-month follow-up MRI was recommended for lesions that remained indeterminate and this rate decreased at the second and third rounds of surveillance to 2.9 per cent and 2.4 per cent respectively. A total of 2.1 per cent of CBEs were thought to be suspicious at the first round of surveillance and CBE was repeated three months later. The corresponding rate for the second and third years of surveillance was 0.4 per cent for CBE. This is potentially suggestive of a learning effect and the availability of previous films for comparison whereby those reading the MRI scans become more skilled as a result of increased experience.

Table 23. Summary of results for studies comparing the specificity of MRI surveillance to mammography surveillance

Study	Sample	Screening modality	Specificity (95% CI)	Statistical significance
Kriege et al. (2004)	1909	MRI XRM	89.8%-99.9%%* 95-100%*	Not tested
Warner et al. (2004)	236	MRI XRM	95.4% (92.9-97.2) 99.8% (98.7-99.9)	
Kuhl et al. (2005b)	529	MRI XRM	97.2 (96.2-98.0) 96.8% (95.7-97.7)	Not tested
MARIBS (Leach et al. 2005))	649	MRI XRM	81% (80-83) 93% (92-95)	p=0.01

* Specificity calculated for a range of BIRADS cut-offs, lower range BIRADS 3 and higher range BIRADS 5

In the studies by Kriege et al. (2004) and MARIBS (Leach et al. 2005), the area under receiver operator characteristic (ROC) curves was studied. The scores for mammography and MRI were remarkably similar in the two studies, and in both studies the difference between mammography and MRI reached statistical significance. In Kriege et al. (2004) the area under the ROC for mammography 0.686, and the corresponding value for MRI was 0.827. The difference between scores was 0.141 (95%CI 0.02-0.262, $p<0.05$). In MARIBS (Leach et al. 2005) the area under the receiver operator characteristic curve for MRI was 0.85 (95%CI 0.84, 0.87) compared to 0.70 (95%CI 0.68, 0.72) for mammography and this difference was statistically significant ($p=0.035$). This would appear to indicate that in both these studies, MRI surveillance could better discriminate between those with breast cancer and those without breast cancer.

In conclusion, MRI appears to be more sensitive than mammography for the detection of breast cancers in women at high risk of breast cancer owing to genetic or family history. The increase in sensitivity is particularly noticeable in women who carry mutations in BRCA1. However, the specificity of MRI is lower than that of mammography, which has implications for resource use and anxiety of those undergoing surveillance.

MRI vs. MRI plus mammography

Two prospective cohort studies were identified (Kuhl et al. 2005b; Leach et al. 2005) in which the combination surveillance protocol utilising both MRI and mammography was compared to MRI used alone. A total of 1,128 women underwent surveillance in these two studies which were generally well conducted. However there were no statistical tests calculated by these investigators. The inclusion criteria in the two studies were broadly similar although MARIBS (Leach et al. 2005) excluded women with a personal history of breast cancer while women with a personal history of breast cancer were included in Kuhl et al. (2005b).

MRI was used in all the included studies and images were taken before and after the bolus injection of contrast enhancement. Data from the use of mammography as a surveillance test were also used in all the included studies. In MARIBS (Leach et al. 2005), the mammography test was performed on the same day as the MRI scan, while in Kuhl et al. (2005b) tests could take place eight weeks apart. It is unknown at this time if that interval between tests is too wide, given the aggressive nature of the tumours found in women with BRCA mutations and other women with family history placing them at high risk of breast cancer.

Measures of diagnostic test accuracy (sensitivity, specificity and positive predictive values) were the principal outcomes examined in the trials (Kuhl et al. 2005b; Leach et al. 2005). None of the studies examined survival outcomes. In addition, there were no data presented on response to treatment as a result of the possible earlier diagnosis of cancerous tumours by MRI surveillance.

The principal results in terms of test accuracy are summarised below in **Table 24**. There appears to be little difference in the sensitivity and specificity in the Kuhl et al. (2005b) trial. However, in the study by MARIBS (Leach et al. 2005) there appears to be a substantial increase in sensitivity compared to using MRI when both MRI and mammography are used together. This is a result that should be interpreted with some caution, as the increase in cancers detected with the use of MRI and mammography is by only four cancers. Indeed all the results are based on small numbers of cancerous lesions being detected. There is little difference in specificity when MRI and mammography are used together in this study, a result that is consistent with that of Kuhl et al. (2005b).

In conclusion, two studies were identified that examined the comparison between MRI surveillance and a combination of MRI and mammography surveillance. There was no evidence in terms of improved survival due to the early detection of cancerous breast lesions. The sensitivities were high in both studies, suggesting that each surveillance regimen is efficacious for detecting tumours in women at high risk of breast cancer owing to family or genetic predisposition. However, it is not clear whether combination surveillance offers any additional benefit over screening with MRI alone. There was little difference in the specificity in each of the trials between MRI alone and MRI plus mammography.

Table 24. Summary of results for MRI plus mammography surveillance compared to MRI surveillance

Study	Sample	Screening modality	Number of cancers detected/total cancers	Sensitivity (95% CI)	Specificity (95%CI)	Statistical significance
Kuhl et al. (2005b)	529	MRI MRI + XRM	39/43 40/43 ,	90.7%(77.9 -97.4) 93%% (80.9-98.5)	97.2%(96.2-98) 96.1%(94.9-97)	Not tested
Kuhl et al. (2005b) data for mutation carriers	43 mutation carriers	MRI MRI +XRM	8/8 8/8	100%(63-100) 100%(63-100)	97.5% (N/R) 94.4% (N/R)	Not tested
MARIBS (Leach et al. 2005)	649	MRI MRI + XRM	27/35 33/35	77% (95%CI 60-90) 94% (95%CI 81,-99)	79% (95%CI 75-83) 77% (95%CI 75-79)	Not tested
MARIBS (2005) (for data on carriers of BRCA1 mutations	82 with BRCA1 mutation	MRI MRI + XRM	12/13 12/13	92% (95%CI 64-100) 92% (95%CI 64-100)	79% (95%CI 75-83) 74% (95%CI69-78)	Not tested

N/R=not reported

Table 25. Primary research studies appraised investigating the accuracy and efficacy of MRI surveillance compared to mammography on outcomes from breast cancer

Source	Study design Evidence grading	Comparison interventions and dates of testing	Sample	Outcomes and verification	Results	Comments
Podo et al. (2002) Italian multi-centre study	Prospective cohort Study III-2 (C1 P2 Q2)	Surveillance protocol: CBE, mammography, US and MRI at yearly intervals. Mammography: standard mediolateral oblique and craniocaudal views were obtained of each breast. Further views taken when necessary. Findings reported using the BIRADS system (1, negative; 2,benign; 3,probably benign; 4, suspicious abnormality and 5, highly suggestive of malignancy) US: performed at a frequency of ≥7.5MHz MRI was performed on coronal and axial planes. One pre-contrast and 5 post-contrast images were taken. Gd-chelate (0.1 mmol/kg) was injected as contrast. MRI was reported using a system that is based on a combination of morphological and enhancement parameters. (0-2 = benign, 3=uncertain, 4-8=malignancy). In the case of non-benign scores (3-8) which were detected only by MRI, the MRI was repeated after 1-2 months	Sample no = 105 patients were enrolled in the first annual round (14 of these women also underwent a second round). Forty (38%) had a previous personal history of breast cancer. Mean age at recruitment 46 yrs, median age 51 years (age range 25-77 years) Mean age at diagnosis was 55.3 years, median 52.5 (range 35-70 years) Recruited from 9 cancer genetics centres within Italy. Inclusion criteria: • Very high risk of breast cancer; • women ≥25yrs age • men ≥50yrs age; • women who had personal history of breast cancer were allowed if unilateral; • Unilateral mammography done if had had a mastectomy and bilateral if had had breast conservation;	Relevant outcomes: Cancer detection rate. Mode of detection Tumour size, stage and node status. Verification of positive findings is by biopsy (either MRI or US guided) and pathology. Verification of negative findings is through follow-up. It is acknowledged that these are preliminary findings and the follow-up is incomplete.	Cancer detection rate: 8 tumours were detected in total, 7 in the prevalent round and 1 in the incident round. 5 of these patients had a previous personal history of breast cancer, 3 were BRCA1 mutation carriers, 3 were BRCA2 mutation carriers and 2 with unknown mutation status. Mode of detection: 7 (88%) were detected by MRI. mammography detected only 1 tumour. (MRI had 1 false positive but mammography had none) Tumour size, stage and node status: 2 invasive ductal carcinomas 2 invasive lobular carcinomas 1 invasive ductal and lobular carcinoma 2 DCIS 1 DCIS and LCIS Tumour size ranged from 3-27mm. There were no node-positive tumours. The follow-up is incomplete and therefore sensitivity and specificity cannot be calculated.	Limitations include: Only the preliminary report of this study. Verification bias, particularly in this study (as acknowledged by the authors) as it is just a preliminary report and sufficient follow-up of negative results has not yet been achieved. This cohort varies from other studies as it is a very high risk group and includes a high proportion of women with a personal history of breast cancer. No comment on women undertaking risk reducing strategies such as on Tamoxifen or having had a bilateral salpingo-oophorectomy. Authors' conclusions: The findings of this study substantiate those of existing studies, that MRI is a more sensitive and more accurate imaging modality than conventional imaging for detecting breast cancer in women at a high risk of this disease (both pre and post menopausal women). A previous personal history of breast cancer was associated with higher probability of breast cancer detection during surveillance. The authors conducted a review of other existing literature and perform a meta-analysis of the results of the studies to date. They note that there are considerable differences in the design of these studies, but state that there are some consistent conclusions. The overarching finding is that MRI is more sensitive and significantly more accurate than conventional imaging in the surveillance of women at a high risk of breast cancer.

Table 25. Primary research studies appraised investigating the accuracy and efficacy of MRI surveillance compared to mammography on outcomes from breast cancer (continued)

Source	Study design Evidence grading	Comparison interventions and dates of testing	Sample	Outcomes and verification	Results	Comments
Podo et al. (2002) Italian Multi-centre study *Continued*		If the lesion was confirmed then a biopsy was undertaken. Pre-menopausal women had MRI within the 2nd week of the menstrual cycle. Dates of surveillance: June 2000 to March 2002. Preliminary report of first phase, 21 months, of the study.	• if on HRT, were included but this was stopped and surveillance not started until been off it for 3 months. Exclusion criteria: • pregnancy; • breast feeding; • current chemotherapy; • terminal illness; • specific contraindications to MRI; Risk stratification: Only recruited subjects who were known BRCA1 or BRCA2 mutation carriers, or had a 1 in 2 probability of being a carrier (first-degree relative who was a proven mutation carrier). 2 women also included whose families had a very high risk or incidence of breast cancer that was likely associated to a non BRCA1 or BRCA2 mutation. Forty of the 105 women also had a personal history of breast cancer.			They point to the need for more extensive, multi-centre and multi-national trials on the evaluation of benefits and costs associated with the introduction of MRI into appropriate surveillance programmes specifically addressed to subjects at high genetic risk of breast cancer. Reviewers' conclusions: This study does appear to show an advantage of MRI surveillance in women at high risk of breast cancer. However, these are only preliminary results of this study and measures of accuracy could not be calculated without further follow-up data. Unfortunately, a further report of this work cannot be found and it is perhaps ongoing. These results are also limited in their external validity by being from a very high risk cohort, especially as a high proportion of women with a personal history of breast cancer were included. The pulling together of results from other studies was hampered by variation in the design of the studies and also the outcomes measured.

Table 25. Primary research studies appraised investigating the accuracy and efficacy of MRI surveillance compared to mammography on outcomes from breast cancer (continued)

Source	Study design Evidence grading	Comparison interventions and dates of testing	Sample	Outcomes and Verification	Results	Comments
Trecate et al. (2003) Italy (NB: Podo is an author on this one as well but we cannot find any further reports from the Podo et al trial.)	Prospective cohort study III-2 (C1 P2 Q3)	Surveillance protocol: Outlined in full in the paper and was dependent on age group. CBE was performed every 6 months for all ages. Mammography was annual and commenced at 25 years with bilateral one-view, and then increased to bilateral double view from 30 years and above. Double-view was performed in craniocaudal and mediolateral oblique projections. One-view was performed in the mediolateral oblique projection for younger women. Annual US was performed alone from 20-25 years, then with mammography from 25-35 years, then 6 months after mammography from 35-40 years and above 40 years only if requested by the radiologist. US was performed with either7.5MHz or 10-12MHZ probes (ATL HDI 3500, Philips).	Sample no = 23 women at high risk of breast cancer (2 cases did not get US). No average age of women given, range was 30-61 years. Inclusion criteria: • BRCA1 or BRCA2 mutation carrier or 1 in 2 probability to be a mutation carrier (on the basis of positive mutational analysis in close relatives). With a negative or positive personal history for breast or ovarian cancer OR • High risk for breast cancer according to criteria specified in paper.	Relevant outcomes: Cancer detection rate. Mode of detection. Tumour size and stage. Verification of positive results was with pathology and verification of negative results was with follow-up. There is no mention of the mean length of follow-up.	Cancer detection: 4 breast cancers were detected overall. Mode of detection: All 4 tumours were detected by MRI. None of the tumours were detected by mammography. It is stated that there were no false-positives or false-negatives for MRI. Tumour size and stage: All 4 tumours were invasive: 2 ductal invasive carcinomas, 1 lobular invasive carcinoma and 1 which was mixed ductal and lobular. 2 occurred in mutation carriers and 2 in women at high risk through family history. Only 2 tumours had the size recorded and these were 10mm and 30mm. No record of nodal status was given. There was no mention of interval tumours.	Limitations include: Small sample size. There are few characteristics given of the women selected other then their risk assessment. There is no information on how they were selected and the characteristics of any women who did not agree to participate. There is no mention of mean age, reproductive history, exogenous hormone use or preventative strategies (i.e. Tamoxifen use or BSO). There is also no indication of which women were having prevalent or incident surveillance screens and for how long they were followed up in the study. There is likely verification bias and this is more likely, the shorter the follow-up period. Authors' conclusions: The authors' conclusions relate to the surveillance strategy as overall, including CBE and US. Breast MRI demonstrated to be a very useful technique for investigating breast disease. It is not influenced by breast density and does not use ionising radiation. For these reasons, it has been proposed to support mammography in the surveillance of BRCA mutated patients. Moreover, according to the reported results, breast MRI seems very helpful in the high-risk patients group.

Table 25. Primary research studies appraised investigating the accuracy and efficacy of MRI surveillance compared to mammography on outcomes from breast cancer (continued)

Source	Study design Evidence grading	Comparison interventions and dates of testing	Sample	Outcomes and verification	Results	Comments
Trecate et al. (2003) Italy *Continued*		MRI was performed annually for all ages for 2 years during the study. A Siemens Vision 1.5 was used with a dedicated double coil. One pre-contrast image and 5 post-contrast images were taken. The contrast agent was Gd-DTPA at 0.1mmol/kg. The method of interpreting the MRI or mammography is not presented. The study was conducted over a 7-month period; however the exact dates are not given.	Risk stratification: As above, either BRCA 1 or 2 carrier, 1 in 2 probability of being a carrier or >50% risk of carrying a susceptibility gene for familial breast cancer on basis of family history. The latter refers to at least 3 cases of breast cancer before 60 years of age, at least 3 cases of breast cancer before 60 years of age and ovarian cancer at any age, or at least 3 cases of breast cancer before 60 years of age and male breast carcinoma at any age. 5 of the women had a personal history of breast cancer, 1 for ovarian cancer and 1 for ovarian and breast cancer (1 had had a mastectomy, but the others had conservative surgery combined with radiation therapy).			We believe the breast MRI can be very useful within this kind of surveillance, with a less invasive approach to the disease. In the case of confirmed good diagnostic results, it could be proposed to be used every other year as an alternative to mammography. Reviewers' conclusions: This study suggests that MRI is a very effective tool for the surveillance of women at high risk of breast cancer. However, the sample is very small and it is difficult to know how long the women were followed up for and this would affect the reliability of the results. There could be false negatives that had not yet come to light. There is also a specific method of risk stratification in this study, which includes women with a personal history of breast cancer (although only if they are BRCA1 or BRCA2 mutation carriers), and this will affect the generalisability of the study. In addition the results are not presented in a very clear manner and it is difficult to determine the overall sensitivity and specificity for all the modalities of surveillance utilised, which would have been valuable information.

Table 25. Primary research studies appraised investigating the accuracy and efficacy of MRI surveillance compared to mammography on outcomes from breast cancer (continued)

Source	Study design Evidence grading	Comparison interventions and dates of testing	Sample	Outcomes and verification	Results	Comments
Morris et al. 2003	Retrospective cohort study III-2 (CX P2 Q2)	Retrospective review of 367 asymptomatic women with normal findings on XRM who were at high risk of developing breast cancer (personal history, lobular carcinoma *in situ*, atypia, or family history of breast cancer). Dates: January 1 2000-December 31, 2001. MRI: performed with the patient prone in a 1.5T commercially available system (Sigma, General Electric Medical Systems, Milwaulkee, WI) using a dedicated surface breast coil. The imaging sequence included a followed by a sagital fat-suppressed T2 weighted sequence (TR/TE 4000/85. A T1-weighted three-dimensional, fat-suppressed fast spoiled gradient-echo sequence (17/2.4; flip angle 35^{o}; bandwidth 31.25 Hz) is then performed before and three times after a rapid bolus injection of 0.1mmol/L of gadopentetate dimeglumine (Magnevist, Berlex, Wayne, NJ) per kg of bodyweight delivered through an in-dwelling IV catheter.	367 asymptomatic women with normal findings on XRM who were at high risk of developing breast cancer (personal history, lobular carcinoma *in situ*, atypia, or family history of breast cancer	Breast MRI exams were interpreted by breast imaging specialists in conjunction with clinical history and other breast imaging studies such as mammograms and sonograms when available. Levels of suspicion reported using the BIRADS scale. MRI-detected lesions referred for biopsy included masses with speculated or irregular margins, irregular shape or heterogeneous or rim enhancement and non-mass lesions showing linear or segmental enhancement. Other lesions were referred for biopsy at the discretion of the interpreting radiologist in conjunction with clinical history and other imaging studies.	Biopsy was recommended for a non-palpable lesion detected on MRI in 64 women, 17% (95%CI 14-22%) of the 367 women who underwent surveillance. Biopsy was performed in 59 women. Biopsy revealed cancer that was occult on XRM and CBE in 14 women, 24% (95%CI14-37%) of the 59 women who had biopsy and 4% (95%CO 2-6%) of the 367 women who had breast MRI surveillance. Biopsy was performed for 79 MRI-detected lesions in 64 women (average 1.2 lesions per woman, range 1-3 lesions per woman). The average size of the lesions that underwent biopsy was 1.0cm (range 0.4-5.9cm). Cancer was identified in 16 lesions in 14 women including 20% (95%CI 12-31%) of 79 lesions that had biopsy. Among these 16 cancers 10 (63%) were DCIS and 6 (38%) were infiltrating cancer.	Median interval from XRM to MRI, in the 59 women who had biopsies, was 14 days (range 0-131 days) and 58 (98%) women had normal findings on XRM within three months of breast MRI. PPV scores presented only for biopsy. Not all results referred for biopsy received the same treatment e.g. different methods of guided biopsy were used. Author's conclusions: Among women with a high risk of developing breast cancer, breast MRI led to a recommendation for a biopsy in 17%. Cancer was found in 24% of the women who underwent biopsy and in 4% of women who had breast MRI surveillance. More than half of the MRI-detected cancers were DCIS. Reviewers' conclusions: Little information presented to draw any accurate conclusions about the accuracy of the MRI surveillance for breast cancer.

Table 25. Primary research studies appraised investigating the accuracy and efficacy of MRI surveillance compared to mammography on outcomes from breast cancer (continued)

Source	Study design Evidence grading	Comparison interventions and dates of testing	Sample	Outcomes and verification	Results	Comments
Morris et al. 2003 *Continued*		Image acquisition started after contrast material injection and saline bolus. Images were obtained sagitally for an acquisition time per volumetric acquisition of less than 3 minutes each. Total imaging time per breast including three contrast-enhanced acquisitions was approximately 20 minutes.		For non-palpable, mammographically occult MRI-detected lesions warranting biopsy, correlation US was recommended at the discretion of the radiologist interpreting the MRI examination if it was thought that the lesion might be evident on US and amenable to US-guided biopsy. If the lesion was not seen on USS, MRI-guided needle localisation for surgical excision was performed.		

Table 25. Primary research studies appraised investigating the accuracy and efficacy of MRI surveillance compared to mammography on outcomes from breast cancer (continued)

Source	Study design Evidence grading	Comparison interventions and dates of testing	Sample	Outcomes and verification	Results	Comments
Hartman et al. (2004) California, USA	Prospective cohort Study III-2 (C1 P2 Q2)	Surveillance protocol: Biannual CBE and annual mammography, breast MRI and ductal lavage. All examinations initially had to be carried out within an 8-week period; this was shortened to 2 weeks since 2002. All mammograms were centralised for the last year of the trial to aim for consistency and quality control. The report does not comment on the views taken or the system of interpretation for the mammograms. MRI examinations were done unilaterally, 1 for each breast, 1-3 days apart. They were timed according to the menstrual cycle. A 1.5 telsa imager was used (Signa LX; General Electrical Medical Systems) with a dedicated 4-coil phased-array breast coil. The contrast agent was 0.1mmol/kg of gadolinium. The interpretation criteria were said to be tailored to each patient's history and imaging findings; a rough guide is given in the paper. Radiologist was not blinded to the genetic status of the patient.	Sample no = 41 women all had an initial surveillance screen, Median age = 42.5 years (range 27-72 years) (data from 2004 paper – not updated in 2005 paper). Recruited from a cancer genetics clinic. Inclusion criteria: Documented BRCA1 or BRCA2 mutation OR >10% risk of developing breast carcinoma at 10 years (based on Claus model); >25 years, or 5 years younger than the earliest age at which a relative was diagnosed with breast cancer. Risk stratification was done by the Claus model (all family history factors). (BRCAPRO was also used but not for the inclusion criteria). 22 women had a known BRCA1 mutation and 6 had a known BRCA2 mutation. The other 18 were all assessed as high risk. 12 patients had a previous history of breast cancer and 3 had a previous history of ovarian malignancy.	Relevant outcomes: Cancer detection. Mode of detection. Verification of positive results was through pathology results and verification of negative results was through follow-up. The mean follow-up time was not given.	Cancer detection: There were no invasive tumours detected during this study, but there was 1 high grade DCIS (BRCA1 mutation carrier) which was detected by MRI and missed by mammography. However, there were 25 abnormal MRIs in the first surveillance round, of which 11 had a (MRI guided) biopsy and 14 required 6 month follow-up MRI (data from 2004 paper, and is not updated in the 2005 paper, although still only 11 women had had biopsies). The rest of the biopsies included several high risk lesions, 2 atypical lobular hyperplasia (ALH) and 2 radial scars. There is no detailed information presented about mammography and CBE. The ductal lavage results are not presented here as this is not being considered in this review. Mode of detection: All lesions except for 1 case of (ALH) were detected by MRI and were not detected by mammography. The ALH case was detected by mammography alone.	Limitations: Small sample size Unclear if these women had received any prior surveillance, so whether the first round was a prevalent or incident screen. Risk stratification models do not take into account any preventative measures such as bilateral salpingo-oophorectomy or tamoxifen use. 25 women came into either or both of these categories and their risk may have been overestimated. (16 had undergone BSO and 7 had been on tamoxifen for at least 6 months, 2 had had BSO and were on tamoxifen)(*see note about change since 2004 paper) Likely verification bias Unclear follow-up time Authors' conclusions: Breast MRI identified high-grade DCIS and high-risk lesions that were missed by CBE and mammography. A larger trial is needed to determine which subgroups of high-risk women will benefit from this surveillance and whether the identification of malignant and high-risk lesions at an early stage will impact breast carcinoma incidence and mortality. The 2005 paper ends by saying that they are continuing this study and aiming to recruit over 500 women in the next 3 years to this protocol. Reviewers' conclusions: This study reports good information on the characteristics of the women in the sample, including Tamoxifen use and BSO, age and risk category.

Table 25. Primary research studies appraised investigating the accuracy and efficacy of MRI surveillance compared to mammography on outcomes from breast cancer (continued)

Source	Study design Evidence grading	Comparison interventions and dates of testing	Sample	Outcomes and Verification	Results	Comments
Hartman et al. (2004) California, USA *Continued*		Dates: enrolment began in September 2001 and reported accrual ended in May 2003 in the 2004 paper, and reported as accrual ending in December 2003 in the 2005 paper (another 5 women recruited in between). It says that the study had moved from one institute to another in this time and concludes by saying that the goal is to recruit more than 500 women in the next 3 years to this study.	25 of the 46 women (54%) had had previous bilateral oophorectomy (BSO) and/or were on Tamoxifen at the time of the initial surveillance. This has increased from 15 in the 2004 paper *(although the total number of women has only increased by 5, so some women already in the trial must have had a BSO or started Tamoxifen during this time). Women were included who had a previous history of breast cancer, however they had to be 1 year post-completion of adjuvant therapy and the previously affected breast did not undergo ductal lavage due to concerns over infection in these circumstances.	.		It suggests that MRI is an efficacious tool in the diagnosis of breast cancer and high-risk lesions in women at high risk of breast cancer, detecting lesions that were not detected by mammography or CBE. From the report it is difficult to determine the efficacy of the other technologies assessed (there is little mentioned on mammography and CBE). The overall cancer detection rate of MRI was lower than reported by other studies and the number of false-positives with MRI was higher. This was discussed by the authors and was felt to be as a result of the women whose risk had been reduced by BSO or Tamoxifen use. These women made up 54% of the sample and studies have shown considerable protective effect of these preventative measures. Therefore the efficacy of the MRI surveillance may have been underestimated by this study. Perhaps women with BSO or using Tamoxifen do not require such intensive surveillance as women at high risk who have not chosen these preventative measures. As the authors conclude, this will need further investigation in bigger, perhaps multi-centre, studies. Their continued research should aid this.

Table 25. Primary research studies appraised investigating the accuracy and efficacy of MRI surveillance compared to mammography on outcomes from breast cancer (continued)

Source	Study design Evidence grading	Comparison interventions and dates of testing	Sample	Outcomes and verification	Results	Comments
Kriege et al. 2004 NEJM	Prospective multicentre cohort study Grade III-2 (C1 P1 Q2)	Clinical breast examination: performed by an experienced physician every six months. Imaging studies performed annually by radiologists. XRM: oblique and cranio-caudal views and if necessary, compression views or magnifications. MRI: Dynamic breast MRI with gadolinium-containing contrast medium according to a standard protocol. Whenever possible, both imaging investigations were performed on the same day or in the same time period, between days 5-15 of the menstrual cycle.	1,909 women with a genetic risk for breast cancer. Mean age 40 years (range 19-72) Within the group of 358 carriers of pathogenic mutations, 276 had BRCA1 mutation, 77 had a BRCA2 mutation, 1 woman had BRCA1 and BRCA2 mutations, 2 had a PTEN mutation and 2 had a TP53 mutation. Inclusion criteria: Cumulative lifetime risk of breast cancer 15% or more owing to a familial or genetic predisposition and age 25-70 years. Women could be tested at an age younger than 25 years if they had a family history of breast cancer being diagnosed before the age of 30 years since testing began at an age 5 years younger than that at which the youngest family member was found to have cancer. Women with symptoms of breast cancer or a personal history of breast cancer were excluded.	Results of both imaging examinations scored in a standardised way according to the BIRADS s on a 5-point scale (1, negative; 2, benign; 3, probably benign; 4, suspicious abnormality; 5, highly suggestive of malignancy). The results of each exam were blinded so that the two examinations were not linked. When one of the imaging exams was a BIRADS 3 or 0 ('need additional imaging evaluation'), further investigation by US with or without fine-needle aspiration was advised, or MRI or XRM was repeated. When one of the two exams was BIRADS 4 or 5, a cytologic or histologic evaluation of a biopsy specimen was performed. When the results of XRM and MRI were negative but the findings on CBE were rated as uncertain or suspicious additional investigations were also performed.	Cancer detection: 51 malignant tumours (44 invasive breast cancers, 6 DCIS and 1 non-Hodgkin's lymphoma) arose. 45 of the breast tumours were screen detected and 5 were interval tumours. The figures were all calculated including the 5 interval tumours but excluding 5 tumours that did not have sufficient data. It is not possible to recalculate these without the interval tumours as it is not clear, once stratified what groups they would be in. Mode of detection 11 tumours were detected by XRM, and 21 by MRI at a BIRADS cut off of 4 and 18 by XRM and 32 by MRI at a BIRADS cut off of 3. Sensitivity (95% CI): XRM 24.4% (12.9 to 39.5%) BIRADS 4 XRM 40.0% (25.7 to 55.7%) BIRADS 3 MRI 46.7% (31.7 to 62.15) BIRADS 4 MRI 71.1% (55.7 to 83.6%)BIRADS 3 Specificity (95% CI): XRM 99.6% (99.4 to 99.8%) BIRADS 4 XRM 94.9% (94.3 to 95.6%) BIRADS 3 MRI 98.9% (98.6 to 99.2%) BIRADS 4 MRI 89.8% (88.8 to 90.7%) BIRADS 3 PPV: XRM 47.8% (26.8 to 69.4%) BIRADS 4 XRM 8.0% (4.8 to 12.3%) BIRADS 3 MRI 32.3% (21.2 to 45.0%)BIRADS 4 MRI 7.15 (4.9 to 9.8%) BIRADS 3	Of the 1,952 women included, 8 withdrew from the study before the first visit and another 35 were excluded because they ultimately proved not to be carriers in a family with a proven mutation and therefore had a less than 15% lifetime risk of developing breast cancer. Of the 1,909 remaining women, 88 (4.6%) left the study or were lost to surveillance before October 2003; 65 of these 88 women underwent prophylactic mastectomy. Another 89 women (4.7%) remained under surveillance but later refused surveillance by MRI because of claustrophobia or other reasons. Area under ROC curve was significantly higher for MRI than for XRM indicating that MRI surveillance could better discriminate between malignant and benign cases. Inclusion of only invasive cancer: the difference between sensitivity of MRI and mammography was even greater than the difference overall. Of the 20 cancers not detected by XRM or CBE, 11 of the 19 invasive tumours were smaller than 10mm and only 1 was associated with a positive node Larger tumours (>2cm diameter) were found more often in women with BRCA1, BRCA2, PTEN, and TP 53 mutations than in the other 2 risk groups in the study, suggesting that more frequent surveillance is needed in these two groups. Authors' conclusions: The surveillance programme used in this study, especially MRI, can detect breast cancer

Table 25. Primary research studies appraised investigating the accuracy and efficacy of MRI surveillance compared to mammography on outcomes from breast cancer (continued)

Source	Study design Evidence grading	Comparison interventions and dates of testing	Sample	Outcomes and verification	Results	Comments
Kriege et al. 2004 NEJM *Continued*		.		The diagnosis of malignant tumours was based on the results of a histologic examination. One of the investigators, an expert pathologist, reviewed all the biopsy specimens that formed the basis for the diagnosis of breast cancer	NPV: XRM 99.1% (98.8 to 99.4%) BIRADS 4 XRM 99.3% (99.0 to 99.5%) BIRADS 3 MRI 99.4% (99.1 to 99.6%) BIRADS 4 MRI 99.6% (99.4 to 99.8%) BIRADS 3 Area under ROC: XRM 0.686 MRI 0.827 Difference between AUCs was statistically significant at 0.141 (95% CI, 0.020 to 0.262, p <0.05). Tumour characteristics: There were 44 invasive tumours and 6 DCIS. The number of tumours less than 10mm in size was significantly higher in the study cohort than in symptomatic women not receiving surveillance in both the National Cancer Registry control group (p<0.001) and the genetic study control group (p=0.04). Lymph nodes were negative in 66.7% (28/42) of the study cohort. This was also significantly higher in the study cohort than the number of node-negative tumours in the National Cancer Registry control group (p<0.001) and the genetic study control group (p=0.001).	At an early stage in women at risk for breast cancer. However, a drawback of MRI is that it has a lower specificity than XRM and as a result, MRI will generate more findings judged as uncertain, which require short-term follow-up or additional investigations. Reviewers' conclusions: A generally well conducted study with conclusions drawn from the data presented above, and the respective surveillance tests performed either on the same day or within a short period of the first surveillance test undertaken.

Table 25. Primary research studies appraised investigating the accuracy and efficacy of MRI surveillance compared to mammography on outcomes from breast cancer (continued)

Source	Study design Evidence grading	Comparison interventions and dates of testing	Sample	Outcomes and verification	Results	Comments
Warner et al. (2004)Ontario and Montreal, Canada	Prospective cohort Study III-2 (C1 P2 Q2)	Study protocol: CBE biannually and mammography, US and MRI all performed annually 4 modalities all performed the same day. (commencing at least 1 year after the woman's last mammogram) CBE coded as normal, suggestive of benign disease, indeterminate, or suspicious of malignancy. Indeterminate CBE exams were repeated after 3 months. Mammography was conventional 4-view film. Further views done when necessary. MRI was performed with 1.5 T magnet (Signa, General Electrical Medical Systems). The first 38 patients in the first year were done in a single-turn elliptical coil after a bolus injection of 0.1mmol/kg of Gd-DTPA. Images were taken in the coronal plane. For the remaining patients, a phased-array coil arrangement was used. This provided sagital images. US used a 7.5MHz transducer (the first 7 patients did not receive US).	Sample no = 236 female BRCA1 and BRCA 2 mutation carriers. Mean age at first surveillance 46.6 years (range 25-65 years) Mean age of diagnosis was 47.4 years (33.4-63 years) Recruited from familial cancer clinics Inclusions: • BRCA 1 or 2 mutation carrier. Exclusions: • past history of unilateral breast cancer if the contra lateral breast not intact • pregnant or lactating women (participation deferred); • history of bilateral breast cancer, currently undergoing chemotherapy or known to have metastatic disease; • women weighing >91kg (technical reasons)	Relevant outcomes: cancer detection rate; mode of detection. tumour stage, size and node status. interval cancers. mortality; sensitivity; specificity; PPV; NPV; ROC curves. NB: the PPV and specificity do not include in the denominator women that had additional diagnostic studies that did not result in biopsy. Verification of positive results was by pathology. Biopsy was undertaken if there was suspicion from any of the four modalities of screening. Verification of a negative result was through follow-up. All patients were followed up for a minimum of 1 year from the date of the last surveillance examination.	Cancer detection: 22 cancers were detected in 21 women (1 bilateral). (7 of these women had previous breast cancer). Mode of detection: 8 by mammography (36%). 17 by MRI (77%). 7 cancers (32%) were detected by MRI alone, 2 cancers (9.1%) were detected by mammography alone; MRI detected 9 of the 12 cancers missed by conventional screening (mammography plus CBE). Tumour stage, size and node status: 6 tumours were DCIS and 16 were invasive (15 infiltrating ductal and 1 invasive lobular). The mean size of the invasive tumours was 11mm at the first round and 13mm at the second round (overall range 5-60mm). 15 cases were node sampled and 2 were node-positive. Interval cancers: There was only 1 interval cancer, detected, in a 40 year old BRCA1 mutation carrier 7 months after her 3rd surveillance screen (retrospectively this tumour was visible on MRI and on mammography at last surveillance)	Limitations include: Likely verification bias. Selected participants are very high risk, being proven mutation carriers and also including those with a prior history of breast cancer. It is not clear which were incident and which were prevalent rounds, and which tumours were detected at which round (a large number of women had had prior mammography). No mention of whether women had had risk reducing measures such as silateral salpingo oophorectomy or Tamoxifen. There was quite a high level of attrition in the study and the characteristics of those women are not outlined. This may have introduced bias. Authors' conclusions: The authors' conclusions relate to the surveillance strategy overall, including CBE and US. This study of BRCA mutation carriers demonstrates that the addition of annual MRI and US to mammography and CBE significantly improves the surveillance for detecting early breast cancers. The use of US did detect additional tumours, but had a high false-positive rate and in light of this its benefit remains to be seen. There was no observed benefit from CBE over and above the 3 imaging modalities.

Table 25. Primary research studies appraised investigating the accuracy and efficacy of MRI surveillance compared to mammography on outcomes from breast cancer (continued)

Source	Study design Evidence grading	Comparison interventions and dates of testing	Sample	Outcomes and Verification	Results	Comments
Warner et al. (2004) Ontario and Montreal, Canada *Continued*		Each imaging modality was read independently by a radiologist and scored on the 5 point BIRADS scale. All lesions with a score of 4 or 5 were biopsied. Pre-menopausal women had surveillance performed mid menstrual cycle to avoid changes due to cyclical hormonal variation. Radiologists were blinded to the results of CBE 31 women left the study before completing 3 rounds, 16 underwent bilateral mastectomy, 3 were too large for MRI machine, 3 stopped due to pregnancy, 4 developed metastatic cancers, 4 were lost to follow-up and 1 did not wish to continue participating. All participants underwent the first round, but only 58% the second and 36% the third (a total of 120 women were still undergoing surveillance when the paper was written).	Risk stratification not really performed as only BRCA mutation carriers included. (all very high risk group) There were 137 (58%) BRCA1 mutation carriers and 99 (42%) BRCA 2 mutation carriers. 31% were Ashkenazi Jews. 30% had a history of breast cancer, 9% a history of ovarian cancer and 60% had no history of cancer or a history of another type of cancer. 85% of the women (*n*=205) had had mammography within the last 15 months and therefore this was an incident rather than a prevalent round for them. 45% were pre-menopausal and 55% were post-menopausal.		Another woman, who elected to have a bilateral mastectomy after breast cancer was found, had a 2mm focus of DCIS in the contra lateral breast which had not shown up at surveillance 2 months earlier. Mortality: All 22 patients who had tumours diagnosed were still alive and disease-free at the time the article was written. It was felt that the cancers detected on the second round were of an earlier stage. The 2 node-positive tumours were detected in the first round. However, it was not exactly clear that the first round was really a prevalent round as a high percentage of women had had prior mammography. It was found that false-positives and false-negatives decreased from the first to the second and then to the third round of surveillance. This is especially seen for the false-positives in MRI, which decreased from 15 to 4 to 1. This may have been due to increasing experience in the radiologists in interpreting these scans. The measures of accuracy are therefore presented by the surveillance modality and by the year of the surveillance. These can be seen in the paper, but overall values for the 3 years are reported here.	MRI based surveillance is likely to become the cornerstone of breast cancer surveillance for BRCA1 and BRCA2 mutation carriers, but it is necessary to demonstrate that this surveillance tool lowers breast cancer mortality before it can be recommended for general use. Reviewers' conclusions: This study demonstrates a greater efficacy of MRI to XRM in the surveillance of BRCA1 and BRCA2 mutation carriers for breast cancer. As the authors suggest, this does not answer whether this translates into reduced mortality. However, the tumours detected did seem to be of an earlier stage and smaller size, with only 2 tumours node-positive. The results of this study are limited to the very high risk population of women who are proven BRCA1 or BRCA2 mutation carriers and including those with a personal history of breast cancer. It may therefore not be generalisable to all women with an increased risk of breast cancer due to a family history. Further studies with larger numbers and longer follow-up, and including women of other risk groups, are required.

Table 25. Primary research studies appraised investigating the accuracy and efficacy of MRI surveillance compared to mammography on outcomes from breast cancer (continued)

Source	Study design Evidence grading	Comparison interventions and dates of testing	Sample	Outcomes and Verification	Results	Comments
Warner et al. (2004) Ontario and Montreal, Canada *Continued*		No direct comparisons were made in this study. Dates of surveillance were between Nov 1997 and March 2003.			Sensitivities of combinations of modalities: XRM + CBE = 45% MRI + CBE + XRM = 86% Measures of Accuracy of individual modalities: Sensitivity (95% CI): XRM= 36% (17.1 TO 59.3%) MRI = 77% (54.6 to 92.2%) MRI was significantly more sensitive than mammography (p=0.02). Specificity (95% CI): XRM = 99.8% (98.7 to 99.9%) MRI = 95% (92.9 to 97.2%) (was 99% in 3rd year) PPV (95% CI): XRM = 89% (51.7 to 99.7%0 MRI = 46% (29.5 to 63.15) NPV (95% CI): XRM = 97% (94.8 to 98.3%) MRI = 99% (97.2 to 99.6%) AUC: XRM = 0.77 MRI = 0.89 XRM + CBE = 0.77 MRI + CBE + XRM = 0.94	

Table 25. Primary research studies appraised investigating the accuracy and efficacy of MRI surveillance compared to mammography on outcomes from breast cancer (continued)

Source	Study design Evidence grading	Comparison interventions and dates of testing	Sample	Outcomes and verification	Results	Comments
Kuhl et al. (2005b) Germany	Prospective Cohort study III-2 (C1 P2 Q2)	Surveillance protocol: Biannual CBE and US and annual XRM and MRI. If abnormalities found on CBE or US at round without XRM or MRI, these additional modalities were used to further investigate this. Surveillance commenced at 30 years or 5 years before the youngest family member affected with the disease. (NB: in first 2 years, women under 30, or 30-39 years with dense breasts did not receive XRM, but this was subsequently abandoned and all women received XRM) Mammography (XRM): Annual conventional film XRM performed with at least 2 views per breast (medio-lateral oblique and caudal-cranial), obtained and interpreted in accordance with German radiological practice guidelines. Diagnoses coded according to the BIRADS diagnostic categories on a 5-point scale (1, negative; 2, benign; 3, probably benign; 4, suspicious abnormality; 5, highly suggestive of malignancy).	Sample no = 529 (out of 590 eligible women, 49 were lost to follow-up after 1 surveillance round and 12 were also excluded as they had a clinical abnormality at initial examination). Inclusion criteria: • asymptomatic women; • personal history of breast cancer included provided that the patient had not undergone bilateral mastectomy, had not received chemotherapy within the previous 12 months and had no metastases. (139 women were included with a personal history of breast cancer). Exclusion criteria: • clinical signs of breast cancer. • chemotherapy within the previous 12 months.	Relevant outcomes: cancer detection; mode of detection; tumour size; tumour stage; node status; interval tumours; sensitivity; specificity; PPV; NPV. Verification of a positive result was achieved by histology (for positive imaging studies). Verification of a negative result was achieved by follow-up for negative imaging studies. If a breast cancer was identified clinically (by palpation) between surveillance rounds or at the 6-month clinical visit, the imaging studies of the previous round were considered false negative. Mean follow-up was 5.3 years (range 2-7 years). A total of 1,542 annual surveillance rounds were completed.	Cancer detection: A total of 43 breast cancers were identified in 41 patients (11 of these women had a prior history of breast cancer), and 40 were said to be detectable by imaging. However, the figures may be misleading as they do not correlate with the interval cancer rate and in some cases may refer to imaging after an interval cancer arose. Mode of detection: XRM identified 14 tumours (only 1 was diagnosed by XRM that wasn't diagnosed by MRI). MRI identified 39 tumours, and XRM plus MRI identified 40 tumours Tumour size, stage and node status (characteristics are presented for XRM and US combined and cannot be separated): Of the 21 cancers detected by XRM and US, 16 were invasive and the rest were DCIS. The invasive cancers had a mean size of 13.9mm and 5 were node-positive. Of the 39 cancers detected by MRI, 31 were invasive and 8 were *in situ* . The invasive tumours had a mean size of 12.4mm and five were node-positive. 14 invasive cancers were detected by MRI that were not detected by XRM or US, with a mean size of 9mm and none of them were node-positive.	Limitations include: CBE and the imaging studies were performed within a time frame of 8 weeks. Few sample characteristics presented, such as OCP or HRT use, or the use of preventative strategies such as Tamoxifen or BSO. Verification bias is likely. Unclear documentation of interval tumours. Lack of blinding to the results of the CBE. Author's conclusions: The authors' conclusions relate to the surveillance strategy as a whole, including CBE and US. If US is used in combination with XRM, it can help compensate for some but by far not for all of the shortcomings of XRM, and it causes a substantial number of false positive diagnoses. If MRI is used for surveillance, XRM proved to be of limited and ultrasound of no additional value. US may however be useful to bridge the relatively long time interval between annual surveillance rounds. Propose that in view of the insufficient diagnostic accuracy of XRM and US, that breast MRI should be considered an integral part of surveillance programmes for women at high familial risk in particular in documented carriers of pathogenic BRCA mutations Reviewer's conclusions: Similar to those of the authors above. MRI is the most effective surveillance modality, especially in women in the highest risk group. It is shown that the sensitivity of XRM decreases as the risk group increases and is especially low in mutation carriers. This is not seen with MRI, which maintains good sensitivity in all risk groups.

Table 25. Primary research studies appraised investigating the accuracy and efficacy of MRI surveillance compared to mammography on outcomes from breast cancer (continued)

Source	Study design Evidence grading	Comparison interventions and dates of testing	Sample	Outcomes and verification	Results	Comments
Kuhl et al. (2005b) Germany *Continued*		Breast MRI: Standard dynamic axial contrast-enhanced breast MRI of both entire breasts was performed on a 1.5T system (NT/INTERA; Philips, Best, the Netherlands) after injection of 0.1mmol/kg body weight gadopentetate dimeglumine (Magnevist, Schering, Berlin, Germany) Ultrasound (US): Performed with 7.5-to 13-MHz probes (Siemens Elegra, GE logic 500 and ATL HDI 5000;Siemens, Erlangen, Germany). The entire breast was systematically examined by the physician who interpreted the study. Diagnoses were scored on a 5-point scale identical to the XRM BIRADS categories Each imaging study was read and scored independently by a different radiologist who had substantial experience with the respective imaging technique.	• women having undergone bilateral mastectomy. • Recruited from high risk clinics in a single gynaecology department Risk Stratification: According to definition of the Consortium on Familial Breast and Ovarian Cancer of the German Cancer Aid, corresponding to a lifetime risk of breast cancer of at least 20% (two or more cases of breast cancer on the same side of the family, including at least two cases with onset before age 50 years, or with breast or ovarian cancer, irrespective of age, families with at least one case of breast cancer diagnosed before 35 years, families with three or more cases of breast cancer on the same side of the family, and women who met the criteria for high familial risk, irrespective of the result of mutational analysis)	Verification of last surveillance round was by continued surveillance in 428 women, telephone interview in 52 women and for 6 women who had prophylactic mastectomy it was by pathology of the specimen. XRM: BIRADS of 4 or 5, biopsy was recommended irrespective of finding in US or MRI. BIRADS 3 was managed by 6-months follow-up until receiving a BIRADS 2 or biopsy clarification. US categorised as BIRADS 3 managed by short-term (6 months) US follow-up. BIRADS 4 or 5 managed by US-guided biopsy (14G, semi-automatic or automatic biopsy gun) except for the following constellation: if an US finding that was suspicious was clearly benign on XRM or MRI no biopsy was performed. MRI: Suspicious scores (4 or 5) were managed by magnetic resonance-guided biopsy.	Interval tumours: The paper states that 40 out of 43 tumours in this cohort were detected by imaging. However, a sentence in the discussion states that the rate of interval cancers was 2% in this cohort. This translates to 10 tumours if it is 2% of the population or 1 tumour if it is 2% of the number of tumours detected overall. The latter is more likely but it is unclear. There is no indication of which risk group these interval tumours were detected in. Comparisons: When stratified by risk groups, the detection rates at both the prevalent and incident rounds were much higher in the mutation carriers than the other 2 risk groups, but these differences are not statistically significant. Sensitivity: XRM 32.6% (19.0 to 48.5%) n = 14/43 MRI 90.7% (77.9 to 97.4%) n = 39/43 MRI+XRM 93.0% (80.9 to 98.5%) n = 40/43 When stratified by risk groups, XRM becomes less sensitive as the lifetime risk of breast cancer increases, with a sensitivity of 25%, for the mutation carrier group. This effect is not seen with MRI which maintains good sensitivity across all risk groups.	The limitations of this study must be taken into account in the interpretation,

Table 25. Primary research studies appraised investigating the accuracy and efficacy of MRI surveillance compared to mammography on outcomes from breast cancer (continued)

Source	Study design Evidence grading	Comparison interventions and dates of testing	Sample	Outcomes and verification	Results	Comments
Kuhl et al. (2005b) Germany *Continued*		The readers were informed about the clinical findings from CBE and the risk status of the patient but were blinded to the results of the respective other imaging modalities. Comparisons are made between the 3 risk groups and the different modalities of surveillance. Dates of study were February 1996 to February 2002.	In women without a personal history of breast cancer the Claus tables were also used to quantify risk. Women were then stratified into 3 risk groups for analysis: Mutation carriers. High lifetime risk (20-40%). Moderate lifetime risk (20%).	Findings categorised as BIRADS 3 short-term follow-up after 6 months was recommended with further management corresponding to that of XRM BIRADS 3 lesions BIRADS 3 categories in all imaging that received short-term follow-up were not considered positive for the calculation of outcomes. Invasive cancer and DCIS were considered a malignant diagnosis but LCIS and atypical ductal hyperplasia were considered to be benign.	A sensitivity of 100% is documented for each risk group (but the denominator for calculating this is smaller than the overall number of women, 34 instead of 43. How this figure is arrived at is unclear. Specificity (95% CI): XRM 96.8% (95.7 to 97.7%) *n* = 1364/1409 MRI 97.2% (96.2 to 98.0%) *n* = 1370/1409 MRI+XRM 96.1% (94.9 to 97.0%) 1354/1409 Stratification by risk group does not appear to affect the specificity. PPV (95% CI): XRM 23.7% (14 to 37%) *n* = 14/59 MRI 50% (38.4 to 61.5%) *n* = 39/78 MRI+XRM 42.1% (32.0 to 52.75) *n* = 40/95 The PPV increases with the increasing risk of breast cancer, this will be affected by the higher incidence in women at higher risk. NPV (95% CI): XRM 97.9% (97.0 to 98.6%) *n* = 1364/1393 MRI 99.7% (99.2 to 99.9%) *n* = 1370/1374 MRI + XRM 99.8% (99.3 to 99.9%) *n* = 1354/1357	

Table 25. Primary research studies appraised investigating the accuracy and efficacy of MRI surveillance compared to mammography on outcomes from breast cancer (continued)

Source	Study design Evidence grading	Comparison interventions and dates of testing	Sample	Outcomes and verification	Results	Comments
Lehman et al. 2005	Prospective international multicentre cohort study, 13 facilities located in the USA and Canada Grade III-2 (C1 P2 Q2)	MRI: protocol parameters included pre-contrast sagittal T2 (4000/80; 256*256) fast spin-echo images with fat suppression and both pre- and post-contrast sagittal T1 (TR≤50/TE≤4.5; 256*128*32-60) three-dimensional gradient-echo images with a 60-degree flip angle. The field of view was restricted to 16-18cm depending on patient size and slices measured <3mm in thickness. T1 images were acquired prior to and immediately after bolus injection of contrast. Study protocol specified that the XRM and the clinical breast examination (CBE) were to be performed within 90 days of the MRI.	390 asymptomatic women at high risk of breast cancer aged 45±9.7 years Inclusion criteria: ≥ 25 years and lifetime risk of breast cancer >25% based on family history or genetic test confirmation; prior history of breast cancer if having contra-lateral breast imaged, and women diagnosed with breast cancer >5 years prior to the study; eligible for bilateral imaging if probability of breast cancer >50% based on the study risk algorithm or positive test for a mutation in BRCA1 or BRCA2. Exclusion criteria: contraindications to MRI examination (claustrophobia, pregnancy, pacemaker, magnetic aneurysm clip or implanted magnetic device), and women who presented with palpable lesions or mammographic abnormalities prior to risk assessment.	Any suspicious MRI enhancing lesions were described based on lesion shape, borders, distribution and internal architecture. The overall assessment was classified on a 5-point scale (1, negative; 2, benign; 3, probably benign; 4, suspicious abnormality; 5, highly suggestive of malignancy). A lesion was identified as malignant if there was a focal mass with irregular or speculated margins, if enhancement was in a ductal distribution if a solid lesion showed rim enhancement or if there was intense regional enhancement in less than one quadrant. Benign lesions were identified as those that had smooth or lobulated margins with internal septations or if the mass was cystic. All lesions given an assessment score of 4 or 5 were recommended for biopsy. A retrospective review was also performed that included all images (MRI and XRM) from patients with cancers diagnosed during the study.	MRI: 27 biopsies were performed as a result of a positive examination. Four of 27 lesions biopsied were diagnosed as malignant and 23 lesions were diagnosed as either benign, atypical ductal hyperplasia (by excisional biopsy) of lobular carcinoma *in situ*. All four women with malignant lesions were diagnosed by MRI. 7 true positives on MRI. 20 false positives on MRI. MRI PPV: 12.9% (95%CI 3.6%, 30%). XRM: 7 women had assessments that were positive only on MRI and 1 woman had an assessment positive on both XRM and MRI. 11 women declined biopsies after positive findings, including 6 women who had BIRADS 4 assessments on MRI (with negative, benign, or probably benign mammograms) and 3 women who had B RADS 4 assessments on XRM (with negative, benign, or probably benign MRI assessments). Two women declined based on a 'probably benign' MRI assessment.	Comparator imaging could have been performed up to 90 days apart. Possibly too long after the MRI, given the aggressive nature of the tumours found in BRCA1 and BRACA2-mutation carriers, to be a true comparison of surveillance procedures. Data not presented for clinical breast examinations. For one woman who had a positive MRI, the lesion did not persist on a subsequent MRI examination. It remains unclear how many of those women who had positive assessments on MRI received another MRI assessment before proceeding to biopsy. Authors' conclusions: MRI surveillance in women at high risk of breast cancer is capable of detecting mammographically and clinically occult breast cancer. However MRI also resulted in 19 false-positive outcomes, leading to benign biopsy results. Reviewer's conclusions: This was a pilot study with no long-term follow-up to identify potential false-negative MRI results or delayed diagnoses when biopsies were declined. In addition, only a single round of surveillance was performed in the study, so there is no data on the interval between surveillance tests. The number of false-positive results is important to consider, particularly given the additional anxiety faced by women who may have seen close family members suffer from breast cancer.

Table 25. Primary research studies appraised investigating the accuracy and efficacy of MRI surveillance compared to mammography on outcomes from breast cancer (continued)

Source	Study design Evidence grading	Comparison interventions and dates of testing	Sample	Outcomes and verification	Results	Comments
MARIBS study group (Leach et al. 2005)	Prospective multicentre cohort study Grade III-2 (C1 P1 Q2)	MRI: Equipment supplied by four manufacturers (GE medical systems, Slough UK; Marconi Medical systems and Philips Medical Systems, Reigate UK; Siemens medical Solutions, Bracknell, UK). All systems had a field strength of 1.0-1.5 Tesla with a dedicated breast coil and with the systems capable of running the agreed national protocol of sequences. The MRI surveillance protocol comprised high spatial resolution T1-weighted sequences before and after contrast medium injection sandwiching a T1-weighted three dimensional coronal dynamic acquisition series with 2 sequences before the bolus IV injection of 0.2mmol per kg bodyweight of gadopentetate dimelglumine (Gd-DTPA; Schering Healthcare, Burgess Hill UK) and at four to six time points after injection. Mammography (XRM) done annually and by preference on the same day as the MRI. Exams took place either in an accredited screening centre or on a family history clinic working to NHSBSP standards.	649 asymptomatic women at high risk of breast cancer aged 35-55 years (median:40 years) Recruited from 22 centres with familial breast cancer clinics in the UK Risk status established by one of the following criteria: known carriers of a deleterious BRCA1, BRCA2 or TP53 mutation, 1st degree relative of someone with BRCA1, BRCA2 or TP53 mutation, strong family history of breast or ovarian cancer, family history consistent with classic Li-Fraumeni syndrome. Exclusion Criteria: previous breast cancer or any other cancer such that expected prognosis was <5 years. Participants who underwent genetic testing with a negative result, and women who developed cancer were excluded from further participation.	For the reporting forms for MRI, a scoring system was used based on morphological and dynamic contrast uptake characteristics previously validated against histology. A worksheet was also developed to ensure consistency of method in the choice of regions of interest and their analysis. XRM was also double reported. Patients recalled because of an indeterminate or suspicious test had either a high temporal resolution study with 0.1mmol per kg Gd-DTPA or a repeat of the initial surveillance protocol, done at a different phase of the menstrual cycle to the initial test. The reporting radiologist and the attending doctor decided the diagnostic pathway.	Among women with more than one surveillance round, there were 1,232 surveillance intervals of 6-54 months in length (median 12 months). 85% of surveillance intervals were between 10-14 months. 35 cancers were diagnosed in 649 women. 6 cancers were detected by XRM only. 19 cancers were detected by MRI alone. Of these 19, four were detected by only one reader (double reading was used throughout). 33 cancers were detected by both XRM and MRI. 2 cancers were interval cancers. Of cancers detected by MRI alone: Three were Grade 1 tumours, five were Grade 2 tumours and 11 were Grade 3 tumours. Seven patients had IDC and ductal carcinoma *in situ*, nine participants had invasive ductal cancer and 2 participants had invasive lobular cancer. All women: Sensitivity of XRM: 40% (95%CI 24, 58) Sensitivity of MRI: 77% (95%CI 60, 90) Sensitivity of XRM + MRI: 94% (95%CI 81, 99) *p*-value (MRI vs. XRM) = 0.01 Specificity of XRM: 93% (95%CI 92, 95) Specificity of MRI: 81% (95%CI 80, 83) Specificity of XRM and MRI: 77	MRI scans done on the same day as other surveillance tests. During the course of the study 30 women who eligible on basis of family history of breast or ovarian cancer became ineligible for further participation because of a negative genetic test. All breast cancers in the BRCA1 and BRACA2 groups were in known mutation carriers. 57 of the 126 women without cancer in the BRCA1 group had not been tested but have a family member with breast/ovarian cancer history. Although the sensitivities quoted for these groups refer exclusively to tested mutation carriers, the specificities do not and should be interpreted as preliminary estimates. Authors' conclusions: Findings indicated that MRI using contrast enhancement is more sensitive than mammography for cancer detection, and that specificities for both procedures was acceptable. Despite a high proportion of Grade 3 cancers found, tumours were small and few women were node-positive in this group. Reviewers' conclusions: Cancers in women who carry a BRCA1 mutation are typically high grade, which may account make the finding of a higher proportion of Grade 3 tumours being detected. The outcomes of this study support the authors' conclusions regarding the sensitivity of MRI as a surveillance test.

Table 25. Primary research studies appraised investigating the accuracy and efficacy of MRI surveillance compared to mammography on outcomes from breast cancer (continued)

Source	Study design Evidence grading	Comparison interventions and dates of testing	Sample	Outcomes and Verification	Results	Comments
MARIBS study group (Leach et al. 2005) *Continued*		Mammographic examinations were either 2-view or 1-view (by medial oblique only)		Recall rates: 10.7% per woman year for MRI. Of 137 supplementary MRI studies, seven were done in women later shown to have cancer and 13 MRI-guided biopsies were taken 62% of suspicious findings on MRI were resolved without invasive procedures and16 women had diagnostic surgery to complete their diagnosis.	*p*-value (MRI vs. XRM) = 0.0001 PPV: XRM: 10.8%: (95%CI 5.8, 17) MRI: 7.3% (95%CI 4.9, 10) Area under ROC: XRM: 0.70 (95%CI 0.68, 0.72) MRI 0.85 (95%CI 0.84, 0.87) P=0.035 BRACA1 mutation/in first-degree relative: Sensitivity: XRM 23% (95%CI 5, 54), MRI 92% (95%CI 64, 100) XRM + MRI 92% (95%CI 64, 100) p-value (XRM vs. MRI)=0.004 Specificity: XRM 92% (95%CI 88, 94) MRI 79% (95%CI 75, 83) XRM + MRI 74% (95%CI69, 78) *p*-value (XRM vs. MRI)=0.0001 BRACA2 mutation/in first-degree relative: Sensitivity: XRM 50% (95%CI 21, 79) MRI 58% (95%CI 28, 84) XRM + MRI 92% (95%CI 62, 100) *p*-value (XRM vs. MRI) 1.0 Specificity: XRM 94% (95%CI 91, 97) MRI 82% (95%CI77-87) XRM + MRI 78% (95%CI 72, 83) *p*-value (XRM vs. MRI) 0.0001	

Chapter 8: Accuracy and efficacy of mammography, ultrasound and MRI

SECONDARY RESEARCH

The search strategy identified only one relevant systematic review of the effectiveness of combination surveillance in women at high risk of breast cancer. This was carried out in Israel and focused specifically on women who were BRCA1 or BRCA2 mutation carriers, although some of the studies included also had high-risk non-mutation carriers as participants (Calderon-Margalit and Paltiel 2004). The methods and conclusions are described in **Table 26**.

The inclusion and exclusion criteria set by Calderon-Margalit et al. (2004) were not explicit, and are not necessarily concordant with the criteria applied in this review. Therefore, the results must be interpreted with caution. The databases searched were Medline and Pubmed, and references of retrieved articles were also obtained. The articles included were published between 1998 and December 2004, and non-English papers were excluded. Five studies were included in total. There were two retrospective cohort studies, one non-randomised trial, one cross-sectional study and one cohort study.

The sample sizes ranged from 12 to 128 mutation carriers and up to 196 total participants. Considerable heterogeneity was demonstrated between these studies in terms of the level of breast cancer risk in the women included and the surveillance protocols. All participants appeared to be at over 15 per cent lifetime risk of developing breast cancer. Three studies used XRM and MRI for surveillance and the other two used XRM, MRI and US for surveillance. It was not clear how many of the studies involved CBE surveillance or recommended regular SBE. Comparisons were made between the modalities of surveillance and various combinations of modalities.

The outcomes of interest were the cancer detection rates, interval tumours and measures of accuracy. The sensitivities were documented as ranging from 42-100 per cent, the specificities from 93-96 per cent, the PPV from 26-66 per cent and the NPV from 95-100 per cent. The authors conclude that none of the included studies investigated the efficacy and effectiveness of surveillance strategies in terms of outcomes such as breast cancer mortality, breast cancer stage and grade or quality of life. Instead they described the internal validity of either a surveillance protocol or specific modalities of surveillance, particularly MRI. They describe how there was considerable heterogeneity in the studies, especially in terms of their choice of gold standard for calculating measures of accuracy. They concluded that the there is a need for more high quality evidence on the efficacy of MRI in the surveillance of breast cancer among BCA mutation carriers and stated that if the sensitivity of MRI proves to be about 100 per cent in detecting breast cancer, then clear criteria need to be defined as to who should receive surveillance in order to increase its PPV and reduce unnecessary interventions and control costs.

Two of the studies which were in the review by Calderon-Margalit et al. (2004), also qualified for inclusion in this review. These two studies are still discussed as they contain information not just on mutation carriers and on combination surveillance strategies that were not fully covered by Calderon-Margalit et al. (2004).

Table 26. Secondary research appraised relevant to accuracy and efficacy of combination surveillance on outcomes from breast cancer

Source	Search method	Criteria for inclusion/exclusion	Results	Comments
Calderon-Margalit, R. and Paltiel, O. (2004)	Search: 1998-2004 Databases searched: Medline, Pubmed and references from relevant articles. Key words: "BRCA1" ,"BRXCA2" with "prevention", "breast cancer", "prophylactic mastectomy", "tamoxifen, "chemoprevention", "screening", "mammography" and "MRI".	A PICOT question is not specified. The purpose was to review the evidence on surveillance for early detection (also bilateral prophylactic mastectomy, prophylactic oophorectomy and chemoprevention) in preventing breast cancer and improving survival of BRCA1 and BRCA2 mutation carriers. Inclusion criteria: None given Exclusion criteria: Non English articles	5 studies were identified of relevance to surveillance. These papers included 2 retrospective cohort studies, 1 non randomised trial, 1 cross sectional study and 1 cohort study. These studies all looked at screening with either MRI, mammography or a combination of both, and 1 included CBE and SBE. Follow-up ranged from none to a median of 3 years. Sample sizes ranged from 109-196 women at high risk. 1 was focused on BRCA mutation carriers specifically (Brekelmans et al.); however, the other studies only analysed mutation carriers as a subgroup of women at high risk. The reference standards are not mentioned. Sensitivities ranged from 42-100% Specificities ranged from 93-96% PPVs ranged from 26-66% NPVs ranged from 95-100%	There was a lack of information provided on the method of appraisal of these studies and the overall quality assessment. It was commented that none of the studies that have investigated screening among BRCA mutation carriers have addressed the efficacy or effectiveness of screening methods in terms of outcome, such as breast cancer mortality, breast cancer stage and grade, or quality of life of breast cancer patients. Instead these studies discussed the internal validity of either a screening protocol or specific screening modalities, mainly MRI. These studies substantially differed in their study population and their choice of a gold standard (necessary for the calculation of sensitivity and specificity). Most studies did not distinguish whether tests were performed as screening tests or were part of a diagnostic work-up in suspected breast cancer patients. There were also few details on the inclusion and exclusion criteria set for selecting papers. Authors' conclusions: The evidence for surveillance for the early detection of breast cancer amongst BRCA1 and BRCA2 carriers is not yet established. Screening with CBE and mammography show lower sensitivity in BRCA1 and BRCA2 carriers than in the general population. Screening with MRI might offer higher sensitivity rates than mammography. There is need for more high quality evidence on the efficacy of MRI in the surveillance of breast cancer among BRCA mutating carriers. If indeed the sensitivity of MRI proves to be about 100% in detecting occult breast cancer, clear criteria should be defined as to who should receive MTI screening in order to increase its PPV, reduce unnecessary procedures and control costs.

Table 26. Secondary research appraised relevant to accuracy and efficacy of combination surveillance on outcomes from breast cancer (continued)

Source	Search method	Criteria for inclusion/exclusion	Results	Comments
Calderon-Margalit, R. and Paltiel, O. (2004) *Continued*				There is little information on the effectiveness of measures, other than prophylactic mastectomy, in preventing breast cancer and improving survival. Therefore, the current body of knowledge does not allow a woman or her physician to confidently make long-term decisions. Reviewers' conclusions: It appears that there is considerable heterogeneity in the small amount of research available, in methods, study populations, screening protocols and reference standards. This substantiates the authors' conclusion that there is not sufficient evidence for any one particular screening modality or regime and that further research is necessary in this high risk population.

PRIMARY RESEARCH

The search identified four eligible primary research studies. Below is an overview of study designs and aspects of quality represented by these studies. Full details of the papers appraised, including methods, key results, limitations and conclusions, are provided in evidence **Table 29**. Studies are presented in chronological order of publication within the tables.

Study design and quality assessments

As discussed, the most robust method of comparing the usefulness for surveillance of diagnostic tests would be an RCT. Consistent with this entire review, no such evidence was identified. All of the four eligible studies were graded evidence level III-2, and all were prospective cohort studies. The studies were all designed to compare the accuracy and effectiveness of combination surveillance, with XRM, US and MRI with or without CBE, US in populations of women at high risk of breast cancer. The results of these studies have been discussed individually in the preceding chapters. This chapter aims to bring these results together and also present the accuracy and efficacy of combination strategies.

These studies were of moderate quality in design and conduct. Several limitations apply across all the studies. They were all likely to be affected by verification bias, because the reference standard for diagnosis was different in the case of a positive surveillance result versus a negative result. Positive surveillance results were followed by biopsy or surgical excision and histopathological confirmation. However, verification of negative results was only possible through clinical follow-up over the surveillance interval. The duration of follow-up varied between the studies. It is possible that interval tumours may not have been detected in case of inadequate follow-up after surveillance and therefore the effectiveness of the test would be overestimated. There were a variety of systems used to classify the surveillance images and differing cut-off points determining an abnormal examination. The level at which this is set would influence the outcomes of the study and also the ability to draw conclusions across studies. The system used and cut-off will be reported for each study if it was documented. The radiologist was not always blinded to the women's risk status in the studies, or to the results of the other modalities of screening. This knowledge may affect their degree of suspicion and therefore the thoroughness with which they carried out the examination.

Study setting

Two studies were undertaken in single centres (Kuhl et al. 2005b; Trecate et al. 2003) and two were multi-centred (Podo et al. 2002; Warner et al. 2004). Podo et al. (2002) recruited participants from genetics centres in Italy, Trecate et al. (2003) recruited participants from the National Cancer Institute in Milan in Italy, Warner et al. (2004) recruited participants from familial cancer centres in Canada and Kuhl et al.(2005b) recruited from high risk breast clinics in one hospital in Germany. As discussed in the chapter on surveillance by XRM, the setting of the study usually determines the prevalence and spectrum of disease in the participant population (Deeks 2001). However, once again this was also determined by the risk stratification that participants underwent.

Risk stratification

The methods of risk stratification varied between the studies. As with the preceding chapters, the risk stratification strategies and the rest of the information for these studies is presented individually.

Podo et al. (2002)

Study sample

This prospective cohort study recruited male and female patients from nine genetics centres within Italy. At the time of publication 105 women had been recruited and no men. Women were included if they were aged 25 years or over and men if they were 50 years or older. The mean age at recruitment was 46 years, with a range of 25-77 years. Risk stratification was performed by criteria specific to this study. Participants had to be known BRCA1 or BRCA2 mutation carriers, or have a one in two probability of being a carrier i.e. have a first-degree relative who was a proven mutation carrier. Two

women were also included whose families had a very high incidence of breast cancer that was likely associated to a mutation other than BRCA1 or BRCA2. Women with a personal history of breast cancer were included if it was unilateral, i.e. 40 in total. They received unilateral surveillance if they had undergone mastectomy and bilateral if they had received breast conserving surgery. If women were on HRT, they were included after stopping treatment for three months. Exclusion criteria were: pregnancy, breast feeding, current chemotherapy, terminal illness and specific contraindications to MRI.

Interventions and comparators

Surveillance consisted of CBE, XRM, US and MRI at yearly intervals. The BIRADS system was used to classify the XRM, but the cut-off for an abnormal screen was not documented. It was not reported whether the radiologists were blinded to the results of other modalities of screening. For MRI, a system of classification was used that was based on a combination of morphological and enhancement patterns. Scores 0-2 were benign, 3 was uncertain and 4-8 was malignant. In the case of non-benign scores (3-8) which were detected only by MRI, the MRI was repeated after one to two months. If the lesion was confirmed then a biopsy was undertaken. The MRI was performed using coronal and axial planes. Contrast enhancement was also used, with Gd-chelate (0.1 mmol/kg) injected. One pre-contrast and five post-contrast images were taken. Pre-menopausal women had MRI within the 2nd week of the menstrual cycle. US was performed with a probe set at a frequency of $\geq$ 7.5 MHz. The study reported on the preliminary phase of this research and therefore the follow-up was incomplete. Only 21 months of the study had been completed at the time of publication.

Outcomes

Cancer detection rate

Eight tumours were detected in total, seven in the prevalent screen and one in the incident screen. This combination strategy gives a detection rate of 76 per 1,000 women under surveillance. Seven tumours were detected by MRI (67 per 1,000 women under surveillance), XRM only detected one tumour (9 per 1,000 women under surveillance) and similarly US only detected one tumour (9 per 1,000 women under surveillance). This is a significant difference between MRI and XRM or US (p=0.03 for MRI versus XRM or US). Five of the tumours were detected in patients with a previous personal history of breast cancer.

Accuracy measures

Accuracy measures were not calculated due to the incomplete follow-up.

Tumour characteristics

There were five invasive tumours, two DCIS and one combined DCIS and LCIS. The tumour size ranged from 3-27mm and none had lymph node involvement. The tumour characteristics were not stratified by mode of detection.

Interval tumours

Interval tumours were not reported due to the incomplete follow-up.

In summary, this study shows a similar performance for XRM and US in the surveillance of this very high-risk group and a significantly higher detection rate for MRI than XRM or US. If MRI had not also been used in this study, the majority of tumours would have remained undetected. This suggests that in such a high-risk group, including mostly mutation carriers and a high proportion having a personal history of breast cancer, surveillance by XRM and US is not adequate. The study is limited by the small sample size, who were very high risk, and by the small number of tumours detected. There need to be further results from this study to comment on the measures of accuracy, interval tumours and thereby the surveillance interval. No further reports from this group were identified in the literature search.

Trecate et al. (2003)

Study sample

This prospective cohort study recruited 23 women at high risk of breast cancer from the National Cancer Institute in Milan, Italy. There was no age restriction and no average age of the cohort was given. The age range was 30-61 years. Risk stratification was specific to this study. The women included were either BRCA1 or BRCA2 mutation carriers, had a one in two probability of being a carrier or over a 50 per cent risk of carrying a susceptibility gene for familial breast cancer based on family history. Women with a personal history of breast cancer were included (six women).

Interventions and comparators

Surveillance depended on the age group of the women. All ages had CBE every six months. Mammography was annual and commenced at 25 years with bilateral one view, and then increased to bilateral double view from 30 years and above. Annual US was performed alone from 20-25 years, then with XRM from 25-35 years, then six months after XRM from 35-40 years and above 40 years only if requested by the radiologist. The US was performed with either 7.5MHz or 10-12 MHZ probes (ATL HDI 3500, Philips). Annual MRI was performed for all ages for two years during the study. A Siemens Vision 1.5 was used with a dedicated double coil. The method of classifying the images was not documented. Follow-up was not documented. It was not reported whether the radiologists interpreting the images was blinded to the results of the other imaging modalities. The study was conducted over a seven-month period but the dates were not given. It is unclear if this work may have been related to the study by Podo et al. (2002). Comparisons were made between the different modalities of surveillance.

Outcomes

Cancer detection rate

Four breast cancers were detected overall. This gives a detection rate of 170 per 1,000 women under surveillance. All tumours were detected by MRI, three were detectable by CBE (130 per 1,000 women under surveillance) but none of the tumours were detected by XRM or US.

Measures of accuracy

No measures of accuracy were calculated in this study.

Tumour characteristics

All four tumours were invasive. Only two tumours had the size recorded and these were 10mm and 30mm. No record of the lymph node status was documented. There was no stratification of tumour characteristics by modality of surveillance.

Interval tumours

No interval tumours were documented.

In summary, this study suggests that MRI may be a useful modality of surveillance for women at very high risk of breast cancer, i.e. mostly mutation carriers with a high proportion having a personal history of breast cancer. The results are extremely limited by the very small sample size, small number of tumours and the lack of detail documented in the publication. The study focuses on very high risk women and may not be generalisable to all women at high risk of breast cancer.

Warner et al. (2004)

Study sample

This prospective cohort study recruited 236 female BRCA1 and BRCA2 mutation carriers from familial cancer centres in southern Ontario and Montreal in Canada. There were no age restrictions and the mean age at first surveillance was 46.6 years, with a range of 25-65 years. Risk stratification was performed by all participants being BRCA1 or BRCA2 mutation carriers. This was therefore a very high risk group, 31 per cent of whom were of Ashkenazi Jewish descent. In addition, 30 per cent had a personal history of breast cancer. Exclusion criteria were: a past history of unilateral breast cancer if the contralateral breast was not intact, pregnant or lactating women, history of bilateral breast cancer

currently undergoing chemotherapy or known to have metastatic disease and women weighing over 91kg (technical reasons). Thirty-one women left the study before completing three rounds, 16 underwent bilateral mastectomy, three were too large for MRI machine, three stopped due to pregnancy, four developed metastatic cancers, four were lost to follow-up and one did not wish to continue participating.

Interventions and comparators

Surveillance consisted of biannual CBE and annual XRM, US and MRI, all performed on the same day. Surveillance commenced at least one year after the woman's last mammogram. CBE was coded as normal, suggestive of benign disease, indeterminate, or suspicious of malignancy. Indeterminate CBE exams were repeated after three months. MRI was performed with 1.5 T magnet (Signa, General Electrical Medical Systems). US used a 7.5MHz transducer (the first seven patients did not receive US). All participants underwent the first surveillance round, but only 58 per cent had the second and 36 per cent the third. BIRADS was used to classify the images and scores of 4 or 5 were biopsied. Each imaging study was read and scored independently by a radiologist specialsed in breast imaging and the radiologists were also blinded to the results of CBE. All patients were followed up for a minimum of one year after their last surveillance examination. Comparisons were drawn between different modalities of surveillance.

Outcomes

Cancer detection rate

Twenty-two cancers were detected overall in 21 women (one woman had bilateral cancer). Seven of these women had a past history of breast cancer. This gives a cancer detection rate of 93 per 1,000 women under surveillance. Two were detected by CBE (8 per 1,000 women under surveillance), eight by XRM (34 per 1,000 women under surveillance), seven by US (30 per 1,000 women under surveillance) and 17 by MRI (72 per 1,000 women under surveillance). Seven tumours were detected by MRI alone, two by XRM alone and two by US alone.

Sensitivity

All the measures of accuracy in the paper are presented individually for each year of surveillance. These results have been combined to give overall results for the three rounds of surveillance. There was not enough raw data to calculate measures of accuracy for CBE.

The sensitivity of XRM, US and MRI respectively were 36 per cent (95% CI, 17.1 to 59.3%), 33 per cent (95% CI, 14.6 to 66.9%) and 77% (95% CI, 54.6 to 92.2%). XRM was significantly more sensitive than either XRM (p=0.02) or US (p=0.006).

Specificity

The specificity of XRM, US and MRI respectively were 99 per cent (95% CI, 98.7 to 99.9%), 96 per cent (95% CI, 93.7 to 97.7%) and 95 per cent (95% CI, 92.9% to 97.2%).

PPV

The PPV of XRM, US and MRI respectively were 88 per cent (95% CI, 51.7 to 99.7%), 29 per cent (95% CI, 12.6 to 51.15) and 46 per cent (95% CI, 29.5 to 63.1%).

NPV

The NPV of XRM, US and MRI respectively were 97 per cent (95% CI, 94.8 to 98.35), 97 per cent (95% CI, 94.5 to 98.2%) and 99 per cent (95% CI, 97.2 to 99.6%).

AUC

The AUCs for XRM, US and MRI respectively were 0.77, 0.65 and 0.89. The AUC for CBE is also given at 0.48 and several combination strategies; CBE and XRM was 0.77, CBE and XRM and US was 0.81, CBE and MRI and US was 0.91, CBE and MRI and XRM was 0.94, CBE and XRM and MRI and US was 0.93. There were no confidence intervals documented for the AUCs.

Tumour characteristics

Sixteen tumours were invasive and six were DCIS. The mean size of the invasive tumours was 11mm at the first surveillance round and 13mm at the second round. Fifteen cases had lymph node sampling

and two were node-positive. The tumour characteristics are not documented stratified by modality of surveillance.

Interval tumours

There was one interval tumour, detected in a 40 year old BRCA1 mutation carrier seven months after her third surveillance round (retrospectively this tumour was visible on MRI and XRM at the previous surveillance visit). Another woman, who elected to have a bilateral mastectomy after breast cancer was found, had a 2mm focus of DCIS in the contralateral breast which had not shown up at surveillance two months earlier.

Mortality

All 22 patients with tumours were still alive and disease-free at the time of publication of the article.

In summary, this study suggests a similar efficacy and accuracy of XRM and US in the surveillance of high risk women and a superior efficacy and accuracy of MRI to XRM and US surveillance in women at high risk of breast cancer. The highest AUC resulted from the combination strategy of CBE with XRM and MRI (0.94). The results of this study are limited to the very high risk population of women who are proven BRCA1 or BRCA2 mutation carriers, including those with a personal history of breast cancer. It is therefore not generalisable to all women with an increased risk of breast cancer due to a family history. Further studies with larger numbers and longer follow-up, and including women of other risk groups are required.

Kuhl et al. (2005b)

Study sample

This prospective cohort study recruited 529 women from high risk clinics in a single hospital in Germany. There was no age restriction and the mean age of the whole cohort was 41.7 years, with a range of 27-59 years. Risk stratification was performed according to the Consortium on Familial Breast and Ovarian Cancer of the German Cancer Aid. All participants had over a 20 per cent lifetime risk of breast cancer. In women that did not have a personal history of breast cancer, the Claus tables were also used to stratify risk. Women with a personal history of breast cancer were included provided the women had not had bilateral mastectomy, had not had chemotherapy within the last 12 months and had no metastases (139 women had a personal history). Another inclusion criterion was being asymptomatic.

Interventions and comparators

Surveillance consisted of biannual CBE and US and annual XRM and MRI. If abnormalities were found on CBE or US at the round without XRM or MRI, these additional modalities were used to further investigate this. Surveillance commenced at 30 years, or five years before the youngest family member affected with the disease. (NB: in the first two years, women under 30, or 30-39 years with dense breasts did not receive XRM, but this was subsequently abandoned and all women received XRM and these data were not included in the calculation of accuracy measures). MRI of both entire breasts was performed on a 1.5T system (NT/INTERA; Philips, Best, the Netherlands). US was performed with 7.5-MHz-13MHz probes. Each imaging study was read and scored independently by a different radiologist who had substantial experience with the respective imaging technique. The radiologists were informed about the clinical findings from CBE and the risk status of the patient but were blinded to the results of the respective other imaging modalities. BIRADS was used to classify the images and scores of 4 or 5 went for biopsy. The mean follow-up time was 5.3 years, with a range of 2-7 years. The number of total annual surveillance rounds for which data on all three imaging modalities was available was 1,452, and this was used in the calculation of accuracy measures. Comparisons are made between the three risk groups and the different modalities of surveillance.

Outcomes

Cancer detection rate

A total of 43 tumours arose in 41 patients during the study period. It is documented that 40 of these were detected by imaging. That gives a cancer detection rate for the overall surveillance strategy of 76 per 1,000 women under surveillance. Eleven (25%) of these patients had a prior history of breast

cancer. CBE identified only one tumour (2 per 1,000 women under surveillance) which was also detected on imaging. Fourteen tumours were detected by XRM (26 per 1,000 women under surveillance). Only one was diagnosed by XRM that was not diagnosed by MRI. Seventeen tumours were detected by US (32 per 1,000 women under surveillance), two of these were at the half yearly CBE and US screen and weren't palpable. Twenty-one tumours were detected by US and XRM combined (40 per 1,000 women under surveillance). Lastly, MRI identified 39 tumours (74 per 1,000 women under surveillance), and XRM and MRI combined identified 40 tumours (76 per 1,000 women under surveillance).

Sensitivity

The overall sensitivity for XRM was 32.6 per cent (95% CI, 19 to 48.5%), for US was 39.5 per cent (95% CI, 25.0 to 55.6%), for XRM and US combined was 48.8 per cent (95% CI, 33.3 to 64.5%), for MRI was 90.7 per cent (95% CI, 77.9 to 97.4%) and for MRI and XRM combined was 93.0 per cent (95% CI, 80.9 to 98.5%).

Overall, there was no apparent difference in sensitivity between XRM and US. MRI was significantly more sensitive than XRM, US or the combination of both ($p< 0.001$).

When stratified by risk groups XRM, US and the combination of XRM+US all become less sensitive as the lifetime risk of breast cancer increases, with sensitivities of 25 per cent, 25 per cent and 37.5 per cent respectively for the mutation carrier group. This effect is not seen with MRI which maintains consistent sensitivity across all risk groups.

Specificity

The overall specificity for XRM was 96.8 per cent (95% CI, 95.7 to 97.7%), for US was 90.5 per cent (95% CI, 88.8 to 92.0%), for XRM and US combined was 89.0 per cent (95% CI, 87.2 to 90.6%), for MRI was 97.2 per cent (95% CI, 96.2 to 98.0%) and for MRI and XRM combined was 96.1 per cent (95% CI, 94.9 to 97.0%).

Overall, MRI offered approximately the same specificity as XRM ($p>0.05$). Both MRI and XRM were significantly more specific than US alone or in combination with XRM ($p<0.001$).

Stratification by risk group or by a past history of breast cancer does not appear to affect the specificity.

PPV

The overall PPV for XRM was 23.7 per cent (95% CI, 1 to 29%), for US was 11.3 per cent (95% CI, 6.7 to 17.4%), for XRM and US combined was 11.9 per cent (95% CI, 7.5 to 17.6%), for MRI was 50 per cent (95% CI, 38.4 to 61.5%) and for MRI and XRM combined was 42.1 per cent (95% CI, 32.0 to 52.7%).

Overall, the PPV was significantly higher for XRM when compared to US or US and XRM combined ($p=0.02$). However MRI had a significantly higher PPV than XRM ($p=0.001$), US ($p<0.001$) and XRM and US combined ($p<0.001$).

Stratification by risk group or by a past history of breast cancer does not appear to affect the PPV either.

Tumour characteristics

Thirty-four tumours were invasive and nine were DCIS. Of the 21 cancers detected by XRM and US, 16 were invasive and the rest were DCIS. The invasive cancers had a mean size of 13.9mm and five were node-positive. Nineteen cancers were detected by MRI that were not detected by XRM or US, 14 of these were invasive and five were DCIS. The invasive tumours had a mean size of 9mm and none were node-positive. The tumour characteristics were stratified by modality of surveillance. There were no significant differences in the characteristics of the tumours detected by XRM or US (p values all >0.05). However, there was a significantly higher proportion of DCIS and minimal cancers detected by MRI than by XRM and US alone or combined (p values < 0.05). Minimal cancers are defined as DCIS or small invasive cancers less than 10mm in size and with negative lymph nodes.

Interval tumours

The interval tumour rate is given as 2 per cent in this cohort. It is unclear if this is a percentage of the women under surveillance or of the tumours that arose. It was also documented that there was one interval cancer that arose between surveillance rounds. However, it was documented that 40 of the 43 cancers were detectable by imaging, which would suggest three interval cancers. These figures were reported in an unclear manner.

In summary, this study suggests that the addition of US to XRM does not significantly improve the sensitivity of surveillance of women at high risk of breast cancer and does significantly reduce the specificity and PPV. MRI has a significantly higher sensitivity and PPV than other modalities and a similar specificity to XRM alone. The combination strategy with the highest sensitivity was MRI and XRM (93%). MRI has a suggested advantage for women in the highest risk groups (mutation carriers) as it does not appear to loose sensitivity as the risk status increases, as XRM, US and XRM combined with US do. The data on interval tumours is somewhat unclear in its documentation. This study included women at high risk who had a personal history of breast cancer, but the majority of the results were not significantly different if stratified by personal history.

Summary

There were four studies identified of relevance to combination surveillance for women at high risk of breast cancer. These were all prospective cohort studies. There were 893 women under surveillance overall in these four studies. There was heterogeneity between the studies in terms of the surveillance strategies, surveillance intervals, the participants' risk status and age. Three studies recruited participants who were either BRCA1 or BRCA2 carriers, or had a 50 per cent chance of being carriers (Podo et al. 2002; Trecate et al. 2003; Warner et al. 2004). Kuhl et al. (2005b) recruited women with over a 20 per cent lifetime risk of breast cancer.

Surveillance consisted of XRM, US and MRI in all the studies. Podo et al. (2002) conducted surveillance with CBE, XRM, US and MRI every year. The other three studies all provided six-monthly CBE. There were six-monthly US examinations conducted by Kuhl et al. (2005b). Trecate et al. (2003) and Warner et al. (2004) conducted annual US. These three studies all performed MRI annually too. Contrast enhanced MRI was used in all four studies, as were US probes with a frequency of 7.5 MHz or over.

The outcomes from these studies were cancer detection rates, measures of accuracy, tumour characteristics and interval tumours. Kuhl et al. (2005b) was the only study to examine the tumour characteristics by modality of surveillance and found no significant differences. Warner et al. (2004) also documented mortality but did not have sufficient follow-up for this to be meaningful.

The cancer detection rates and measures of accuracy of the individual and combined surveillance strategies are summarised in **Tables 27** and **28**.

Table 27. Cancer detection rates with combination screening strategies in women at high risk of breast cancer

Study	Cancer Detection Rate				
	CBE	XRM	US	MRI	XRM +MRI
Podo et al. (2002)	N/R	9 per1,000 w/s	9 per 1,000 w/w	67 per 1,000 w/s	N/R
Trecate et al. (2003)	130 per 1,000 w/s	0	0	170 per 1,000 w/s	N/R
Warner et al. (2004)	8 per 1,000 w/s	34 per 1,000w/s	30 per 1,000 w/s	72 per 1,000 w/s	N/R
Kuhl et al. (2005b)	2 per 1,000 w/s	26 per 1,000 w/s	32 per 1,000 w/s (XRM + US, 40 per 1,000 w/s)	74 per 1,000 w/s	76 per 1,000 w/s

w/s = women under surveillance N/R = not reported

Table 28. Measures of accuracy for combination screening strategies in women at high risk of breast cancer

Study	Modality of screening					
	XRM	US	XRM + US	MRI	MRI + XRM	Overall
Warner et al. (2004)						
Sensitivity	36.3% (17.1-59.3%)	33% (14.6-66.9%)	64% (with CBE)	77.3% (54.6-92.2%)	86% (with CBE)	95%†
Specificity	99.8% (98.7-99.9%)	96% (93.7-97.7%)	N/R	95.4% (92.9-97.2%)	N/R	
PPV	88.9% (51.7-99.7%)	29% (12.6-51.15)	N/R	45.9% (29.5-63.1%)	N/R	
NPV	96.9% (94.8-98.3%)	97% (94.5-98.2%)	N/R	98.8% (97.2-99.6%)	N/R	
AUC	0.77	0.65	N/R	N/R	N/R	
Kuhl et al. (2005b)						
Sensitivity	32.6% (19.0-48.5%)	39.5% (25.0-55.6%)	48.8% (33.3-64.5%)	90.7% (77.9-97.4%)	93.0% (80.9-98.5%)	93.0%‡
Specificity	96.8% (95.7-97.7%)	90.5% (88.8-92.0%)	89.0% (87.2-90.6%)	97.2% (96.2-98.0%)	96.1% (94.9-97.0%)	
PPV	23.7% ()	11.3% (6.7-17.4%)	11.9% (6.7-17.4%)	50.0% (38.4-61.5%)	42.1% (32.0-52.7%)	
NPV	97.9% (97.0-98.6%)	98.0% (97.1-98.7%)	98.0% (97.1-98.7%)	99.7% (99.2-99.9%)	99.8% (99.3-99.9%)	

N/R = not reported, Overall =CBE+XRM+US+MRI, † no 95%CI reported, ‡is equivalent to MRI+XRM (no additional tumours detected with overall strategy)

The cancer detection rates are higher in all the studies for MRI compared with XRM or US. The measures of accuracy in the study by Warner et al. (2004) show that MRI is significantly more sensitive then either XRM ($p=0.01$) or US ($p=0.009$) alone. The specificity, PPV and NPV are not significantly different between MRI and US, but the specificity and PPV are significantly lower for MRI than XRM ($p<0.01$ and 0.02 respectively). The results from Kuhl et al. (2005b) estimate that the sensitivity of MRI is significantly better than XRM ($p<0.001$) or US alone ($p<0.001$) and the combination of XRM and US ($p<0.01$). Kuhl et al. (2005b) also demonstrate that MRI surveillance maintains equivalently good sensitivity throughout all risk groups, including mutation carriers. The PPV for MRI is higher than that of XRM and US. There is no apparent difference between the sensitivity of MRI and the combination of MRI and XRM.

Warner et al. (2004) also calculated the AUC for various surveillance strategies. The results were 0.65 for US alone; 0.77 for XRM alone or XRM and US combined; 0.81 for CBE, XRM and US; 0.89 for MRI; 0.91 for CBE, MRI and US; 0.93 for CBE, XRM, US and MRI; and 0.94 for CBE, XRM and MRI.

In conclusion, MRI appears to be significantly more sensitive than XRM, US or the combination or XRM and US in the surveillance of women at high risk of breast cancer. It may also be especially effective for women at highest risk, i.e. mutation carriers, as its sensitivity does not decrease as the risk

status increases. However, the specificity and PPV of MRI may be lower than XRM (Warner et al. 2004). This is due to false-positive examinations and as discussed previously has implications for resource use and anxiety in women involved in the surveillance programme. Surveillance with MRI and XRM appears to offer little advantage over MRI alone, although the MARIBS study, in the previous chapter, did suggest an advantage of this combined strategy (Leach et al. 2005). It has been suggested that breast imaging with MRI is still early in its development and that as radiologists gain experience and increase the number of breast MRIs they are reading, and have previous films available for comparison in incidence rounds, that the number of false positives will substantially decrease (Robson 2004; Warner et al. 2004), as occurred with screening XRM. Further research is required to determine whether this will be the case.

Table 29. Primary research studies appraised investigating the accuracy and efficacy of combination surveillance strategies on outcomes from breast cancer

Source	Study design Evidence grading	Comparison interventions and dates of testing	Sample	Outcomes and verification	Results	Comments
Podo et al. (2002) Italian multi-centre study	Prospective cohort study III-2 (C1 P2 Q2)	Surveillance protocol: CBE, mammography, US and MRI at yearly intervals. Mammography: Standard mediolateral oblique and craniocaudal views were obtained of each breast. Further views taken when necessary. Findings reported using the BIRADS system (1, negative; 2, benign; 3, probably benign; 4, suspicious abnormality and 5, highly suggestive of malignancy). US: performed at a frequency of ≥7.5MHz. MRI: Performed on coronal and axial planes. One pre-contrast and 5 post-contrast images were taken. Gd-chelate (0.1 mmol/kg) was injected as contrast. MRI was reported using a system that is based on a combination of morphological and enhancement parameters. (0-2 = benign, 3=uncertain, 4-8=malignancy). In the case of non-benign scores (3-8) which were detected only by MRI, the MRI was repeated after 1-2 months. If the lesion was confirmed then a biopsy was undertaken.	Sample no = 105 patients were enrolled in the first annual round (14 of these women also underwent a second round). Forty (38%) had a previous personal history of breast cancer. Mean age at recruitment 46 years, median age 51 years (age range 25-77 years). Mean age at diagnosis was 55.3 years, median 52.5 (range 35-70 years) Recruited from 9 cancer genetics centres within Italy. Inclusion criteria: • very high risk of breast cancer; • women ≥25 years age; • men ≥50 years age • women who had personal history of breast cancer were allowed if unilateral. Unilateral mammography done if had had a mastectomy and bilateral if had had breast conservation;	Relevant outcomes: Cancer detection rate. Mode of detection. Tumour size, stage and node status. Verification of positive findings is by biopsy (either MRI or US guided) and pathology. Verification of negative findings is through follow-up – it is acknowledged that these are preliminary findings and the follow-up is incomplete.	Cancer detection rate: 8 tumours were detected in total; 7 in the prevalent round and 1 in the incident round. 5 of these patients had a previous personal history of breast cancer 3 were BRCA1 mutation carriers, 3 were BRCA2 mutation carriers and 2 with unknown mutation status. Mode of detection: 7 (88%) were detected by MRI. Both mammography and US detected only 1 tumour. (MRI had 1 false positive but mammography and US had none) Tumour size, stage and node status: 2 invasive ductal carcinomas 2 invasive lobular carcinomas 1 invasive ductal and lobular carcinoma 2 DCIS 1 DCIS and LCIS Tumour size ranged from 3-27mm. There were no node-positive tumours. The follow-up is incomplete and therefore sensitivity and specificity cannot be calculated.	Limitations: Only the preliminary report of this study. Verification bias, particularly in this study (as acknowledged by the authors) as it is just a preliminary report and sufficient follow-up of negative results has not yet been achieved. This cohort varies from other studies as it is a very high risk group and includes a high proportion of women with a personal history of breast cancer. No comment on women undertaking risk reducing strategies such as on Tamoxifen or having had a bilateral salpingo-oophorectomy. Authors' conclusions: The findings of this study substantiate those of existing studies, that MRI is a more sensitive and more accurate imaging modality than conventional imaging for detecting breast cancer in women at a high risk of this disease (both pre- and post-menopausal women). A previous personal history of breast cancer was associated with higher probability of breast cancer detection during surveillance. The authors conducted a review of other existing literature and performed a meta-analysis of the results of the studies to date. They note that there are considerable differences in the design of these studies, but state that there are some consistent conclusions. The overarching finding is that MRI is more sensitive and significantly more accurate than conventional imaging in the surveillance of women at a high risk of breast cancer.

Table 29. Primary research studies appraised investigating the efficacy and accuracy of combination surveillance strategies on outcomes from breast cancer (continued)

Source	Study design Evidence grading	Comparison interventions and dates of testing	Sample	Outcomes and verification	Results	Comments
Podo et al. (2002) Italian Multi-centre study *Continued*		Pre-menopausal women had MRI within the 2nd week of the menstrual cycle. Dates of surveillance: June 2000 to March 2002 (preliminary report of first phase, 21 months, of the study).	• if on HRT, were included but this was stopped and surveillance not started until been off it for 3 months Exclusion Criteria: • pregnancy; • breast feeding; • current chemotherapy; • terminal illness; • specific contraindications to MRI. Risk stratification: Only recruited subjects who were known BRCA1 or BRCA2 mutation carriers, or had a 1 in 2 probability of being a carrier (first-degree relative who was a proven mutation carrier). 2 women also included whose families had a very high risk or incidence of breast cancer that was likely associated to a non BRCA1 or BRCA2 mutation. 40 of the 105 women also had a personal history of breast cancer.			They point to the need for more extensive, multi-centre and multi-national trials on the evaluation of benefits and costs associated with the introduction of MRI into appropriate surveillance programmes specifically addressed to subjects at high genetic risk of breast cancer. Reviewers' conclusions: This study does appear to show an advantage of MRI surveillance in women at high risk of breast cancer. However, these are only preliminary results of this study and measures of accuracy could not be calculated without further follow-up data. Unfortunately a further report of this work cannot be found and it is perhaps ongoing. These results are also limited in their external validity by being from a very high risk cohort, especially as a high proportion of women with a personal history of breast cancer were included. The pulling together of results from other studies was hampered by variation in the design of the studies and also the outcomes measured.

Table 29. Primary research studies appraised investigating the accuracy and efficacy of combination surveillance strategies on outcomes from breast cancer (continued)

Source	Study design Evidence grading	Comparison interventions and dates of testing	Sample	Outcomes and verification	Results	Comments
Trecate et al. (2003) Italy (NB: Podo is an author on this one as well but we cannot find any further reports from the Podo et al trial.)	Prospective cohort study III-2 (C1 P2 Q3)	Surveillance protocol: Outlined in full in the paper and was dependent on age group, CBE was performed every 6 months for all ages. Mammography was annual and commenced at 25 years with bilateral one-view, and then increased to bilateral double-view from 30 years and above. Double-view was performed in craniocaudal and mediolateral oblique projections. One-view was performed in the mediolateral oblique projection for younger women. Annual US was performed alone from 20-25 years, then with mammography from 25-35 years, then 6 months after mammography from 35-40 years and above 40 years only if requested by the radiologist. US was performed with either7.5MHz or 10-12MHZ probes (ATL HDI 3500, Philips). MRI was performed annually for all ages for 2 years during the study. A Siemens Vision 1.5 was used with a dedicated double coil.	Sample no = 23 women at high risk of breast cancer (2 cases did not get US). No average age of women given, range was 30-61 years. Inclusion criteria: • BRCA1 or BRCA2 mutation carrier or 1 in 2 probability to be a mutation carrier on the basis of positive mutational analysis in close relatives. • With a negative or positive personal history for breast or ovarian cancer. OR • High risk for breast cancer according to criteria specified in paper. Risk stratification: As above, either BRCA1 or BRCA2 carrier, 1 in 2 probability of being a carrier or >50% risk of carrying a susceptibility gene for familial breast cancer on basis of family history.	Relevant outcomes: Cancer detection rate. Mode of detection. Tumour size and stage. Verification of positive results was with pathology and verification of negative results was with follow-up. There is no mention of the mean length of follow-up.	Cancer detection: 4 breast cancers were detected overall. Mode of detection: All 4 tumours were detected by MRI, 3 were detectable by CBE but none of the tumours were detected by mammography or US examination (although 1 woman did not receive an US). It is stated that there were no false-positives or false-negatives for MRI. Tumour size and stage: All 4 tumours were invasive: 2 ductal invasive carcinomas, 1 lobular invasive carcinoma and 1 which was mixed ductal and lobular. 2 occurred in mutation carriers and 2 in women at high risk through family history. Only 2 tumours had the size recorded and these were 10mm and 30mm. No record of nodal status was given. There was no mention of interval tumours.	Limitations included: Small sample size. There are few characteristics given of the women selected other then their risk assessment. There is no information on how they were selected and the characteristics of any women who did not agree to participate. There is no mention of mean age, reproductive history, exogenous hormone use or preventative strategies (i.e. Tamoxifen use or BSO). There is also no indication of which women were having prevalent or incident surveillance screens and for how long they were followed up in the study. There is likely verification bias and this is more likely, the shorter the follow-up period. Authors' conclusions: Breast MRI demonstrated to be a very useful technique for investigating breast disease. It is not influenced by breast density and does not use ionising radiation. For these reasons, it has been proposed to support mammography in the surveillance of BRCA mutated patients. Moreover, according to the reported results, breast MRI seems very helpful in the high-risk patients group. We believe the breast MRI can be very useful within this kind of surveillance, with a less invasive approach to the disease. In the case of confirmed good diagnostic results, it could be proposed to be used every other year as an alternative to mammography. Reviewers' conclusions: This study suggests that MRI is a very effective tool for the surveillance of women at high risk of breast cancer.

Table 29. Primary research studies appraised investigating the accuracy and efficacy of combination surveillance strategies on outcomes from breast cancer (continued)

Source	Study design Evidence grading	Comparison interventions and dates of testing	Sample	Outcomes and Verification	Results	Comments
Trecate et al. (2003) Italy *Continued*		One pre-contrast image and 5 post-contrast images were taken. The contrast agent was Gd-DTPA at 0.1mmol/kg. The method of interpreting the MRI or mammography is not presented. The study was conducted over a 7-month period; however the exact dates are not given.	The latter refers to at least 3 cases of breast cancer before 60 years of age, at least 3 cases of breast cancer before 60 years of age and ovarian cancer at any age, or at least 3 cases of breast cancer before 60 years of age and male breast carcinoma at any age. 5 of the women had a personal history of breast cancer, 1 for ovarian cancer and 1 for ovarian and breast cancer. (1 had had a mastectomy, but the others had conservative surgery combined with radiation therapy).			However, the sample is very small and it is difficult to know how long the women were followed up for and this would affect the reliability of the results. There could be false-negatives that had not yet come to light. There is also a specific method of risk stratification in this study, which includes women with a personal history of breast cancer, although only if they are BRCA1 or BECA2 mutation carriers, and this will affect the generalisability of the study. In addition, the results are not presented in a very clear manner and it is difficult to determine the overall sensitivity and specificity for all the modalities of surveillance utilised, which would have been valuable information.

Table 29. Primary research studies appraised investigating the accuracy and efficacy of combination surveillance strategies on outcomes from breast cancer (continued)

Source	Study design Evidence grading	Comparison interventions and dates of testing	Sample	Outcomes and verification	Results	Comments
Warner et al. (2004) Ontario and Montreal, Canada	Prospective cohort Study III-2 (C1 P2 Q2)	Study protocol: CBE biannually and mammography, US and MRI all performed annually 4 modalities all performed the same day. (commencing at least 1 year after the woman's last mammogram) CBE coded as normal, suggestive of benign disease, indeterminate, or suspicious of malignancy. Indeterminate CBE exams were repeated after 3 months. Mammography was conventional 4-view film. Further views done when necessary. MRI was performed with 1.5 T magnet (Signa, General Electrical Medical Systems). The first 38 patients in the first year were done in a single-turn elliptical coil after a bolus injection of 0.1mmol/kg of Gd-DTPA. Images were taken in the coronal plane. For the remaining patients, a phased-array coil arrangement was used. This provided sagital images. US used a 7.5MHz transducer (the first 7 patients did not receive US).	Sample no = 236 female BRCA1 and BRCA 2 mutation carriers. Mean age at first surveillance 46.6 years (range 25-65 years) Mean age of diagnosis was 47.4 years (33.4-63 years) Recruited from Familial cancer clinics. Inclusions: • BRCA 1 or BRCA2 mutation carrier. Exclusions: • past history of unilateral breast cancer if the contra lateral breast not intact; • pregnant or lactating women (participation deferred); • history of bilateral breast cancer, currently undergoing chemotherapy or known to have metastatic disease; • women weighing >91kg (technical reasons.)	Relevant outcomes: Cancer detection rate. Mode of Detection. Tumour stage, size and node status. Interval cancers. Mortality. Sensitivity. Specificity. PPV. NPV. ROC curves. NB: the PPV and specificity do not include in the denominator women that had additional diagnostic studies that did not result in biopsy. Verification of positive results was by pathology, biopsy was undertaken if there was suspicion from any of the four modalities of surveillance. Verification of a negative result was through follow-up. All patients were followed up for a minimum of 1 year from the date of the last surveillance examination.	Cancer detection: 22 cancers were detected in 21 women (1 bilateral). (7 of these women had previous breast cancer). Mode of detection: 2 were detected by CBE (9.1%) 8 by mammography (36%) 7 by US (33%) 17 by MRI (77%) 7 cancers (32%) were detected by MRI alone, 2 cancers (9.1%) were detected by mammography alone, 2 were detected by US alone (9.5% though not all women had undergone US testing). Therefore, MRI detected 9 of the 12 cancers missed by conventional surveillance (mammography plus CBE). Tumour stage, size and node status: 6 tumours were DCIS and 16 were invasive (15 infiltrating ductal and 1 invasive lobular). The mean size of the invasive tumours was 11mm at the first round and 13mm at the second round. (overall range 5-60mm). 15 cases were node sampled and 2 were node-positive. Interval cancers: There was only 1 interval cancer, detected in a 40 year old BRCA1 mutation carrier 7 months after her 3rd surveillance screen (retrospectively this tumour was visible on MRI and on mammography at last surveillance visit).	Limitations: Likely verification bias. Selected participants are very high risk, being proven mutation carriers and also including those with a prior history of breast cancer. It is not clear which were incident and which were prevalent rounds and which tumours were detected at which round (a large number of women had had prior mammography). No mention of whether women had had risk reducing measures such as bilateral salpingo oophorectomy or Tamoxifen. There was a quite high level of attrition in the study and the characteristics of those women are not outlined. This may have introduced bias. Authors' conclusions: This study of BRCA mutation carriers demonstrates that the addition of annual MRI and US to mammography and CBE significantly improves the surveillance for detecting early breast cancers. The use of US did detect additional tumours, but had a high false-positive rate and in light of this its benefit remains to be seen. There was no observed benefit from CBE over and above the 3 imaging modalities. MRI-based surveillance is likely to become the cornerstone of breast cancer surveillance for BRCA1 and BRCA2 mutation carriers, but it is necessary to demonstrate that this surveillance tool lowers breast cancer mortality before it can be recommended for general use.

Table 29. Primary research studies appraised investigating the accuracy and efficacy of combination surveillance strategies on outcomes from breast cancer (continued)

Source	Study design Evidence grading	Comparison interventions and dates of testing	Sample	Outcomes and Verification	Results	Comments
Warner et al. (2004) Ontario and Montreal, Canada *Continued*		Each imaging modality was read independently by a radiologist and scored on the 5 point BIRADS scale. All lesions with a score of 4 or 5 were biopsied. Pre-menopausal women had surveillance performed mid menstrual cycle to avoid changes due to cyclical hormonal variation. Radiologists were blinded to the results of CBE. 31 women left the study before completing 3 rounds, 16 underwent bilateral mastectomy, 3 were too large for MRI machine, 3 stopped due to pregnancy, 4 developed metastatic cancers, 4 were lost to follow-up and 1 did not wish to continue participating. All participants underwent the first round, but only 58% the second and 36% the third (a total of 120 women were still undergoing surveillance when the paper was written). No direct comparisons were made in this study.	Risk stratification not really performed as only BRCA mutation carriers included. (all very high risk group). There were 137 (58%) BRCA1 mutation carriers and 99 (42%) BRCA 2 mutation carriers. 31% were Ashkenazi Jews. 30% had a history of breast cancer, 9% a history of ovarian cancer and 60% had no history of cancer or a history of another type of cancer. 85% of the women (*n*=205) had had mammography within the last 15 months and therefore this was an incident rather than a prevalent round for them. 45% were pre-menopausal and 55% were post-menopausal.		Another woman, who elected to have a bilateral mastectomy after breast cancer was found, had a 2mm focus of DCIS in the contra lateral breast which had not shown up on screening 2 months earlier. Mortality: All 22 patients who had tumours diagnosed were still alive and disease-free at the time the article was written. It was felt that the cancers detected on the second round were of an earlier stage. The 2 node-positive tumours were detected in the first round. However, it was not exactly clear that the first round was really a prevalent round as a high percentage of women had had prior mammography. It was found that false-positives and false-negatives decreased from the first to the second and then to the third round of surveillance. This is especially seen for the false-positives in MRI, which decreased from 15 to 4 to 1. This may have been due to increasing experience in the radiologists in interpreting these scans. The measures of accuracy are therefore presented by the modality of surveillance and by the year of the surveillance. These can be seen in the paper, but overall values for the 3 years are reported here.	Reviewers' conclusions: This study demonstrates a greater efficacy in a combined approach, using all 4 modalities, in the surveillance of BRCA1 and BRCA2 mutation carriers for breast cancer. As the authors suggest, this does not answer whether this translates into reduced mortality. However, the tumours detected did seem to be of an earlier stage and smaller size, with only 2 tumours node-positive. The results of this study are limited to the very high risk population of women who are proven BRCA1 or BRCA2 mutation carriers and including those with a personal history of breast cancer. It may therefore not be generalisable to all women with an increased risk of breast cancer due to a family history. Further studies with larger numbers and longer follow-up, and including women of other risk groups are required.

Table 29. Primary research studies appraised investigating the accuracy and efficacy of combination surveillance strategies on outcomes from breast cancer (continued)

Source	Study design Evidence grading	Comparison interventions and dates of testing	Sample	Outcomes and Verification	Results	Comments
Warner et al. (2004) Ontario and Montreal, Canada *Continued*		Dates of surveillance were between Nov 1997 and March 2003.			Sensitivities of combinations of modalities: MRI + CBE + XRM+ US = 95% MRI + CBE + XRM = 86% XRM + CBE = 45% CBE + XRM + US = 64% (no 95% CI reported for these) Measures of accuracy of individual modalities: Sensitivity (95% CI): XRM = 36% (17.1 to 59.3%) US = 33% 14.6 to 56.9%) MRI = 77% (54.6 to 92.2%) MRI was significantly more sensitive than either mammography (p=0.02) or US (p=0.006). Specificity (95% CI): XRM = 99.8% (98.7 to 99.9%) US = 96% (93.7 to 97.7%) MRI = 95% (92.9 to 97.2%) (was 99% in 3rd year) PPV (95% CI): XRM = 89% (51.7 to 99.7%) US = 29% (12.6 to 51.1%) MRI = 46% (29.5 to 63.1%) NPV (95% CI): XRM = 97% (94.8 to 98.3%) US = 97% (94.5 to 98.2%) MRI = 99% (97.2 to 99.6%)	

Table 29. Primary research studies appraised investigating the accuracy and efficacy of combination surveillance strategies on outcomes from breast cancer (continued)

Source	Study design Evidence grading	Comparison interventions and dates of testing	Sample	Outcomes and verification	Results	Comments
Warner et al. (2004) Ontario and Montreal, Canada *Continued*					AUC: XRM = 0.77 US = 0.65 MRI = 0.89 CBE = 0.48 MRI + CBE + XRM + US = 0.93 MRI + CBE + XRM = 0.94 MRI + CBE + US = 0.91 CBE + XRM + US = 0.81 XRM + CBE = 0.77	

Table 29. Primary research studies appraised investigating the accuracy and efficacy of combination surveillance strategies on outcomes from breast cancer (continued)

Source	Study design Evidence grading	Comparison interventions and dates of testing	Sample	Outcomes and verification	Results	Comments
Kuhl, et al. (2005b) Germany	Prospective cohort study. III-2 (C1 P2 Q2)	Surveillance protocol: Biannual CBE and US and annual XRM and MRI. If abnormalities found on CBE or US at round without XRM or MRI, these additional modalities were used to further investigate this. Surveillance commenced at 30 years, or 5 years before the youngest family member affected with the disease. (NB: in first 2 years, women under 30, or 30-39 years with dense breasts, did not receive XRM, but this was subsequently abandoned and all women received XRM). Mammography (XRM): Annual conventional film XRM performed with at least 2 views per breast (medio-lateral oblique and caudal-cranial), obtained and interpreted in accordance with German radiological practice guidelines. Diagnoses coded according to the BI-RADStm diagnostic categories on a 5-point scale (1, negative; 2, benign; 3, probably benign; 4, suspicious abnormality; 5, highly suggestive of malignancy).	Sample no = 529 (out of 590 eligible women; 49 were lost to follow-up after 1 surveillance round and 12 were also excluded as they had a clinical abnormality at initial examination) Inclusion criteria: • asymptomatic women; • personal history of breast cancer included provided that the patient had not undergone bilateral mastectomy, had not received chemotherapy within the previous 12 months and had no metastases (139 women were included with a personal history of breast cancer) Exclusion criteria: • clinical signs of breast cancer; • chemotherapy within the previous 12 months; • women having undergone bilateral mastectomy.	Relevant outcomes: Cancer detection. Mode of detection. Tumour size. Tumour stage. Node status. Interval tumours. Sensitivity. Specificity. PPV. NPV. Verification of a positive result was achieved by histology (for positive imaging studies). Verification of a negative result was achieved by follow-up (for negative imaging studies. If a breast cancer was identified clinically by palpation between surveillance rounds or at the 6-month clinical visit the imaging studies of the previous round were considered false-negative. Mean follow-up was 5.3 years (range 2-7 years). (A total of 1,542 annual surveillance rounds were completed	Cancer detection: A total of 43 breast cancers were identified in 41 patients (11 of these women had a prior history of breast cancer), 40 of these were said to be detectable by imaging. Mode of detection: CBE identified only one tumour (also detected on imaging). XRM identified 14 tumours (only 1 was diagnosed by XRM that wasn't diagnosed by MRI). US identified 17 tumours (2 of these were at the half-yearly CBE and US screen and they were not palpable). US + XRM detected 21 tumours. MRI identified 39 tumours. XRM + MRI identified 40 tumours. Tumour size, stage and node status: Of the 21 cancers detected by XRM and US, 16 were invasive and the rest were DCIS. The invasive cancers had a mean size of 13.9mm and 5 were node-positive. Of the 39 tumours detected by MRI, 31 were invasive and 8 were *in situ*. The invasive tumours had a mean size of 12.4mm and five were node-positive. 14 invasive cancers were detected by MRI that were not detected by XRM or US; these had a mean size of 9mm and none of them were node-positive.	Limitations include: CBE and the imaging studies were performed within a time frame of 8 weeks. Few sample characteristics presented, such as OCP or HRT use, or the use of preventative strategies such as tamoxifen or BSO. Verification bias is likely. Interval tumours are unclearly reported. Lack of blinding to the results of the CBE . Author's conclusions: If US is used in combination with XRM, it can help compensate for some but by far not for all of the shortcomings of XRM, and it causes a substantial number of false-positive diagnoses. If MRI is used for surveillance, XRM proved to be of limited and ultrasound of no additional value. US may however be useful to bridge the relatively long time interval between annual surveillance rounds. Propose that in view of the insufficient diagnostic accuracy of XRM and USS, that breast MRI should be considered an integral part of surveillance programmes for women at high familial risk in particular in documented carriers of pathogenic BRCA mutations. Reviewer's conclusions: Similar to those of the authors above. US surveillance improves that of CBE and XRM alone, but MRI is the most effective, especially in women in the highest risk group. This is because MRI does not lose sensitivity as the risk status increases, as US, XRM and US + XRM do. The combination of XRM and MRI does not seem to offer much advantage over MRI alone. The limitations of this study must be taken into account in the interpretation.

Table 29. Primary research studies appraised investigating the accuracy and efficacy of combination surveillance strategies on outcomes from breast cancer (continued)

Source	Study design Evidence grading	Comparison interventions and dates of testing	Sample	Outcomes and verification	Results	Comments
Kuhl et al. (2005b) Germany *Continued*		Breast MRI: Standard dynamic axial contrast-enhanced breast MRI of both entire breasts was performed on a 1.5T system (NT/INTERA; Philips, Best, the Netherlands) after injection of 0.1mmol/kg body weight gadopentetate dimeglumine (Magnevist, Schering, Berlin, Germany) Ultrasound (US): Performed with 7.5-to 13-MHz probes (Siemens Elegra, GE logic 500 and ATL HDI 5000; Siemens, Erlangen, Germany). The entire breast was systematically examined by the physician who interpreted the study. Diagnoses were scored on a 5-point scale identical to the XRM BIRADS categories. Each imaging study was read and scored independently by a different radiologist who had substantial experience with the respective imaging technique. The readers were informed about the clinical findings from CBE and the risk status of the patient but were blinded to the results of the respective other imaging modalities.	• Recruited from high risk clinics in a single Gynaecology Department Risk Stratification: According to definition of the Consortium on Familial Breast and Ovarian Cancer of the German Cancer Aid corresponding to a lifetime risk of breast cancer of at least 20% (two or more cases of breast cancer on the same side of the family, including at least two cases with onset before age 50 years, or with breast or ovarian cancer, irrespective of age, families with at least one case of breast cancer diagnosed before 35 years, families with three or more cases of breast cancer on the same side of the family, and women who met the criteria for high familial risk, irrespective of the result of mutational analysis) In women without a personal history of breast cancer the Claus tables were also used to quantify risk.	). Verification of last surveillance round was by continued surveillance in 428 women, telephone interview in 52 women and for 6 women who had prophylactic mastectomy it was by pathology of the specimen. XRM: BIRADS of 4 or 5, biopsy was recommended irrespective of finding in US or MRI. BIRADS 3 was managed by 6-months follow-up until receiving a BIRADS 2 or biopsy clarification. US categorised as BIRADS 3 managed by short-term (6 months) US follow-up. BIRADS 4 or 5 managed by US-guided biopsy (14G, semi-automatic or automatic biopsy gun) except for the following constellation: if an US finding that was suspicious was clearly benign on XRM or MRI no biopsy was performed. MRI: Suspicious scores (4 or 5) were managed by magnetic resonance-guided biopsy.	Interval tumours: The paper states that 40 out of 43 tumours in this cohort were detected by imaging. However, a sentence in the discussion states that the rate of interval cancers was 2% in this cohort. This translates to 10 tumours if it is 2% of the total population or 1 tumour if it is 2% of the total number of tumours detected. The latter is more likely but it is unclear. Comparisons: When stratified by risk groups, the detection rates, at both the prevalent and incident rounds, were much higher in the mutation carriers than the other 2 risk groups, but these differences are not statistically significant. Sensitivity (95% CI): XRM 32.6% (19.0 to 48.5%) n = 14/43 US 39.5% (25.0 to 55.6%) n = 17/43 XRM+US 48.8% (33.3 to 64.5%) n = 21/43 MRI 90.7% (77.9 to 97.45) n = 39/43 MRI+XRM 93.0% (80.9 to 98.5%) n = 40/43 When stratified by risk groups XRM, US and the combination of XRM+US all become less sensitive as the lifetime risk of breast cancer increases, with sensitivities of 25%, 25% and 37.5% respectively for the mutation carrier group.	

Table 29. Primary research studies appraised investigating the accuracy and efficacy of combination surveillance strategies on outcomes from breast cancer (continued)

Source	Study design Evidence grading	Comparison interventions and dates of testing	Sample	Outcomes and verification	Results	Comments
Kuhl et al. (2005b) Germany *Continued*		Comparisons are made between the 3 risk groups and the different modalities of surveillance. Dates of study were February 1996 to February 2002.	Women were then stratified into 3 risk groups for analysis: Mutation carriers. High lifetime risk (20-40%). Moderate lifetime risk (20%.)	Findings categorized as BI-RADS 3 short-term follow-up after 6 months was recommended with further management corresponding to that of XRM BIRADS 3 lesions. BIRADS 3 categories in all imaging that received short-term follow-up were not considered positive for the calculation of outcomes. Invasive cancer and DCIS were considered a malignant diagnosis but LCIS and atypical ductal hyperplasia were considered to be benign.	This effect is not seen with MRI which maintains good sensitivity across all risk groups. Sensitivity of 100% is given for MRI and MRI+ XRM for all risk groups. However, the denominators is smaller than the entire group of women and it is unclear why this figure was used (34 instead of 43) Specificity: XRM 96.8% (95.7 to 97.7%) n = 1364/1409 US 90.5% (88.8 to 92.0%) n = 1275/1409 XRM+US 89.0% (87.2 to 90.6%) n = 1254/1409 MRI 97.2% (96.2 to 98.0%) n = 1370/1409 MRI+XRM 96.1% (94.9 to 97.0%) 1354/1409 Stratification by risk group does not appear to affect the specificity. PPV: XRM 23.7% (14 to 37%) n = 14/59 US 11.3% (6.7 to 17.4%) n = 17/151 XRM+US 11.9% (7.5 to 17.6%) n = 21/176 MRI 50% (38.4 to 61.5%) n = 39/78 MRI+XRM 42.1% (32.0 to 52.75) n = 40/95 The PPV increases with the increasing risk of breast cancer, this will be affected by the higher incidence in women at higher risk.	

Table 29. Primary research studies appraised investigating the accuracy and efficacy of combination surveillance strategies on outcomes from breast cancer (continued)

Source	Study design Evidence grading	Comparison interventions and dates of testing	Sample	Outcomes and verification	Results	Comments
Kuhl et al. (2005b) Germany *Continued*					NPV (95% CI) XRM 97.9% (97.0 to 98.6%) n = 1364/1393 US 98% (97.1 to 98.7%0 n = 1275/1301 XRM + US 98.3% (97.4 to 98.95) n = 1254/1276 MRI 99.7% (99.2 to 99.9%0 n = 1370/1374 MRI + XRM 99.8% (99.3 to 99.9%) n = 1354/1357	

Chapter 9: Discussion

SUMMARY OF EVIDENCE

This report systematically reviewed the international evidence for the surveillance of women at high risk of breast cancer.

Approximately 2,780 articles were identified by the search strategy. From 156 articles identified as potentially eligible for inclusion, a final group of 38 papers were selected for appraisal. There were four systematic reviews. There were no randomised controlled trials included in the systematic reviews or identified by the search strategy. The main results are presented below by modality of surveillance.

One systematic review and 24 primary studies were identified which looked at the accuracy and efficacy of surveillance with XRM in women at high risk of breast cancer. The strategies in these studies consisted of CBE and XRM, with or without additional modalities of surveillance. There was considerable heterogeneity between the studies in terms of the surveillance conducted, the level of risk of the participants, the age groups included and the inclusion or exclusion of women with a past history of breast cancer. The studies were frequently limited by the small number of participants and the relatively few tumours that arose during the study period. The heterogeneity between studies prevented any meta-analysis of the results.

Three principal comparisons were used in these studies to assess the efficacy of surveillance with XRM in high-risk women. The first was to demonstrate that surveillance with XRM had a higher cancer detection rate and was more accurate than CBE alone. Overall there were higher cancer detection rates and sensitivities with XRM surveillance compared with CBE. This is logical as tumours are only detectable by CBE once they have reached approximately 10mm in size (Hughes et al. 1999). There were two studies in which CBE surveillance performed better than XRM (Gui et al. 2001; Trecate et al. 2003). However, the results of these studies were unreliable due to the method of analysis of the results and the small sample size respectively.

The second comparison was of cancer detection rates from the study population and those of established breast screening programmes for women of all risk groups over the age of 50 years. Eight studies demonstrated similar rates of detection (six to the NHSBSP, one to the Italian BSP and one to the Dutch BSP). The assumption behind this comparison was that if surveillance in women at high risk of breast cancer detected cancers at an equivalent rate to established BSPs then it should be equally acceptable to adopt. This does not consider the potential harms of commencing XRM in younger women and exposing them to ionising radiation over a considerable period of time, especially as women at high risk require more regular surveillance due to the aggressive nature of tumours in this population. In women who are mutation carriers this is of particular concern as it is thought that these mutations affect DNA repair mechanisms and place them at higher risk of radiation-induced tumours.

The third comparison made in these studies was of tumour characteristics between a surveillance population and a population that didn't receive surveillance. Only two studies performed such a comparison (Kollias et al. 1998; Macmillan 2000). Both these studies demonstrated a significantly higher proportion of *in situ* tumours in the surveillance population as opposed to the population without surveillance. However, rather than early diagnosis this could potentially represent overdiagnosis of lesions that would never have been diagnosed in the women's lifetimes without surveillance. Macmillan et al. (2000) also demonstrated a significantly higher proportion of tumours with a good prognostic index in the surveillance group compared with the population without surveillance. No significant difference in the lymph node status could be demonstrated in either study. The assumption behind this comparison is that detecting tumours at an earlier stage can lead to early treatment and that this may translate to a decrease in mortality. However, consideration needs to be taken of whether any demonstrated survival advantage may be a product of lead-time bias or length bias. The stage of breast cancer at diagnosis has certainly been shown to relate to prognosis (Bland et al. 1998) and a decrease in mortality has been demonstrated in women over the age of 50 years as a result of XRM screening (Tabar et al. 2000). However, it cannot be assumed that this would necessarily occur in this population

of high-risk women. The natural history of tumour in high-risk women, and their response to treatment, may differ from tumours in women at average risk.

All of the aforementioned were intermediate outcomes. Very few studies looked at outcomes such as survival or mortality. In those that did, the results were unreliable due to the short period of follow-up and the small numbers involved. One study (Maurice et al. 2006) did suggest a significant decrease in mortality associated with the surveillance of women at high risk of breast cancer.

Overall, surveillance with XRM appeared superior to CBE alone, has a cancer detection rate equivalent to established BSPs and may detect tumours at an earlier stage. There was still a high number of interval tumours arising in these studies, in some cases almost twice as many as were detected by XRM (Gui et al. 2001). This suggests that in women at high risk of breast cancer, surveillance with XRM and CBE alone is not adequate and additional modalities of surveillance are required.

Ultrasound has traditionally been used as a diagnostic test to examine the breast rather than for screening or surveillance. However, with improving technology and the availability of higher frequency probes, a role for US in surveillance has been considered. Nine studies were identified of relevance to the accuracy and efficacy of surveillance of women at high risk of breast cancer with US. Four of these compared the cancer detection rate and measures of accuracy of US to CBE alone. All nine of these compared US surveillance with XRM. Four of these studies included MRI in their surveillance strategies and one study (Sim et al. 2004) re-analysed data on US from an existing study of XRM, US and MRI (Stoutjesdijk et al. 2001). Selection bias may have been introduced to the study by Sim et al. (2004), as the use of US in the original study appears to have been more diagnostic than surveillant. Two studies (Crystal et al. 2003; Kolb et al. 1998) examined US as an adjunct test in women who had normal findings on XRM and CBE. This is likely to have overestimated the efficacy of detecting early breast cancer by US surveillance of women at high risk.

Surveillance with US appears to be more sensitive than surveillance with CBE alone in women at high risk of breast cancer. However, the sensitivity is still relatively low. The low sensitivity of surveillance with CBE and US alone suggests that there would be a high rate of interval cancers. This cannot be assessed due to the design of these studies, which include XRM in their surveillance protocols. If this was the case it would suggest that CBE and US alone are not adequate for the surveillance of women at high risk of breast cancer and that other modalities of surveillance are required in addition, or instead, of this strategy.

Surveillance in women at high risk of breast cancer with US has equivalent sensitivity and NPV to XRM, but US has a lower specificity and PPV. The combination of XRM and US has an increased sensitivity over either modality alone, but still maintains the lower specificity and PPV associated with US. There was no evidence in terms of improved survival due to the early detection of cancerous breast lesions. US has the advantages of not using ionising radiation for surveillance and being a useful tool for biopsy. However, the number of false-positives generated is a disadvantage as it would lead to anxiety and a higher rate of invasive investigations. Due to this, US may remain a diagnostic tool and other modalities of surveillance, if available and affordable, may be required in women at high risk of breast cancer.

One such imaging technology which has been suggested for the surveillance of women at high risk of breast cancer owing to family history or genetic predisposition is MRI. The use of MRI of the breast has been suggested in other clinical situations such as pre-operative evaluation in patients with localised breast cancer, and in the evaluation of suspicious mammography findings referred for biopsy. The role of MRI in the evaluation of patients with known or suspected breast cancer remains controversial. The objective of this review as it pertains to the evaluation of MRI was to ascertain the effectiveness and accuracy of MRI surveillance women at high risk of breast cancer.

No studies were found that compared MRI surveillance with no surveillance at all. Therefore, it is unknown if there is any benefit of surveillance with MRI over no surveillance at all in these women in terms of improved survival. There is also no information on whether the detection of cancerous breast lesions by MRI surveillance leads to improvements in response to cancer treatment, on the basis that smaller cancers detected early are more amenable to treatment.

Four studies were identified that compared surveillance with MRI to surveillance by clinical breast examination (CBE). Surveillance with MRI appears to be superior to CBE for the detection of breast cancer in women at high risk of breast cancer owing to family history or genetic predisposition. The specificity scores for MRI were also high, indicating that the testing in these two studies is associated with a relatively small number of false-positive test results. At this point, there are no data on whether the early detection of cancerous lesions by MRI surveillance or by CBE confers any benefits in terms of survival.

Two systematic reviews and 10 primary studies were identified that compared surveillance with MRI to surveillance using mammography. Of the 10 primary studies, eight were prospective cohort studies and two were retrospective cohort studies and a total of 4,428 women underwent surveillance. MRI was used in all the included studies and images were taken before and after the bolus injection of contrast enhancement. Data from the use of mammography as a surveillance test were also used in all the included studies. The MRI surveillance tests were performed annually; although in one study (Kriege et al. 2004) clinical breast examination was performed at six-month intervals. Mammography was usually performed within stipulated timeframes from the MRI test in the studies that compared MRI to mammography. In MARIBS (Leach et al. 2005), the mammography test was performed on the same day as the MRI scan, while in Lehman et al. (2005), the comparator test was to take place within 90 days of the first test. In Kuhl et al. (2005b) the comparator tests could take place eight weeks apart. It is unknown at this time if that interval between tests is too wide, given the aggressive nature of the tumours found in women with BRCA mutations and other women with family history placing them at high risk of breast cancer. Also unknown is the optimum surveillance interval and whether women at high risk of breast cancer should receive surveillance annually, or six-monthly.

Measures of diagnostic test accuracy (sensitivity, specificity and positive predictive values) were outcomes examined in the more recent trials (Kriege et al. 2004; Kuhl et al. 2005b; Leach et al. 2005; Warner et al. 2004). Data were also reported for particular subgroups such as women with BRCA mutations and also at varying levels of risk for breast cancer based on family history in Kuhl et al. (2005b) and MARIBS (Leach et al. 2005). There were few measures of effectiveness studied in the included trials comparing XRM and MRI. None of the studies examined survival outcomes, aside from Warner et al. (2004), who stated that all the women undergoing surveillance were still alive at the time of writing. In addition, there were no data presented on response to treatment as a result of the possible earlier diagnosis of cancerous tumours by MRI surveillance.

Surveillance with MRI appears to be associated with substantially higher sensitivity scores than for mammography in terms of detecting cancers in women at high risk of breast cancer owing to familial or genetic history. However the results are based on a relatively small numbers of cancers detected so should be interpreted with some caution.

The difference in sensitivity between MRI and mammography is particularly pronounced in BRCA carriers in the studies where this has been examined. The MARIBS study (Leach et al. 2005) included data for women who had a first-degree relative with a confirmed positive genetic test for a mutation in the BRCA1 gene. Given that these women have a higher absolute risk in the age-range studied in MARIBS (Leach et al. 2005) than the other risk groups, MRI surveillance might be particularly useful in this group. In women with BRCA2 mutations, the gain was smaller and not statistically significant. The data in the Kuhl et al. (2005b) study relating to mutation carriers is also useful but should be interpreted with some caution as the results are based on a very low number of cancers detected.

In the studies in which MRI was examined alongside mammography, the specificity of MRI was relatively high, although in most cases lower than that for mammography. It may be that the increase in MRI scans being examined by the investigators over time resulted in a decrease in false positive results at later time periods within the studies. For example, in the study by Warner et al. (2004) after the first round of surveillance, 16.5 per cent of participants underwent a diagnostic MRI scan to clarify the status of an indeterminate or possibly suspicious lesion. The rates of referrals for either a follow-up diagnostic scan or a biopsy reduced over subsequent rounds of surveillance. This is potentially suggestive of a learning effect whereby those reading the MRI scans become more skilled as a result of increased experience, and the availability of previous films for comparison, resulting in a decrease in false positive results over the course of the study period.

In the studies by Kriege et al. (2004) and MARIBS (Leach et al. 2005) the area under receiver operator characteristic (ROC) curves was studied. The scores for mammography and for MRI were remarkably similar in these two studies, and in both studies the difference between mammography and MRI reached statistical significance. In Kriege et al. (Kriege et al. 2004) the area under the ROC for mammography 0.686, and the corresponding value for MRI was 0.827. The difference between scores was 0.141 (95%CI 0.02-0.262, $p<0.05$). In MARIBS (Leach et al. 2005) the area under the receiver operator characteristic curve for MRI was 0.85 (95%CI 0.84, 0.87) compared to 0.70 (95%CI 0.68, 0.72) for mammography and this difference was statistically significant ($p=0.035$). This would appear to indicate that in both these studies, MRI screening could better discriminate between those with and without breast cancer.

In conclusion, MRI appears to be more sensitive than mammography for the detection of breast cancers in women at high risk of breast cancer owing to genetic or family history. The increase in sensitivity is particularly noticeable in women who carry mutations in BRCA1. However, the specificity of MRI is lower than that of mammography, which has implications for resource use and anxiety of those undergoing surveillance.

For the comparison between stand-alone MRI surveillance and a combination regimen of MRI surveillance with mammography, two prospective cohort studies were identified. A total of 1,128 women received surveillance in these two studies, which were generally well conducted. However, there were no statistical tests calculated by these investigators. The inclusion criteria in the two studies were broadly similar.

There appears to be little difference in the sensitivity and specificity in the Kuhl et al. (2005b) trial between MRI alone and MRI plus mammography. However in the study by MARIBS (Leach et al. 2005) there appears to be a substantial increase in sensitivity when using MRI compared with MRI and mammography combined. This is a result that should be interpreted with some caution, as there were only four additional cancers detected with the use of MRI and mammography. Indeed all the results are based on small numbers of cancerous lesions being detected. There is little difference in specificity when MRI and mammography are used together in this study, a result that is consistent with that of Kuhl et al. (2005b). This would suggest that the use of two imaging modalities does not reduce the numbers of false-positive results that were observed when MRI was used alone.The studies were also not powered to detect meaningful differences between stand-alone surveillance with MRI and combination surveillance with mammography and MRI, in the mutation-carrying subsets of the enrolled populations.

In conclusion, two studies were identified that examined the comparison between MRI surveillance and a combination of MRI and mammography surveillance. There was no evidence in terms of improved survival due to the early detection of cancerous breast lesions. The sensitivities were high in both studies, suggesting that each surveillance regimen is efficacious for detecting tumours in women at high risk of breast cancer owing to family or genetic predisposition. However it is not clear whether combination surveillance offers any additional benefit over surveillance with MRI alone. There was little difference in the specificity in each of the trials between MRI alone and MRI plus mammography.

Four of the studies on surveillance with MRI included US as well as XRM in their strategies (Kuhl et al. 2005b; Podo et al. 2002; Trecate et al. 2003; Warner et al. 2004). The outcomes from these studies were cancer detection rates, measures of accuracy, tumour characteristics and interval tumours. Kuhl et al. (2005b) was the only study to examine the tumour characteristics by modality of surveillance and found no significant differences. Warner et al. (2004) also documented mortality but did not have sufficient follow-up for this to be meaningful.

The cancer detection rates were higher in all the studies for MRI compared with XRM or US. Only two studies reported measures of accuracy. The study by Warner et al. (2004) suggests that MRI is significantly more sensitive then either XRM ($p=0.02$) or US ($p=0.006$) alone. The specificity, PPV and NPV are not significantly different between MRI and US, but the specificity and PPV are significantly lower for MRI than XRM ($p<0.01$ and 0.02 respectively). The results from Kuhl et al. (2005b) also suggest that the sensitivity of MRI is significantly better than XRM ($p<0.001$) or US alone ($p<0.001$) and the combination of XRM and US ($p<0.01$). The PPV for MRI is higher than that of XRM and US. There is no apparent difference between the sensitivity of MRI and the combination of MRI and XRM shown in the study by Kuhl et al. (2005b). This differs from the results of the MARIBS

study (Leach et al. 2005). MRI also appears to be especially advantageous in women at the highest risk (mutation carriers) as, unlike XRM and US and their combination, its sensitivity does not decrease with increased risk status.

Warner et al. (2004) also calculated the AUC for various screening strategies. The results were 0.65 for US alone; 0.77 for XRM alone or XRM and US combined; 0.81 for CBE, XRM and US; 0.89 for MRI; 0.91 for CBE, MRI and US; 0.93 for CBE, XRM, US and MRI; and 0.94 for CBE, XRM and MRI.

Overall, MRI appears to be significantly more sensitive than XRM, US or the combination or XRM and US in the surveillance of women at high risk of breast cancer. However, the specificity and PPV may be lower than XRM (Warner et al. 2004). As discussed for all studies of MRI, these results are based on a very small number of cancers detected and this reduces their reliability. The low specificity and PPV of MRI surveillance is due to false-positive examinations and as discussed previously has implications for resource use and anxiety in women involved in the screening programme. It has been suggested that breast screening with MRI is still early in its development and that as radiologists gain experience and increase the number of breast MRIs they are reading, that the number of false-positives will substantially decrease (Robson 2004; Warner et al. 2004). Further research is required to determine whether this will be the case.

In conclusion, MRI alone or in combination with other surveillance modalities appears to be a promising for the surveillance of women at high risk of breast cancer. However, improved cancer detection does not necessarily translate to a decrease in mortality. More research with larger numbers of participants and longer follow-up is required to truly assess the performance of MRI, and combination strategies, for the surveillance of women at high risk of breast cancer. In addition to its accuracy, MRI has the advantage of not using ionising radiation. The drawbacks of MRI are primarily related to the potential harm of false-positive diagnoses, cost and availability. If the introduction of a surveillance strategy for women at high risk of breast cancer with MRI was to be contemplated, a more complete assessment would need to be carried out. This should include the potential benefit from surveillance versus the potential physical and psychological harm caused by the test, diagnostic procedures and treatment; the health care system being capable of supporting all the necessary elements of the surveillance pathway, including diagnosis, follow-up and evaluation; consideration of social and ethical issues and consideration of cost-benefit issues.

References

Banks, E., Reeves, G., Beral, V., Bull, D., Crossley, B., Simmonds, M., Hilton, E., et al. (2004). Influence of personal characteristics of individual women on sensitivity and specificity of mammography in the Million Women Study: Cohort study. *British Medical Journal*, 329, 477-479.

Bartella, L., & Morris, E. A. (2006). Advances in breast imaging: magnetic resonance imaging. *Current Oncology Reports*, 8, 7-13.

Berry, D. A., Parmigiani, G., Sanchez, J., Schildkraut, J., & Winer, E. (1997). Probability of carrying a mutation of breast-ovarian cancer gene BRCA1 based on family history. *Journal of the National Cancer Institute*, 89, 227-238.

Bland, K. I., Menck, H. R., Scott-Conner, C. E., Morrow, M., Winchester, D. J., & Winchester, D. P. (1998). The National Cancer Data Base 10-year survey of breast carcinoma treatment at hospitals in the United States. *Cancer*, 83, 1262-1273.

Brekelmans, C. T. M., Seynaeve, C., Bartels, C. C. M., Tilanus-Linthorst, M. M. A., Meijers-Heijboer, E. J., Crepin, C. M. G., van Geel, A. N., et al. (2001). Effectiveness of breast cancer surveillance in BRCA1/2 gene mutation carriers and women with high familial risk. *Journal of Clinical Oncology*, 19, 924-930.

Calderon-Margalit, R., & Paltiel, O. (2004). Prevention of breast cancer in women who carry BRCA1 or BRCA2 mutations: a critical review of the literature. *International Journal of Cancer*, 112, 357-364.

Chart, P. L., & Franssen, E. (1997). Management of women at increased risk for breast cancer: preliminary results from a new program. *CMAJ Canadian Medical Association Journal*, 157, 1235-1242.

Claus, E. B. (2001). Risk models used to counsel women for breast and ovarian cancer: a guide for clinicians. *Familial Cancer*, 1, 197-206.

Claus, E. B., Risch, N., & Thompson, W. D. (1994). Autosomal dominant inheritance of early-onset breast cancer. Implications for risk prediction. *Cancer*, 73, 643-651.

Crystal, P., Strano, S. D., Shcharynski, S., & Koretz, M. J. (2003). Using sonography to screen women with mammographically dense breasts. *AJR American Journal of Roentgenology*, 181, 177-182.

Daly, M. B., Axilbund, J. E., Bryant, E., Buys, S., Eng, C., Friedman, S., Esserman, L. J., et al. (2006). Genetic/familial high-risk assessment: breast and ovarian. *Journal of the National Comprehensive Cancer Network*, 4, 156-176.

Daly, P. A. (2004). Hereditary cancer: guidelines in clinical practice - General overview. *Annals of Oncology*, 15, iv121-iv125.

Deeks, J. (2001). Systematic reviews of evaluations of diagnostic and screening tests. In M. Egger, G. Davey Smith & D. G. Altman (Eds.), *Systematic reviews in health care: meta-analysis in context* (pp. 248-282). London: BMJ Books.

Duffy, S. W., Tabar, L., Vitak, B., & Warwick, J. (2006). Tumor size and breast cancer detection: what might be the effect of a less sensitive screening tool than mammography? *Breast Journal*, 12 Suppl 1, S91-95.

Eberl, M. M., Fox, C. H., Edge, S. B., Carter, C. A., & Mahoney, M. C. (2006). BI-RADS classification for management of abnormal mammograms. *Journal of the American Board of Family Medicine: JABFM*, 19, 161-164.

Elmore, J. G., Reisch, L. M., Barton, M. B., Barlow, W. E., Rolnick, S., Harris, E. L., Herrinton, L. J., et al. (2005). Efficacy of breast cancer screening in the community according to risk level. *Journal of the National Cancer Institute*, 97, 1035-1043.

Federico, M., Maiorana, A., Mangone, L., Turchetti, D., Canossi, B., Cortesi, L., Romagnoli, R., et al. (1999). Identification of families with hereditary breast and ovarian cancer for clinical and mammographic surveillance: The Modena Study Group proposal. *Breast Cancer Research & Treatment*, 55, 213-221.

Fentiman, I. S., & D'Arrigo, C. (2004). Pathogenesis of breast carcinoma. *International Journal of Clinical Practice*, 58, 35-40.

Frank, T. S., & Critchfield, G. C. (2001). Identifying and managing hereditary risk of breast and ovarian cancer. *Clinics in Perinatology*, 28, 395-406.

Greene, F. L., Page, D. L., & Fleming, I. D. (2002). *AJCC cancer staging manual. American Joint Committee on Cancer*. New York: Springer.

Gui, G. P., Hogben, R. K., Walsh, G., A'Hern, R., & Eeles, R. (2001). The incidence of breast cancer from screening women according to predicted family history risk: does annual clinical examination add to mammography? *European Journal of Cancer*, 37, 1668-1673.

Gui, G. P., Kadayaprath, G., Darhouse, N., Self, J., Ward, A., A'Hern, R., & Eeles, R. (2006). Clinical outcome and service implications of screening women at increased breast cancer risk from a family history. *European Journal of Surgical Oncology* 32, 719-724. Epub ahead of print Jun 16.

Halapy, E., Chiarelli, A. M., Klar, N., & Knight, J. A. (2005). Accuracy of breast screening among women with and without a family history of breast and/or ovarian cancer. *Breast Cancer Research & Treatment*, 90, 299-305.

Halapy, E. E., Chiarelli, A. M., Klar, N., & Knight, J. (2004). Breast screening outcomes in women with and without a family history of breast and/or ovarian cancer. *Journal of Medical Screening*, 11, 32-38.

Hartman, A. R., Daniel, B. L., Kurian, A. W., Mills, M. A., Nowels, K. W., Dirbas, F. M., Kingham, K. E., et al. (2004). Breast magnetic resonance image screening and ductal lavage in women at high genetic risk for breast carcinoma. *Cancer*, 100, 479-489.

Heywang, S. H., Wolf, A., Pruss, E., Hilbertz, T., Eiermann, W., & Permanetter, W. (1989). MR imaging of the breast with Gd-DTPA: use and limitations. *Radiology*, 171, 95-103.

Hou, M. F., Chuang, H. Y., Ou-Yang, F., Wang, C. Y., Huang, C. L., Fan, H. M., Chuang, C. H., et al. (2002). Comparison of breast mammography, sonography and physical examination for screening women at high risk of breast cancer in Taiwan. *Ultrasound in Medicine & Biology*, 28, 415-420.

Hughes, K. S., Papa, M. Z., Whitney, T., & McLellan, R. (1999). Prophylactic mastectomy and inherited predisposition to breast carcinoma. *Cancer*, 86, 2502-2516.

Irwig, L., Houssami, N., Armstrong, B., & Glasziou, P. (2006). Evaluating new screening tests for breast cancer. *BMJ*, 332, 678-679.

Irwig, L., Houssami, N., & van Vliet, C. (2004). New technologies in screening for breast cancer: a systematic review of their accuracy. *British Journal of Cancer*, 90, 2118-2122.

Kaiser, W. A., & Zeitler, E. (1989). MR imaging of the breast: fast imaging sequences with and without Gd-DTPA. Preliminary observations. *Radiology*, 170, 681-686.

Kerlikowske, K., Carney, P. A., Geller, B., Mandelson, M. T., Taplin, S. H., Malvin, K., Ernster, V., et al. (2000). Performance of screening mammography among women with and without a first-degree relative with breast cancer. *Annals of Internal Medicine*, 133, 855-863.

Kolb, T. M., Lichy, J., & Newhouse, J. H. (1998). Occult cancer in women with dense breasts: detection with screening US--diagnostic yield and tumor characteristics. *Radiology*, 207, 191-199.

Kollias, J., Sibbering, D. M., Blamey, R. W., Holland, P. A., Obuszko, Z., Wilson, A. R., Evans, A. J., et al. (1998). Screening women aged less than 50 years with a family history of breast cancer. *European Journal of Cancer*, 34, 878-883.

Kriege, M., Brekelmans, C. T., Boetes, C., Besnard, P. E., Zonderland, H. M., Obdeijn, I. M., Manoliu, R. A., et al. (2004). Efficacy of MRI and mammography for breast-cancer screening in women with a familial or genetic predisposition. *New England Journal of Medicine*, 351, 427-437.

Kriege, M., Brekelmans, C. T. M., Boetes, C., Besnard, P. E., Zonderland, H. M., Obdeijn, I. M., Manoliu, R. A., et al. (2005). Efficacy of magnetic resonance imaging and mammography for breast cancer screening in women with a familial or genetic predisposition. *Obstetrical & Gynecological Survey*, 60, 107-109.

Kuhl, C. K., Kuhn, W., & Schild, H. (2005a). Management of women at high risk for breast cancer: new imaging beyond mammography. *Breast*, 14, 480-486.

Kuhl, C. K., Schmutzler, R. K., Leutner, C. C., Kempe, A., Wardelmann, E., Hocke, A., Maringa, M., et al. (2000). Breast MR imaging screening in 192 women proved or suspected to be carriers of a breast cancer susceptibility gene: preliminary results. *Radiology*, 215, 267-279.

Kuhl, C. K., Schrading, S., Leutner, C. C., Morakkabati-Spitz, N., Wardelmann, E., Fimmers, R., Kuhn, W., et al. (2005b). Mammography, breast ultrasound, and magnetic resonance imaging for surveillance of women at high familial risk for breast cancer. *Journal of Clinical Oncology*, 23, 8469-8476.

Kuschel, B., Hauenstein, E., Kiechle, M., & Meindl, A. (2006). Hereditary breast and ovarian cancer - Current clinical guidelines in Germany. *Breast Care*, 1, 8-14.

Lalloo, F., Boggis, C. R., Evans, D. G., Shenton, A., Threlfall, A. G., & Howell, A. (1998). Screening by mammography, women with a family history of breast cancer. *European Journal of Cancer*, 34, 937-940.

Leach, M. O., Boggis, C. R., Dixon, A. K., Easton, D. F., Eeles, R. A., Evans, D. G., Gilbert, F. J., et al. (2005). Screening with magnetic resonance imaging and mammography of a UK population at high familial risk of breast cancer: a prospective multicentre cohort study (MARIBS). [erratum appears in Lancet. 2005 May 28-Jun 3;365(9474):1848]. *Lancet*, 365, 1769-1778.

Lehman, C. D., Blume, J. D., Weatherall, P., Thickman, D., Hylton, N., Warner, E., Pisano, E., et al. (2005). Screening women at high risk for breast cancer with mammography and magnetic resonance imaging. *Cancer*, 103, 1898-1905.

Lucassen, A., Watson, E., & Eccles, D. (2001). Evidence based case report: advice about mammography for a young woman with a family history of breast cancer. *British Medical Journal*, 322, 1040-1042.

Macmillan, R. D. (2000). Screening women with a family history of breast cancer--results from the British Familial Breast Cancer Group. *European Journal of Surgical Oncology*, 26, 149-152.

Mark, D. (2003). Magnetic resonance imaging of the breast in screening women considered to be at high genetic risk of breast cancer. *Blue Cross Blue Shield TEC Assessment Program*, 18, 1-14.

Markham, B. A., Hutchison, B., Birch, S., Goldsmith, L. J., & Evans, C. E. (1997). Casting the screening net: separating big fish from little fish. *Health Policy*, 42, 171-184.

Maurice, A., Evans, D. G., Shenton, A., Ashcroft, L., Baildam, A., Barr, L., Byrne, G., et al. (2006). Screening younger women with a family history of breast cancer - does early detection improve outcome? *European Journal of Cancer*, 42, 1385-1390.

McIntosh, A., Shaw, C., Evans, G., Turnbull, N., Bahar, N., Barclay, M., Easton, D., et al. (2004). *Clinical guidelines and evidence review for the classification and care of women at risk of familial breast cancer*. London: National Collaborating Centre for Primary Care/University of Sheffield.

Moller, P., Reis, M. M., Evans, G., Vasen, H., Haites, N., Anderson, E., Steel, C. M., et al. (1999). Efficacy of early diagnosis and treatment in women with a family history of breast cancer. European Familial Breast Cancer Collaborative Group. *Disease Markers*, 15, 179-186.

Morris, E. A., Liberman, L., Ballon, D. J., Robson, M., Abramson, A. F., Heerdt, A., & Dershaw, D. D. (2003). MRI of occult breast carcinoma in a high-risk population. *AJR American Journal of Roentgenology*, 181, 619-626.

Moss, S., Waller, M., Anderson, T. J., Cuckle, H., & Trial Management Group (2005). Randomised controlled trial of mammographic screening in women from age 40: predicted mortality based on surrogate outcome measures. *British Journal of Cancer*, 92, 955-960.

Murday, V., Pears, R., Ball, J., Eeles, R., & Hodgson, S. (2004). An audit of screening for familial breast cancer before 50 years in the South Thames Region - have we got it right? *Familial Cancer*, 3, 29-34.

Myles, J., Duffy, S., Nixon, R., Boggis, C., Howell, A., Shenton, A., & Evans, G. (2001). Initial results of a study into the effectiveness of breast cancer screening in a population identified to be at high risk. *Revue d Epidemiologie et de Sante Publique*, 49, 471-475.

National Collaborating Centre for Primary Care (2006). Familial breast cancer. The classification and care of women at risk of familial breast cancer in primary, secondary and tertiary care. Update. London: National Institute for Clinical Excellence. Available from: http://www.nice.org.uk/guidance/cg41/guidance/pdf/English Accessed on 24.11.06

National Health and Medical Research Council (2000). *How to use the evidence: assessment and application of scientific evidence*. Canberra: NHMRC.

National Health Committee (2003). *Screening to improve health in New Zealand: criteria to assess screening programmes*. Wellington: National Health Committee.

New Zealand Guidelines Group (2004). *Surveillance and management of groups at increased risk of colorectal cancer*. Wellington: New Zealand Guidelines Group.

New Zealand Health Information Service (2006a). Cancer: new registrations and deaths 2002. Wellington: NZHIS. Available from: http://www.nzhis.govt.nz/publications/cancer02.pdf Accessed on 22.11.06

New Zealand Health Information Service (2006b). Selected morbidity data for publicly funded hospitals 1 July 2002 to 30 June 2003. Wellington: NZHIS. Available from: http://www.nzhis.govt.nz/publications/morbidity02-03.pdf Accessed on 22.11.06

Nixon, R. M., Pharoah, P., Tabar, L., Krusemo, U. B., Duffy, S. W., Prevost, T. C., & Chen, H. H. (2000). Mammographic screening in women with a family history of breast cancer: some results from the Swedish two-county trial. *Revue d Epidemiologie et de Sante Publique*, 48, 325-331.

O'Driscoll, D., Warren, R., MacKay, J., Britton, P., & Day, N. E. (2001). Screening with breast ultrasound in a population at moderate risk due to family history. *Journal of Medical Screening*, 8, 106-109.

Pharoah, P. D. P., Stratton, J. F., & Mackay, J. (1998). Screening for breast and ovarian cancer: the relevance of family history. *British Medical Bulletin*, 54, 823-838.

Pichert, G., Bolliger, B., Buser, K., & Pagani, O. (2003). Evidence-based management options for women at increased breast/ovarian cancer risk. *Annals of Oncology*, 14, 9-19.

Podo, F., Sardanelli, F., Canese, R., D'Agnolo, G., Natali, P. G., Crecco, M., Grandinetti, M. L., et al. (2002). The Italian multi-centre project on evaluation of MRI and other imaging modalities in early detection of breast cancer in subjects at high genetic risk. *Journal of Experimental & Clinical Cancer Research*, 21, 115-124.

Rijnsburger, R. (2005). *Effects and costs of breast cancer screening in women with a familial or genetic predisposition.* Thesis. Rotterdam: Department of Public Health, Erasmus MC, University Medical Centre.

Rizzatto, G. J. (2001). Towards a more sophisticated use of breast ultrasound. *European Radiology*, 11, 2425-2435.

Robson, M. (2004). Breast cancer surveillance in women with hereditary risk due to BRCA1 or BRCA2 mutations. *Clinical Breast Cancer*, 5, 260-268; discussion 269-271.

Robson, M. E. (2002). Clinical considerations in the management of individuals at risk for hereditary breast and ovarian cancer. *Cancer Control*, 9, 457-465.

Scheuer, L., Kauff, N., Robson, M., Kelly, B., Barakat, R., Satagopan, J., Ellis, N., et al. (2002). Outcome of preventive surgery and screening for breast and ovarian cancer in BRCA mutation carriers. *Journal of Clinical Oncology*, 20, 1260-1268.

Sim, L. S., Hendriks, J. H., & Fook-Chong, S. M. (2004). Breast ultrasound in women with familial risk of breast cancer. *Annals of the Academy of Medicine, Singapore*, 33, 600-606.

Singletary, S. E., & Connolly, J. L. (2006). Breast cancer staging: working with the sixth edition of the AJCC Cancer Staging Manual. *CA: a Cancer Journal for Clinicians*, 56, 37-47; quiz 50-31.

Smith, J. A., & Andreopoulou, E. (2004). An overview of the status of imaging screening technology for breast cancer. *Annals of Oncology*, 15, i18-i26.

Spratt, J. S., Meyer, J. S., & Spratt, J. A. (1995). Rates of growth of human solid neoplasms: Part I. *Journal of Surgical Oncology*, 60, 137-146.

Spratt, J. S., Meyer, J. S., & Spratt, J. A. (1996). Rates of growth of human neoplasms: Part II. *Journal of Surgical Oncology*, 61, 68-83.

Statistics New Zealand (2002). 2001 census: national summary (2001) - reference report. Table 1-population summary by sex, for the census night and census usually resident population counts, 1926-2001. Wellington: Statistics New Zealand. Available from: http://www.stats.govt.nz/census/2001-census-statistics/2001-national-summary/default.htm Accessed on 22.11.06

Stoutjesdijk, M. J., Boetes, C., Jager, G. J., Beex, L., Bult, P., Hendriks, J. H., Laheij, R. J., et al. (2001). Magnetic resonance imaging and mammography in women with a hereditary risk of breast cancer. *Journal of the National Cancer Institute*, 93, 1095-1102.

Tabar, L., Vitak, B., Chen, H. H., Duffy, S. W., Yen, M. F., Chiang, C. F., Krusemo, U. B., et al. (2000). The Swedish Two-County Trial twenty years later. Updated mortality results and new insights from long-term follow-up. *Radiologic Clinics of North America*, 38, 625-651.

Thomas, D. B., Gao, D. L., Ray, R. M., Wang, W. W., Allison, C. J., Chen, F. L., Porter, P., et al. (2002). Randomized trial of breast self-examination in Shanghai: final results. *Journal of the National Cancer Institute*, 94, 1445-1457.

Tilanus-Linthorst, M., Verhoog, L., Obdeijn, I. M., Bartels, K., Menke-Pluymers, M., Eggermont, A., Klijn, J., et al. (2002). A BRCA1/2 mutation, high breast density and prominent pushing margins of a tumor independently contribute to a frequent false-negative mammography.[erratum appears in Int J Cancer 2002 Dec 20;102(6):665]. *International Journal of Cancer*, 102, 91-95.

Trecate, G., Vergnaghi, D., Bergonzi, S., De Simone, T., Fengoni, E., Costa, C., Spatti, G., et al. (2003). Breast MRI screening in patients with increased familial and/or genetic risk for breast cancer: A preliminary experience. *Tumori*, 89, 125-131.

Warner, E., Plewes, D. B., Hill, K. A., Causer, P. A., Zubovits, J. T., Jong, R. A., Cutrara, M. R., et al. (2004). Surveillance of BRCA1 and BRCA2 mutation carriers with magnetic resonance imaging, ultrasound, mammography, and clinical breast examination. *JAMA*, 292, 1317-1325.

Warner, E., Plewes, D. B., Shumak, R. S., Catzavelos, G. C., Di Prospero, L. S., Yaffe, M. J., Goel, V., et al. (2001). Comparison of breast magnetic resonance imaging, mammography, and ultrasound for surveillance of women at high risk for hereditary breast cancer. *Journal of Clinical Oncology*, 19, 3524-3531.

Appendix 1: Search strategies

SEARCH STRATEGIES

Medline

1 breast neoplasms/ (64688)
2 (breast adj (cancer or neoplas$ or carcino$ or malignan$ or adenocarcino$)).tw. (56204)
3 1 or 2 (74919)
4 (screen$ or surveillance).mp. (194488)
5 exp Mammography/ (7999)
6 exp Ultrasonography/ (72153)
7 exp Magnetic Resonance Imaging/ (111737)
8 (breast adj3 examination).tw. (1014)
9 mass screening/ (25873)
10 mri.tw. (43525)
11 or/4-10 (378723)
12 exp Genetic Predisposition to Disease/ (27536)
13 familial.mp. (27339)
14 Genes, BRCA1/ (2370)
15 ((high or increas$) adj2 risk$).tw. (106958)
16 ((high or increas$) adj2 rate).tw. (38498)
17 ((high or increas$) adj2 incidence).tw. (22926)
18 ((extra or heighten$) adj (risk$ or rate or incidence)).tw. (352)
19 (family adj2 history).tw. (13750)
20 first degree relative$.tw. (3118)
21 family/ (14771)
22 family health/ (9380)
23 ge.fs. (886620)
24 Genes, BRCA2/ (964)
25 (brca1 or brca2).tw. (3960)
26 or/12-25 (1056047)
27 3 and 11 and 26 (3905)
28 limit 27 to english (3638)
29 (letter or news).pt. (319885)
30 28 not 29 (3483)
31 randomized controlled trial.pt. (122201)
32 meta-analysis.pt. (10826)
33 randomized controlled trials/ or meta-analysis/ (38551)
34 controlled clinical trials/ or controlled clinical trial.pt. (28785)
35 exp clinical trials/ or clinical trial.pt. (307900)
36 random allocation/ or (random$ adj2 allocat$).tw. (26564)
37 single blind method/ or double blind method/ (49315)
38 (clinic$ adj trial$).tw. (62969)
39 ((singl$ or doubl$ or trebl$ or tripl$) adj (blind$ or mask$ or dumm$)).tw. (39592)
40 (systematic$ adj3 (review$ or overview)).tw. (9987)
41 (meta-analy$ or metaanaly$).tw. (12831)
42 exp review literature/ (2381)
43 (hand search$ or relevant journals or manual search$ or selection criteria or data extraction).ab. (10222)
44 or/31-43 (391968)
45 letter.pt. (254413)
46 case report.tw. (52552)
47 (historical article or review of reported cases or review, multicase).pt. (62607)
48 or/45-47 (365602)

49 animal/ (1362984)
50 human/ (3622684)
51 49 not (49 and 50) (917974)
52 44 not (48 or 51) (360257)
53 exp epidemiologic studies/ (531360)
54 exp case control studies/ (216952)
55 exp cohort studies/ (316769)
56 cross-sectional studies/ (47919)
57 (case control or cohort analy$ or cross sectional).tw. (65979)
58 (longitudinal or retrospective).tw. (115298)
59 (cohort adj (study or studies)).tw. (21249)
60 ((follow-up or observational) adj (study or studies)).tw. (20548)
61 or/53-60 (584587)
62 30 and (52 or 61) (1050)
63 "sensitivity and specificity"/ (122231)
64 predictive value of tests/ (50193)
65 (false positive or false negative).mp. (19196)
66 (positive predictive value or ppv or negative predictive value or npv).mp. (11689)
67 interval cancer$.mp. (183)
68 diagnostic accuracy.mp. (6393)
69 likelihood function$.mp. (6996)
70 Comparative Study/ (546517)
71 or/63-70 (678899)
72 30 and 71 (769)
73 72 or 62 (1509)
74 (ovarian or prostate).ti. (40908)
75 breast.ti. (51200)
76 74 not 75 (39761)
77 73 not 76 (1485)

Embase

1 exp Breast Cancer/ (75627)
2 (breast adj (cancer or neoplas$ or carcino$ or malignan$ or adenocarcino$)).mp. (80787)
3 1 or 2 (81902)
4 cancer screening/ (16850)
5 (screen$ or surveillance).tw. (154172)
6 exp mammography/ (11067)
7 exp breast examination/ (13162)
8 echography/ (46401)
9 (ultrasound$ or ultrasonography).mp. (79429)
10 exp nuclear magnetic resonance imaging/ (130521)
11 mri.tw. (45667)
12 (breast adj3 examination).tw. (1038)
13 or/4-12 (380959)
14 3 and 13 (16416)
15 exp genetic predisposition/ or exp genetic susceptibility/ (26321)
16 Familial Cancer/ (3903)
17 brca1 protein/ or brca2 protein/ (3278)
18 (bcra1 or bcra2).tw. (11)
19 family/ (18011)
20 familial.tw. (25555)
21 first degree relative$.tw. (3102)
22 (family adj2 history).tw. (14506)
23 ((high or increas$ or extra or heighten$) adj2 (risk$ or rate or incidence)).tw. (186824)
24 GENETICS/ or CANCER GENETICS/ (12204)
25 or/15-24 (269721)
26 14 and 25 (3102)
27 "SENSITIVITY AND SPECIFICITY"/ (27571)
28 (positive predictive value or negative predictive value or ppv or npv).tw. (11566)
29 diagnostic accuracy/ or diagnostic error/ or diagnostic value/ (131181)

30 (false positive or false negative).mp. (13867)
31 likelihood ratio$.mp. (2382)
32 interval cancer$.tw. (214)
33 comparative study/ or intermethod comparison/ (135027)
34 or/27-33 (275824)
35 26 and 34 (490)
36 limit 35 to english (433)
37 letter.pt. (206526)
38 36 not 37 (428)
39 clinical trial/ (317135)
40 randomized controlled trial/ (93875)
41 randomization/ (17620)
42 single blind procedure/ or double blind procedure/ (46076)
43 crossover procedure/ (13679)
44 placebo/ (49736)
45 (randomized controlled trial$ or randomised controlled trial$).tw. (18071)
46 rct.tw. (1323)
47 (random$ adj2 allocat$).tw. (6432)
48 ((singl$ or doubl$ or tripl$ or trebl$) adj (blind$ or mask$ or dummy)).tw. (41526)
49 prospective study/ (49320)
50 case study/ (2174)
51 case report.tw. (55290)
52 abstract report/ or letter/ (206921)
53 or/50-52 (263204)
54 or/39-49 (385078)
55 54 not 53 (372572)
56 exp meta-analysis/ (23022)
57 (meta-analy$ or metaanaly$).tw. (12248)
58 (systematic$ adj3 (review$ or overview)).tw. (9676)
59 (reference list$ or manual search$ or hand search$ or relevant journals or bibliograph$).tw. (5687)
60 (data extraction or selection criteria or medline or embase or cinahl or psychlit or psychinfo).ab. (17431)
61 review.pt. (524122)
62 or/56-61 (549524)
63 (letter or editorial).pt. (331063)
64 animal/ (884)
65 human/ (2931740)
66 64 not (64 and 65) (501)
67 63 or 66 (331537)
68 62 not 67 (547133)
69 clinical study/ (6869)
70 case control study/ (11685)
71 family study/ (4060)
72 longitudinal study/ (11331)
73 retrospective study/ (61125)
74 prospective study/ (49320)
75 cohort analysis/ (32341)
76 (cohort adj (study or studies)).mp. (20958)
77 (case control adj (study or studies)).mp. (22009)
78 (observational adj (study or studies)).tw. (10599)
79 (epidemiologic$ adj (study or studies)).tw. (18442)
80 (follow-up adj (study or studies)).tw. (10738)
81 (cross sectional adj (study or studies)).tw. (14396)
82 or/69-81 (221914)
83 26 and (55 or 68 or 82) (1374)
84 83 or 38 (1594)
85 (ovarian or prostate).ti. (40358)
86 breast.ti. (49436)
87 85 not (85 and 86) (39240)

88 84 not 87 (1548)

Cochrane Register of Controlled Trials

1 breast neoplasms/ (4258)
2 (breast adj (cancer or neoplas$ or carcino$ or malignan$ or adenocarcino$)).tw. (7499)
3 1 or 2 (8127)
4 (screen$ or surveillance).mp. (8897)
5 exp Mammography/ (401)
6 exp Ultrasonography/ (3706)
7 exp Magnetic Resonance Imaging/ (1983)
8 (breast adj3 examination).tw. (197)
9 mass screening/ (1184)
10 mri.tw. (1354)
11 or/4-10 (14972)
12 exp Genetic Predisposition to Disease/ (258)
13 familial.mp. (967)
14 Genes, BRCA1/ (18)
15 ((high or increas$) adj2 risk$).tw. (10649)
16 ((high or increas$) adj2 rate).tw. (7479)
17 ((high or increas$) adj2 incidence).tw. (2634)
18 ((extra or heighten$) adj (risk$ or rate or incidence)).tw. (30)
19 (family adj2 history).tw. (657)
20 first degree relative$.tw. (176)
21 family/ (480)
22 family health/ (196)
23 ge.fs. (4357)
24 Genes, BRCA2/ (10)
25 (brca1 or brca2).tw. (34)
26 or/12-25 (25598)
27 3 and 11 and 26 (134)
28 limit 27 to english [Limit not valid; records were retained] (134)
29 (letter or news).pt. (4321)
30 28 not 29 (133)
31 limit 30 to yr=1996-2006 (103)

Current Contents

1. (breast cancer or breast carcino* or breast malignan* or breast adnocarcino* or breast neoplas*)
2. (Mammography or ultrasound or ultrasonograpy or magnetic resonance imaging or mri)
3. (Surveillance or screen*)
4. (breast SAME examination)
5. #1 and (#2 or #3 or #4)
6. ((high or heighten* or increas* or extra) SAME (risk* or rate or incidence))
7. (family history or genetic predisposition or brca1 or brca2)
8. (first degree relative* or familial or genetic)
9. #6 or #7 or #8
10. #5 and #9
11. (positive predictive value or negative predictive value or ppv or npv)
12. (false positive or false negative or diagnostic accuracy)
13. (interval cancer$ or likelihood ratio$ or comparative study)
14. (control* SAME (trial or study))
15. ((systematic* SAME review) or metaanaly* or meta-analy*)
16. ((cohort or longitudinal or follow-up) SAME study)
17. (clinical SAME (study or trial))
18. #11 or #12 or #13 or #14 or #15 or #16 or #17
19. #10 and #18

SEARCHES FROM OTHER SOURCES

In databases and all other sources without controlled vocabulary combinations of the index terms and additional keywords from the above strategies, were used in the search.

Appendix 2: Sources searched

SOURCES SEARCHED

Bibliographic databases

Current Contents
Cochrane Central Register of Controlled Trials
Embase
Medline
PubMed (last 60 days only)

Review databases

Cochrane Database of Systematic Reviews
Database of Abstracts of Reviews of Effects (DARE)
Health Technology Assessment Database
NHS Economic Evaluation Database

Guidelines

National Institute for Clinical Excellence (NICE) http://www.nice.org.uk/
US Guidelines Clearing House http://www.guidelines.gov
UK National Library for Health Guidelines Finder http://www.library.nhs.uk/guidelinesfinder/
TRIP database http://www.tripdatabase.com/

Other

National Cancer Institute (US) http://www.cancer.gov/
Statistics New Zealand http://www.stats.govt.nz
New Zealand Health Information Service http://www.nzhis.govt.nz
National Breast Cancer Centre (Australia) http://www.nbcc.org.au/

Reports supplied by the Dutch Health Insurance Board (College voor Zorgverzekeringen) http://www.cvz.nl *(see Acknowledgement section)*
References of retrieved papers

Appendix 3

METHODOLOGY FOR CALCULATING DIAGNOSTIC TEST PERFORMANCE

The diagnostic test performance includes consideration of validity and reliability of the test. Specifically, sensitivity, specificity, and positive and negative predictive values were calculated when possible to assess the validity of each screening test. These measures were calculated based on presentation of results as shown in **Table 30.**

Table 30. Assessment of validity of a diagnostic test

		Reference test or true disease state		
		Positive	**Negative**	**Total**
Diagnostic test	Positive	a	b	n_p
	Negative	c	d	n_n
Total sample size		n_1	n_2	

Based on **Table 30** above, measures of validity were calculated using the following formulae:

Sensitivity = a/(a+c)
= a/n_1

Specificity = d/(b+d)
= d/n_2

Positive predictive value (PPV) = a/(a+b)
= a/n_p

Negative predictive value (NPV) = d/(c+d)
= d/n_n

Appendix 4

NHMRC LEVELS OF EVIDENCE

The strength of evidence presented in the selected studies was assessed and classified using the dimensions of evidence defined by the National Health and Medical Research Council (NHMRC, 2000). A modified description of the designations of the levels of evidence is shown below.

Level of evidence	Study design
I	Evidence obtained from a systematic review of all relevant randomised controlled trials
II	Evidence obtained from at least one properly-designed randomised controlled trial
III-1	Evidence obtained from well-designed pseudorandomised controlled trials (alternate allocation or some other method)
III-2	Evidence obtained from comparative studies (including systematic reviews of such studies) with concurrent controls and allocation not randomised, cohort studies, case-control studies, or interrupted time series with a control group
III-3	Evidence obtained from comparative studies with historical control, two or more single arm studies, or interrupted time series without a parallel control group
IV	Evidence obtained from case-series, either post-test or pre-test/post-test

Appendix 5

RETRIEVED STUDIES EXCLUDED FOR REVIEW

Anonymous (2005). Use of magnetic resonance imaging in breast oncology. *Journal of the American College of Surgeons*, 200, 742.

Antill, Y. C., Shanahan, M., & Phillips, K. A. (2005). The integrated, multidisciplinary clinic: A new model for the ongoing management of women at high genetic risk for breast and ovarian cancer. *Cancer Forum*, 29, 107-110.

Bartella, L., Liberman, L., Morris, E. A., & Dershaw, D. D. (2006). Nonpalpable mammographically occult invasive breast cancers detected by MRI. *AJR American Journal of Roentgenology*, 186, 865-870.

Bartella, L., & Morris, E. A. (2006). Advances in breast imaging: magnetic resonance imaging. *Current Oncology Reports*, 8, 7-13.

Basil, J. B., & Rader, J. S. (2000). Management of women at risk for malignancy. *Current Opinion in Oncology*, 12, 508-513.

Beattie, M., & Kerlikowske, K. (2005). The addition of MRI greatly improved the sensitivity of screening for early breast cancer in mutation carriers. *Evidence-Based Obstetrics & Gynecology*, 7, 100-102.

Beck, J. D., & La Fargue, M. M. (2004). Breast cancer risk assessment: identifying and managing high risk patients. *Seminars in Breast Disease*, 7, 153-158.

Beckmann, M. W., Schnurch, H. G., Bodden-Heidrich, R., Mosny, D. S., Crombach, G., Nitz, U., Achnoula, M., et al. (1996). Early cancer detection programmes for women at high risk for breast and ovarian cancer: a proposal of practical guidelines. *European Journal of Cancer Prevention*, 5, 468-475.

Benichou, J. (2001). Development of a protocol for evaluation of mammographic surveillance services in women under 50 with a family history of breast cancer by J Mackay, C Rogers, H Fielder, R Blamey, D Macmillan, C Boggis, J Brown, PD Pharoah, S Moss, NE Day, J Myles, J Austoker, J Gray, J Cuzick and SW Duffy. *Journal of Epidemiology & Biostatistics*, 6, 373-374.

Berg, W. (2003). Randomized screening study of breast ultrasonography and mammography in women at high risk for breast cancer. *PDQ, ACRIN-6666*, 6666.

Berrington de Gonzalez, A., & Reeves, G. (2005). Mammographic screening before age 50 years in the UK: comparison of the radiation risks with the mortality benefits. *British Journal of Cancer*, 93, 590-596.

Bisanti, L., & Russo, A. (2001). Development of a protocol for evaluation of mammographic surveillance services in women under 50 with a family history of breast cancer by J Mackay, C Rogers, H Fielder, R Blamey, D Macmillan, C Boggis, J Brown, PD Pharoah, S Moss, NE Day, J Myles, J Austoker, J Gray, J Cuzick and SW Duffy. *Journal of Epidemiology & Biostatistics*, 6, 371-372.

Boetes, C. (2002). The evaluation of women with familial risk of breast cancer. *Journal of Experimental & Clinical Cancer Research*, 21, 97-101.

Boetes, C., & Stoutjesdijk, M. (2001). MR imaging in screening women at increased risk for breast cancer. *Magnetic Resonance Imaging Clinics of North America*, 9, 357-372.

Brekelmans, C. T., Bartels, C. C., Crepin, E., van Geel, A. N., Meijers-Heijboer, H., Seynaeve, C., Tilanus-Linthorst, M. M., et al. (1999). Breast cancer screening in high-risk women. Rotterdam Committee of Medical and Genetic Counseling. *Disease Markers*, 15, 34-36.

Brown, J., Buckley, D., Coulthard, A., Dixon, A. K., Dixon, J. M., Easton, D. F., Eeles, R. A., et al. (2000a). Magnetic resonance imaging screening in women at genetic risk of breast cancer: imaging and analysis protocol for the UK multicentre study. UK MRI Breast Screening Study Advisory Group. *Magnetic Resonance Imaging*, 18, 765-776.

Brown, J., Coulthard, A., Dixon, A. K., Dixon, J. M., Easton, D. F., Eeles, R. A., Evans, D. G. R., et al. (2000b). Protocol for a national multi-centre study of magnetic resonance imaging screening in women at genetic risk of breast cancer. *Breast*, 9, 78-82.

Chang-Claude, J. (2001). Inherited genetic susceptibility to breast cancer. *IARC Scientific Publications*, 154, 177-190.

Chang, J., & Elledge, R. M. (2001). Clinical management of women with genomic BRCA1 and BRCA2 mutations. *Breast Cancer Research & Treatment*, 69, 101-113.

Chiarelli, A. M., Halapy, E., Nadalin, V., Shumak, R., O'Malley, F., & Mai, V. (2006). Performance measures from 10 years of breast screening in the Ontario Breast Screening Program, 1990/91 to 2000. *European Journal of Cancer Prevention*, 15, 34-42.

Choi, S. (2005). Should women at high risk of breast cancer be screened with MRI? *CMAJ: Canadian Medical Association Journal*, 173, 32.

Daly, M. B., Axilbund, J. E., Bryant, E., Buys, S., Eng, C., Friedman, S., Esserman, L. J., et al. (2006). Genetic/familial high-risk assessment: breast and ovarian. *Journal of the National Comprehensive Cancer Network*, 4, 156-176.

Daly, P. A. (2004). Hereditary cancer: Guidelines in clinical practice - general overview. *Annals of Oncology*, 15, iv121-iv125.

DeMichele, A., & Weber, B. L. (2002). Risk management in BRCA1 and BRCA2 mutation carriers: Lessons learned, challenges posed. *Journal of Clinical Oncology*, 20, 1164-1166.

Domchek, S. M., Armstrong, K., & Weber, B. L. (2006). Clinical management of BRCA1 and BRCA2 mutation carriers. *Nature Clinical Practice Oncology*, 3, 2-3.

Eeles, R. A. (2000). Future possibilities in the prevention of breast cancer. Intervention strategies in BRCA1 and BRCA2 mutation carriers. *Breast Cancer Research*, 2, 283-290.

Elmore, J. G., Armstrong, K., Lehman, C. D., & Fletcher, S. W. (2005). Screening for breast cancer. *JAMA*, 293, 1245-1256.

Elwood, J. M. (1999). Public health aspects of breast cancer gene testing in Canada. Part 1: risks and interventions. *Chronic Diseases in Canada*, 20, 3-13.

Emery, J., Lucassen, A., & Murphy, M. (2001). Common hereditary cancers and implications for primary care. *Lancet*, 358, 56-63.

Evans, D. G. R., & Lalloo, F. (2002). Risk assessment and management of high risk familial breast cancer. *Journal of Medical Genetics*, 39, 865-871.

Evans, D. G. R., Lalloo, F., Shenton, A., Boggis, C., & Howell, A. (2001). Uptake of screening and prevention in women at very high risk of breast cancer. *Lancet*, 358, 889-890.

Frank, T. S., & Critchfield, G. C. (2001). Identifying and managing hereditary risk of breast and ovarian cancer. *Clinics in Perinatology*, 28, 395-406.

Friedenson, B. (2000). Is mammography indicated for women with defective BRCA genes? Implications of recent scientific advances for the diagnosis, treatment, and prevention of hereditary breast cancer. *Medscape General Medicine*, 2, E9.

Fries, M. H., Hailey, B. J., Flanagan, J., & Licklider, D. (2004). Outcome of five years of accelerated surveillance in patients at high risk for inherited breast/ovarian cancer: report of a phase II trial. *Military Medicine*, 169, 411-416.

Fries, M. H., Holt, C., Carpenter, I., Carter, C. L., Daniels, J., Flanagan, J., Murphy, K., et al. (2002). Guidelines for evaluation of patients at risk for inherited breast and ovarian cancer: Recommendations of the Department of Defense Familial Breast/Ovarian Cancer Research Project. *Military Medicine*, 167, 93-98.

Garber, J. E. (2003). Breast cancer screening: a final analysis? *CA: a Cancer Journal for Clinicians*, 53, 138-140.

Gilbert, F. J. (1999). New screening techniques for breast cancer (MRI). *Disease Markers*, 15, 115-116.

Gilbert, F. J. (2005). Screening for breast cancer in women at moderate and high risk. *Clinical Oncology (Royal College of Radiologists)*, 17, 240-243.

Gilhuijs, K. G., Deurloo, E. E., Muller, S. H., Peterse, J. L., & Schultze Kool, L. J. (2002). Breast MR imaging in women at increased lifetime risk of breast cancer: clinical system for computerized assessment of breast lesions initial results. *Radiology*, 225, 907-916.

Goodwin, P. J. (2000). Management of familial breast cancer risk. *Breast Cancer Research & Treatment*, 62, 19-33.

Gordon, P. B. (2002). Ultrasound for breast cancer screening and staging. *Radiologic Clinics of North America*, 40, 431-441.

Goscin, C. P., Berman, C. G., & Clark, R. A. (2001). Magnetic resonance imaging of the breast. *Cancer Control*, 8, 399-406.

Grann, V. R., Jacobson, J. S., Thomason, D., Hershman, D., Heitjan, D. F., & Neugut, A. I. (2002). Effect of prevention strategies on survival and quality-adjusted survival of women with BRCA1/2 mutations: an updated decision analysis. *Journal of Clinical Oncology*, 20, 2520-2529.

Gregory, H., Milner, B., Haites, N. E., Caligo, M., Chang-Claude, J., Eccles, D. M., Evans, D. G., et al. (2000). Current breast screening services available throughout Europe to families at increased genetic risk for breast cancer. *CME Journal of Gynecologic Oncology*, 5, 248-253.

Griffith, G. L., Edwards, R. T., Gray, J., Wilkinson, C., Turner, J., France, B., & Bennett, P. (2004). Estimating the survival benefits gained from providing national cancer genetic services to women with a family history of breast cancer. *British Journal of Cancer*, 90, 1912-1919.

Hamann, U. (2000). Hereditary breast cancer: high risk genes, genetic testing and clinical implications. *Clinical Laboratory*, 46, 447-461.

Harms, S. E. (1998). Breast magnetic resonance imaging. *Seminars in Ultrasound CT & MRI*, 19, 104-120.

Hartman, A. R. (2005). The problems with risk selection; scientific and psychosocial aspects. *Recent Results in Cancer Research*, 166, 125-144.

Heywang-Kobrunner, S. H., Viehweg, P., Heinig, A., & Kuchler, C. (1997). Contrast-enhanced MRI of the breast: accuracy, value, controversies, solutions. *European Journal of Radiology*, 24, 94-108.

Hollingsworth, A. B., Singletary, S. E., Morrow, M., Francescatti, D. S., O'Shaughnessy, J. A., Hartman, A. R., Haddad, B., et al. (2004). Current comprehensive assessment and management of women at increased risk for breast cancer. *American Journal of Surgery*, 187, 349-362.

Hughes, K. S., Papa, M. Z., Whitney, T., & McLellan, R. (1999). Prophylactic mastectomy and inherited predisposition to breast carcinoma. *Cancer*, 86, 2502-2516.

Hughes, K. S., Roche, C. A., Whitney, T., & McLellan, R. (2000). The management of women at high risk of experiencing hereditary breast and ovarian cancer: the Lahey guidelines. *Disease Management & Health Outcomes*, 7, 201-215.

Jatoi, I., & Anderson, W. F. (2005). Cancer screening. *Current Problems in Surgery*, 42, 620-682.

Julian, T. B. (2001). MRI: a role in clinical trials. *Journal of Magnetic Resonance Imaging*, 13, 837-841.

Kerlikowske, K., Grady, D., Barclay, J., Sickles, E. A., & Ernster, V. (1996). Effect of age, breast density, and family history on the sensitivity of first screening mammography. *JAMA*, 276, 33-38.

Kinkel, K., & Vlastos, G. (2001). MR imaging: breast cancer staging and screening. *Seminars in Surgical Oncology*, 20, 187-196.

Kneeshaw, P. J., Turnbull, L. W., & Drew, P. J. (2003). Current applications and future direction of MR mammography. *British Journal of Cancer*, 88, 4-10.

Kossoff, M. B. (2000). Ultrasound of the breast. *World Journal of Surgery*, 24, 143-157.

Kriege, M., Brekelmans, C. T., Boetes, C., Rutgers, E. J., Oosterwijk, J. C., Tollenaar, R. A., Manoliu, R. A., et al. (2001). MRI screening for breast cancer in women with familial or genetic predisposition: design of the Dutch National Study (MRISC). *Familial Cancer*, 1, 163-168.

Kriege, M., Brekelmans, C. T., Obdeijn, I. M., Boetes, C., Zonderland, H. M., Muller, S. H., Kok, T., et al. (2006a). Factors affecting sensitivity and specificity of screening mammography and MRI in women with an inherited risk for breast cancer. *Breast Cancer Research and Treatment,* 100, 109-19. Epub ahead of print Jun 22.

Kriege, M., Brekelmans, C. T. M., Boetes, C., Besnard, P. E., Zonderland, H. M., Obdeijn, I. M., Manoliu, R. A., et al. (2005a). Efficacy of magnetic resonance imaging and mammography for breast cancer screening in women with a familial or genetic predisposition. *Obstetrical & Gynecological Survey*, 60, 107-109.

Kriege, M., Brekelmans, C. T. M., Boetes, C., Muller, S. H., Zonderland, H. M., Obdeijn, I. M., Manoliu, R. A., et al. (2006b). Differences between first and subsequent rounds of the MRISC breast cancer screening program for women with a familial or genetic predisposition. *Cancer*, 106, 2318-2326.

Kriege, M., Brekelmans, C. T. M., & Klijn, J. G. M. (2005b). MRI screening for breast cancer in women with a familial or genetic predisposition. *Imaging Decisions*, 9, 11-18.

Kuhl, C. K. (2002). High-risk screening: multi-modality surveillance of women at high risk for breast cancer (proven or suspected carriers of a breast cancer susceptibility gene). *Journal of Experimental & Clinical Cancer Research*, 21, 103-106.

Kuhl, C. K. (2004). Screening of women with hereditary risk of breast cancer. *Clinical Breast Cancer*, 5, 269-271.

Kuhl, C. K., Kuhn, W., & Schild, H. (2005). Management of women at high risk for breast cancer: new imaging beyond mammography. *Breast*, 14, 480-486.

Kuhl, C. K., Schmutzler, R. K., Leutner, C. C., Kempe, A., Wardelmann, E., Hocke, A., Maringa, M., et al. (2000). Breast MR imaging screening in 192 women proved or suspected to be carriers of a breast cancer susceptibility gene: preliminary results. *Radiology*, 215, 267-279.

Kuschel, B., Hauenstein, E., Kiechle, M., & Meindl, A. (2006). Hereditary breast and ovarian cancer - Current clinical guidelines in Germany. *Breast Care*, 1, 8-14.

Lai, M. S., Yen, M. F., Kuo, H. S., Koong, S. L., Chen, T. H., & Duffy, S. W. (1998). Efficacy of breast-cancer screening for female relatives of breast-cancer-index cases: Taiwan multicentre cancer screening (TAMCAS). *International Journal of Cancer*, 78, 21-26.

Law, J., & Faulkner, K. (2001). Cancers detected and induced, and associated risk and benefit, in a breast screening programme. *British Journal of Radiology*, 74, 1121-1127.

Leach, M. O., Eeles, R. A., Turnbull, L. W., Dixon, A. K., Brown, J., Hoff, R. J., Coulthard, A., et al. (2002). The UK national study of magnetic resonance imaging as a method of screening for breast cancer (MARIBS). *Journal of Experimental & Clinical Cancer Research*, 21, 107-114.

Lee, C. H., & Weinreb, J. C. (2004). The use of magnetic resonance imaging in breast cancer screening. *Journal of the American College of Radiology*, 1, 176-182.

Lehman, C. D., Schnall, M. D., Kuhl, C. K., & Harms, S. E. (2004). Report of the Working Groups on Breast MRI: report of the High-Risk Screening Group. *Breast Journal*, 10 Suppl 2, S9-S12.

Levavi, H., & Sabah, G. (2003). BRCA susceptibility genes - a review of current conservative management of BRCA mutation carriers. *European Journal of Gynaecological Oncology*, 24, 463-466.

Liberman, L. (2004). Breast cancer screening with MRI--what are the data for patients at high risk? *New England Journal of Medicine*, 351, 497-500.

Mackay, J., Rogers, C., Fielder, H., Blamey, R., Macmillan, D., Boggis, C., Brown, J., et al. (2001). Development of a protocol for evaluation of mammographic surveillance services in women under 50 with a family history of breast cancer. *Journal of Epidemiology & Biostatistics*, 6, 365-369; discussion 371-365.

McDonald, S., Saslow, D., & Alciati, M. H. (2004). Performance and reporting of clinical breast examination: a review of the literature. *CA: a Cancer Journal for Clinicians*, 54, 345-361.

Meurisse, M., Defechereux, T., Meurisse, N., Bataille, Y., & Hamoir, E. (2002). New challenges in the treatment of early breast cancer or surgery for early breast cancer ... can less be more? *Acta Chirurgica Belgica*, 102, 97-109.

Moller, P., Evans, D. G., Haites, N. E., Vasen, H., Reis, M. M., Andersonk, E., Apold, J., et al. (2000). Women at risk for inherited breast or ovarian cancer: Guidelines for follow-up. *CME Journal of Gynecologic Oncology*, 5, 269-271.

Moller, P., Evans, G., Haites, N., Vasen, H., Reis, M. M., Anderson, E., Apold, J., et al. (1999). Guidelines for follow-up of women at high risk for inherited breast cancer: consensus statement from the Biomed 2 Demonstration Programme on Inherited Breast Cancer. *Disease Markers*, 15, 207-211.

Morris, E. A. (2003). Screening for breast cancer with MRI. *Seminars in Ultrasound, CT & MR*, 24, 45-54.

Myles, J. P., Nixon, R. M., Duffy, S. W., Tabar, L., Boggis, C., Evans, G., Shenton, A., et al. (2003). Bayesian evaluation of breast cancer screening using data from two studies. *Statistics in Medicine*, 22, 1661-1674.

Orel, S. G., & Schnall, M. D. (1999). High Risk Screening Working Group report. *Journal of Magnetic Resonance Imaging*, 10, 995-1005.

Padhani, A. R., Ah-See, M. L., & Makris, A. (2005). MRI in the detection and management of breast cancer. *Expert Review of Anticancer Therapy*, 5, 239-252.

Pharoah, P. D. P., Stratton, J. F., & Mackay, J. (1998). Screening for breast and ovarian cancer: the relevance of family history. *British Medical Bulletin*, 54, 823-838.

Pichert, G., Bolliger, B., Buser, K., & Pagani, O. (2003). Evidence-based management options for women at increased breast/ovarian cancer risk. *Annals of Oncology*, 14, 9-19.

Plevritis, S. K., Kurian, A. W., Sigal, B. M., Daniel, B. L., Ikeda, D. M., Stockdale, F. E., & Garber, A. M. (2006). Cost-effectiveness of screening BRCA1/2 mutation carriers with breast magnetic resonance imaging. *JAMA*, 295, 2374-2384.

Prance, S. (2004). Surveillance of BRCA1 and BRCA2 mutation carriers with magnetic resonance imaging, ultrasound, mammography, and clinical breast examination. *Women's Oncology Review*, 4, 315-316.

Quan, M. L., & Petrek, J. A. (2003). Clinical implications of hereditary breast cancer. *Advances in Surgery*, 37, 197-212.

Rizzatto, G. J. (2001). Towards a more sophisticated use of breast ultrasound. *European Radiology*, 11, 2425-2435.

Robson, M. (2004). Breast cancer surveillance in women with hereditary risk due to BRCA1 or BRCA2 mutations. *Clinical Breast Cancer*, 5, 260-268; discussion 269-271.

Robson, M. E. (2002). Clinical considerations in the management of individuals at risk for hereditary breast and ovarian cancer. *Cancer Control*, 9, 457-465.

Robson, M. E., & Offit, K. (2004). Breast MRI for women with hereditary cancer risk. *Journal of the American Medical Association*, 292, 1368-1370.

Rothenberg, B. (2006). Computer-aided detection of malignancy with magnetic resonance imaging of the breast. *Blue Cross Blue Shield TEC Assessment Program*, 21, 1-23.

Rubinstein, W. S., Latimer, J. J., Sumkin, J. H., Huerbin, M., Grant, S. G., & Vogel, V. G., 3rd (2006). A study of 0.5 Tesla dedicated magnetic resonance imaging for the detection of breast cancer in young, high-risk women. *BMC Womens Health*, 6, 10.

Schnall, M. D. (2001). Application of magnetic resonance imaging to early detection of breast cancer. *Breast Cancer Research*, 3, 17-21.

Schnall, M. D. (2003). Breast MR imaging. *Radiologic Clinics of North America*, 41, 43-50.

Smith, J. A., & Andreopoulou, E. (2004). An overview of the status of imaging screening technology for breast cancer. *Annals of Oncology*, 15, i18-i26.

Smith, R. A., Saslow, D., Andrews Sawyer, K., Burke, W., Costanza, M. E., Evans, I. W., Foster Jr, R. S., et al. (2003). American Cancer Society Guidelines for breast cancer screening: update 2003. *CA: a Cancer Journal for Clinicians*, 53, 141-169.

Strobel, E. S., & Fritschk, E. (2002). Hereditary premenopausal breast cancer. *Onkologie*, 25, 24-27.

Tabar, L., Vitak, B., Chen, H. H., Duffy, S. W., Yen, M. F., Chiang, C. F., Krusemo, U. B., et al. (2000). The Swedish Two-County Trial twenty years later. Updated mortality results and new insights from long-term follow-up. *Radiologic Clinics of North America*, 38, 625-651.

Tilanus-Linthorst, M., Verhoog, L., Obdeijn, I. M., Bartels, K., Menke-Pluymers, M., Eggermont, A., Klijn, J., et al. (2002). A BRCA1/2 mutation, high breast density and prominent pushing margins of a tumor independently contribute to a frequent false-negative mammography.[erratum appears in Int J Cancer 2002 Dec 20;102(6):665]. *International Journal of Cancer*, 102, 91-95.

Tilanus-Linthorst, M. M., Obdeijn, I. M., Bartels, K. C., de Koning, H. J., & Oudkerk, M. (2000). First experiences in screening women at high risk for breast cancer with MR imaging. *Breast Cancer Research & Treatment*, 63, 53-60.

Trafalis, D. T. P., & Athanassiou, A. E. (2005). A guideline for the management of women at substantially increased risk of breast cancer development. *Journal of BUON*, 10, 443-458.

Tranin, A. S. (2005). Hereditary Breast and ovarian cancer: providing personalized medicine. *AWHONN Lifelines*, 9, 372-376.

Turchetti, D., Cortesi, L., Federico, M., Romagnoli, R., & Silingardi, V. (2002). Hereditary risk of breast cancer: not only BRCA. *Journal of Experimental & Clinical Cancer Research*, 21, 17-21.

Unic, I., Stalmeier, P. F., Peer, P. G., & van Daal, W. A. (1997). A review on family history of breast cancer: screening and counseling proposals for women with familial (non-hereditary) breast cancer. *Patient Education & Counseling*, 32, 117-127.

Vasen, H. F., Tesfay, E., Boonstra, H., Mourits, M. J., Rutgers, E., Verheyen, R., Oosterwijk, J., et al. (2005). Early detection of breast and ovarian cancer in families with BRCA mutations. *European Journal of Cancer*, 41, 549-554.

Warmuth, M. A., Sutton, L. M., & Winer, E. P. (1997). A review of hereditary breast cancer: from screening to risk factor modification. *American Journal of Medicine*, 102, 407-415.

Warner, E., & Causer, P. A. (2005). MRI surveillance for hereditary breast-cancer risk. *Lancet*, 365, 1747-1749.

Warner, E., Heisey, R. E., Goel, V., Carroll, J. C., & McCready, D. R. (1999). Hereditary breast cancer. Risk assessment of patients with a family history of breast cancer. *Canadian Family Physician*, 45, 104-112.

Warner, E., Plewes, D. B., Hill, K. A., Causer, P. A., Zubovits, J. T., Jong, R. A., Cutrara, M. R., et al. (2004). Surveillance of BRCA1 and BRCA2 mutation carriers with magnetic resonance imaging, ultrasound, mammography, and clinical breast examination. *JAMA*, 292, 1317-1325.

Warner, E., Plewes, D. B., Shumak, R. S., Catzavelos, G. C., Di Prospero, L. S., Yaffe, M. J., Goel, V., et al. (2001). Comparison of breast magnetic resonance imaging, mammography, and ultrasound for surveillance of women at high risk for hereditary breast cancer. *Journal of Clinical Oncology*, 19, 3524-3531.

Warren, R. (2001). Screening women at high risk of breast cancer on the basis of evidence. *European Journal of Radiology*, 39, 50-59.

Warren, R. M., Pointon, L., Caines, R., Hayes, C., Thompson, D., Leach, M. O., & UK MRI Breast Screening Study (2002). What is the recall rate of breast MRI when used for screening asymptomatic women at high risk? *Magnetic Resonance Imaging*, 20, 557-565.

Warren, R. M. L., Pointon, L., Thompson, D., Hoff, R., Gilbert, F. J., Padhani, A., Easton, D., et al. (2005). Reading protocol for dynamic contrast-enhanced MR images of the breast: Sensitivity and specificity analysis. *Radiology*, 236, 779-788.

Willey, S. C. (2004). Breast magnetic resonance image screening and ductal lavage in women at high genetic risk of breast carcinoma. *Women's Oncology Review*, 4, 153-154.

Zielinski, S. L. (2004). MRI more sensitive than mammography for women at high risk of breast cancer. *Journal of the National Cancer Institute*, 96, 1275.

Appendix 6

RETRIEVED STUDIES APPRAISED FOR REVIEW

Banks, E., Reeves, G., Beral, V., Bull, D., Crossley, B., Simmonds, M., Hilton, E., et al. (2004). Influence of personal characteristics of individual women on sensitivity and specificity of mammography in the Million Women Study: Cohort study. *British Medical Journal*, 329, 477-479.

Brekelmans, C. T. M., Seynaeve, C., Bartels, C. C. M., Tilanus-Linthorst, M. M. A., Meijers-Heijboer, E. J., Crepin, C. M. G., van Geel, A. N., et al. (2001). Effectiveness of breast cancer surveillance in BRCA1/2 gene mutation carriers and women with high familial risk. *Journal of Clinical Oncology*, 19, 924-930.

Calderon-Margalit, R., & Paltiel, O. (2004). Prevention of breast cancer in women who carry BRCA1 or BRCA2 mutations: a critical review of the literature. *International Journal of Cancer*, 112, 357-364.

Chart, P. L., & Franssen, E. (1997). Management of women at increased risk for breast cancer: preliminary results from a new program. *CMAJ Canadian Medical Association Journal*, 157, 1235-1242.

Crystal, P., Strano, S. D., Shcharynski, S., & Koretz, M. J. (2003). Using sonography to screen women with mammographically dense breasts. *AJR American Journal of Roentgenology*, 181, 177-182.

Elmore, J. G., Armstrong, K., Lehman, C. D., & Fletcher, S. W. (2005). Screening for breast cancer. *JAMA*, 293, 1245-1256.

Federico, M., Maiorana, A., Mangone, L., Turchetti, D., Canossi, B., Cortesi, L., Romagnoli, R., et al. (1999). Identification of families with hereditary breast and ovarian cancer for clinical and mammographic surveillance: The Modena Study Group proposal. *Breast Cancer Research & Treatment*, 55, 213-221.

Gui, G. P., Hogben, R. K., Walsh, G., A'Hern, R., & Eeles, R. (2001). The incidence of breast cancer from screening women according to predicted family history risk: does annual clinical examination add to mammography? *European Journal of Cancer*, 37, 1668-1673.

Gui, G. P., Kadayaprath, G., Darhouse, N., Self, J., Ward, A., A'Hern, R., & Eeles, R. (2006). Clinical outcome and service implications of screening women at increased breast cancer risk from a family history. *European Journal of Surgical Oncology,* 32, 719-724. Epub ahead of print Jun 16.

Halapy, E., Chiarelli, A. M., Klar, N., & Knight, J. A. (2005). Accuracy of breast screening among women with and without a family history of breast and/or ovarian cancer. *Breast Cancer Research & Treatment*, 90, 299-305.

Halapy, E. E., Chiarelli, A. M., Klar, N., & Knight, J. (2004). Breast screening outcomes in women with and without a family history of breast and/or ovarian cancer. *Journal of Medical Screening*, 11, 32-38.

Hartman, A. R., Daniel, B. L., Kurian, A. W., Mills, M. A., Nowels, K. W., Dirbas, F. M., Kingham, K. E., et al. (2004). Breast magnetic resonance image screening and ductal lavage in women at high genetic risk for breast carcinoma. *Cancer*, 100, 479-489.

Hou, M. F., Chuang, H. Y., Ou-Yang, F., Wang, C. Y., Huang, C. L., Fan, H. M., Chuang, C. H., et al. (2002). Comparison of breast mammography, sonography and physical examination for screening women at high risk of breast cancer in Taiwan. *Ultrasound in Medicine & Biology*, 28, 415-420.

Irwig, L., Houssami, N., & van Vliet, C. (2004). New technologies in screening for breast cancer: a systematic review of their accuracy. *British Journal of Cancer*, 90, 2118-2122.

Kerlikowske, K., Carney, P. A., Geller, B., Mandelson, M. T., Taplin, S. H., Malvin, K., Ernster, V., et al. (2000). Performance of screening mammography among women with and without a first-degree relative with breast cancer. *Annals of Internal Medicine*, 133, 855-863.

Kolb, T. M., Lichy, J., & Newhouse, J. H. (1998). Occult cancer in women with dense breasts: detection with screening US--diagnostic yield and tumor characteristics. *Radiology*, 207, 191-199.

Kollias, J., Sibbering, D. M., Blamey, R. W., Holland, P. A., Obuszko, Z., Wilson, A. R., Evans, A. J., et al. (1998). Screening women aged less than 50 years with a family history of breast cancer. *European Journal of Cancer*, 34, 878-883.

Kriege, M., Brekelmans, C. T., Boetes, C., Besnard, P. E., Zonderland, H. M., Obdeijn, I. M., Manoliu, R. A., et al. (2004). Efficacy of MRI and mammography for breast-cancer screening in women with a familial or genetic predisposition. *New England Journal of Medicine*, 351, 427-437.

Kuhl, C. K., Schrading, S., Leutner, C. C., Morakkabati-Spitz, N., Wardelmann, E., Fimmers, R., Kuhn, W., et al. (2005b). Mammography, breast ultrasound, and magnetic resonance imaging for surveillance of women at high familial risk for breast cancer. *Journal of Clinical Oncology*, 23, 8469-8476.

Lalloo, F., Boggis, C. R., Evans, D. G., Shenton, A., Threlfall, A. G., & Howell, A. (1998). Screening by mammography, women with a family history of breast cancer. *European Journal of Cancer*, 34, 937-940.

Leach, M. O., Boggis, C. R., Dixon, A. K., Easton, D. F., Eeles, R. A., Evans, D. G., Gilbert, F. J., et al. (2005). Screening with magnetic resonance imaging and mammography of a UK population at high familial risk of breast cancer: a prospective multicentre cohort study (MARIBS). [erratum appears in Lancet. 2005 May 28-Jun 3;365(9474):1848]. *Lancet*, 365, 1769-1778.

Lehman, C. D., Blume, J. D., Weatherall, P., Thickman, D., Hylton, N., Warner, E., Pisano, E., et al. (2005). Screening women at high risk for breast cancer with mammography and magnetic resonance imaging. *Cancer*, 103, 1898-1905.

Lucassen, A., Watson, E., & Eccles, D. (2001). Evidence based case report: Advice about mammography for a young woman with a family history of breast cancer. *British Medical Journal*, 322, 1040-1042.

Macmillan, R. D. (2000). Screening women with a family history of breast cancer--results from the British Familial Breast Cancer Group. *European Journal of Surgical Oncology*, 26, 149-152.

Mark, D. (2003). Magnetic resonance imaging of the breast in screening women considered to be at high genetic risk of breast cancer. *Blue Cross Blue Shield TEC Assessment Program*, 18, 1-14.

Maurice, A., Evans, D. G., Shenton, A., Ashcroft, L., Baildam, A., Barr, L., Byrne, G., et al. (2006). Screening younger women with a family history of breast cancer - does early detection improve outcome? *European Journal of Cancer*, 42, 1385-1390.

Moller, P., Reis, M. M., Evans, G., Vasen, H., Haites, N., Anderson, E., Steel, C. M., et al. (1999). Efficacy of early diagnosis and treatment in women with a family history of breast cancer. European Familial Breast Cancer Collaborative Group. *Disease Markers*, 15, 179-186.

Morris, E. A., Liberman, L., Ballon, D. J., Robson, M., Abramson, A. F., Heerdt, A., & Dershaw, D. D. (2003). MRI of occult breast carcinoma in a high-risk population. *AJR American Journal of Roentgenology*, 181, 619-626.

Murday, V., Pears, R., Ball, J., Eeles, R., & Hodgson, S. (2004). An audit of screening for familial breast cancer before 50 years in the South Thames Region - have we got it right? *Familial Cancer*, 3, 29-34.

Myles, J., Duffy, S., Nixon, R., Boggis, C., Howell, A., Shenton, A., & Evans, G. (2001). Initial results of a study into the effectiveness of breast cancer screening in a population identified to be at high risk. *Revue d Epidemiologie et de Sante Publique*, 49, 471-475.

Nixon, R. M., Pharoah, P., Tabar, L., Krusemo, U. B., Duffy, S. W., Prevost, T. C., & Chen, H. H. (2000). Mammographic screening in women with a family history of breast cancer: some results from the Swedish two-county trial. *Revue d Epidemiologie et de Sante Publique*, 48, 325-331.

O'Driscoll, D., Warren, R., MacKay, J., Britton, P., & Day, N. E. (2001). Screening with breast ultrasound in a population at moderate risk due to family history. *Journal of Medical Screening*, 8, 106-109.

Podo, F., Sardanelli, F., Canese, R., D'Agnolo, G., Natali, P. G., Crecco, M., Grandinetti, M. L., et al. (2002). The Italian multi-centre project on evaluation of MRI and other imaging modalities in early detection of breast cancer in subjects at high genetic risk. *Journal of Experimental & Clinical Cancer Research*, 21, 115-124.

Scheuer, L., Kauff, N., Robson, M., Kelly, B., Barakat, R., Satagopan, J., Ellis, N., et al. (2002). Outcome of preventive surgery and screening for breast and ovarian cancer in BRCA mutation carriers. *Journal of Clinical Oncology*, 20, 1260-1268.

Sim, L. S., Hendriks, J. H., & Fook-Chong, S. M. (2004). Breast ultrasound in women with familial risk of breast cancer. *Annals of the Academy of Medicine, Singapore*, 33, 600-606.

Stoutjesdijk, M. J., Boetes, C., Jager, G. J., Beex, L., Bult, P., Hendriks, J. H., Laheij, R. J., et al. (2001). Magnetic resonance imaging and mammography in women with a hereditary risk of breast cancer. *Journal of the National Cancer Institute*, 93, 1095-1102.

Trecate, G., Vergnaghi, D., Bergonzi, S., De Simone, T., Fengoni, E., Costa, C., Spatti, G., et al. (2003). Breast MRI screening in patients with increased familial and/or genetic risk for breast cancer: A preliminary experience. *Tumori*, 89, 125-131.

Warner, E., Plewes, D. B., Hill, K. A., Causer, P. A., Zubovits, J. T., Jong, R. A., Cutrara, M. R., et al. (2004). Surveillance of BRCA1 and BRCA2 mutation carriers with magnetic resonance imaging, ultrasound, mammography, and clinical breast examination. *JAMA*, 292, 1317-1325.