EVIDENCE RELEVANT TO GUIDELINES FOR THE INVESTIGATION OF BREAST SYMPTOMS

SECOND EDITION

FEBRUARY 2006

PREPARED BY THE NATIONAL BREAST CANCER CENTRE

FUNDED BY THE AUSTRALIAN GOVERNMENT
DEPARTMENT OF HEALTH AND AGEING
CONTENTS

List of tables 5
List of figures 6
Acknowledgements 7
Executive summary 8
Introduction 11
Methods 13
Searching the literature 13
Appraising quality and validity 15
Objective 1: Appraisal of international guidelines 17
Objective 2: The accuracy and clinical utility of the triple test in women presenting with a breast lump or asymmetrical prominence 19
The accuracy of a positive and negative triple test 20
The accuracy of combinations of components of the triple test 21
The accuracy of combined results for mammography and fine needle aspiration cytology 21
Objective 3: The value of core biopsy versus fine needle aspiration cytology as a component of the triple test in diagnosing breast cancer in women presenting with a breast lump or asymmetrical prominence 24
Diagnostic performance 25
Choosing between core biopsy and fine needle aspiration cytology 25
Objective 4: The value of ultrasound in diagnosing breast cancer in women presenting with a breast lump or asymmetrical prominence 28
Study quality 29
Diagnostic performance 29
Age-specific results 31
Objective 5: The clinical significance of nipple discharge and the utility of cytology in diagnosing breast cancer

The clinical significance of nipple discharge

The diagnostic accuracy of cytology of discharge smears in the diagnosis of breast cancer

Conclusions

References

List of abbreviations

APPENDIX A: Search strategies

APPENDIX B: Websites

APPENDIX C: List of Included Studies

APPENDIX D: Second-stage Exclusions

APPENDIX E: Tables for ultrasound section
LIST OF TABLES

Table 1. Electronic databases used in the search of the primary literature 13
Table 2. Entry criteria used to assess eligibility of published material 14
Table 3. Structure and content of the AGREE Instrument 16
Table 4. Items used to assess the quality of diagnostic test studies 17
Table 5. Standardised domain scores for clinical practice guidelines according to AGREE criteria 19
Table 6. True positive rate and false positive rate for the triple test and each component 21
Table 7. Distribution of results for clinical examination, mammography and fine needle aspiration with women with and without cancer 22
Table 8. Distribution of results for mammography and fine needle aspiration of women with and without cancer 24
Table 9. Reported diagnostic performance of studies examining the accuracy of core biopsy versus FNAC 26
Table 10. Reported diagnostic performance of mammography and ultrasonography in the diagnosis of breast cancer 31
Table 11. True and false positive rates for mammography with or without ultrasonography by age group from Zonderland and colleagues 32
Table 12. Studies examining the association between nipple discharge and breast cancer 33
Table 13. Distributions of nipple discharge characteristics as percentages of all discharges and of those cases found to have cancer 34
Table 14. Probability of cancer by age and characteristic of discharge 34
Table 15. Reported diagnostic performance of cytology of nipple discharge in the diagnosis of breast cancer 35
LIST OF FIGURES

**Figure 1.** The probability of breast cancer based on results of the triple test 21

**Figure 2.** The probabilities of breast cancer based on results of individual components of the triple test 23

**Figure 3.** The probabilities of breast cancer based on results of mammography and fine needle aspiration cytology 24
ACKNOWLEDGEMENTS

The National Breast Cancer Centre gratefully acknowledges the support of all the individuals and groups who contributed to the development of this report.

Working Group
This report was developed with input from a multidisciplinary Working Group:

Dr Jane Armes (The Royal College of Pathologists of Australasia)
Dr Kathleen Burns (Australian Divisions of General Practice)
Dr Bronwyn Kennedy (The Royal Australian College of General Practitioners)
Dr Marjorie Kossoff (The Royal Australian and New Zealand College of Radiologists)
Ms Ros Lawson (Breast Cancer Network Australia)
Dr Warwick Lee (radiologist with expertise in breast imaging)
Mr David Oliver (Royal Australasian College of Surgeons)
Dr Wendy Raymond (The Royal College of Pathologists of Australasia)
Dr Julie Thompson (general practitioner with expertise in rural health)

This report was developed by Dr Elmer V Villanueva. Dr Adele Weston of Health Technology Analysts P/L provided technical input into the systematic review of evidence on ultrasonography. Ms Caroline Nehill and Dr Helen Zorbas contributed to the conception and design of the report and provided substantial critical comments on its intellectual content.

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

This report builds, replicates and extends the findings of the National Breast Cancer Centre (NBCC) review Evidence Relevant to Guidelines for the Investigation of Breast Symptoms, First Edition released in 1997. The review was used to support a number of NBCC publications, including The Investigation of a New Breast Symptom: A Guide for General Practitioners (the Guide).

The present edition has the following specific objectives:

1. To assess guidelines produced by international agencies for the quality of their evidence base.
2. To examine the evidence of the accuracy and clinical utility of the triple test in women presenting with a breast lump or asymmetrical prominence.
3. To evaluate the value of core biopsy versus fine needle aspiration cytology (FNAC) as a component of the triple test in women presenting with a breast lump or asymmetrical prominence.
4. To evaluate the value of ultrasound in diagnosing breast cancer in women presenting with a breast lump or asymmetrical prominence.
5. To examine the clinical significance of nipple discharge in terms of the crude and age-specific probabilities of breast cancer and the use and interpretation of clinical features and cytology in the assessment of cancer risk.

Standard techniques of systematic reviewing were applied to evaluate the evidence published since 1996, while retaining as much as possible of the approach used in the first edition.

APPRAISAL OF INTERNATIONAL GUIDELINES

Three guidelines on the diagnosis of breast disease were identified. These were released by the following organisations: NHS Cancer Screening Programmes (Cancer Research UK), Institute for Clinical Systems Improvement (USA) and the Brigham and Women’s Hospital (USA). While each showed strengths in specific areas, none adequately described the guidelines development process or provided sufficient coverage of issues relevant to the Australian setting. Guideline development groups focusing on the management of women with breast symptoms in the general practice setting should consider these guidelines in the light of local patterns of care, differences in health and financing systems, and end user and patient behaviour.
THE ACCURACY AND CLINICAL UTILITY OF THE CLASSIC TRIPLE TEST

The triple test result may significantly modify the pre-test probability of a patient. The triple test is considered positive if any one of its component tests results in a positive result (i.e. “indeterminate”, “suspicious” or “malignant”). A positive finding on one of the components of the triple test will steadily increase the probability of cancer across the entire range of pre-test probabilities. A negative triple test suggests that the probability of breast cancer is less than 1%, unless the pre-test probability is greater than 60%.

The components of the triple test may be found to be positive or negative in a correlated manner. In the presence of positive FNAC, a single additional clinical or mammographic finding increases the odds of breast cancer by approximately equivalent amounts. Similarly, isolated clinical or mammographic findings in the presence of a negative FNAC result will decrease the odds of breast cancer by equivalent amounts. There is very little change in the post-test odds if clinical and mammographic findings agree but both disagree with FNAC results.

THE VALUE OF CORE BIOPSY VERSUS FINE NEEDLE ASPIRATION CYTOLOGY AS A COMPONENT OF THE TRIPLE TEST

Two studies conducted in patients who were referred for breast symptoms or triple test assessment showed that core biopsy is at least as sensitive and specific as FNAC in the diagnostic setting.

CHOOSING BETWEEN CORE BIOPSY AND FINE NEEDLE ASPIRATION CYTOLOGY

Core biopsy and FNAC are complementary procedures. The decision to use FNAC, core biopsy or both will be influenced by various factors including the size and clinical characteristics of the lesion, the characteristics of the lesion identified on imaging, the likelihood of achieving a definitive diagnosis, the woman’s ability to tolerate more than one procedure, the expertise of the clinician performing the procedure, the preference of the managing clinician, the availability of pathologists with experience in cytology, the need for hormone receptor assay or tumour marker studies in inoperable tumours, consideration of subsequent surgical management, and the need for a rapid result.
THE VALUE OF ULTRASOUND IN DIAGNOSING BREAST CANCER

The quality of the evidence from five studies is variable. The sensitivity of mammography increases with age. Three studies indicate that the use of ultrasound in addition to mammography improves sensitivity and specificity when compared to the use of mammography alone. The results of two studies suggest that the sensitivity of ultrasound alone is superior to that of mammography alone in younger women. Differences in false positive rates are not consistent across all age groups.

It is unlikely that the addition of ultrasound will appreciably add to the detection of breast cancer in older women who have undergone the triple test. In younger women, ultrasound is the preferred imaging technique due to the increased sensitivity and specificity in comparison to mammography. There have been no studies that prospectively address the question of the optimal age at which one imaging modality is superior to another. Expert consensus opinion suggests that mammography and ultrasound be used as complementary modalities for the evaluation of symptomatic women, particularly those aged between 35 and 50 years.

THE CLINICAL SIGNIFICANCE OF NIPPLE DISCHARGE

Nipple discharge most commonly presents with serous or milky characteristics in women younger than 60 years. Bloody discharge accounts for approximately one-tenth of all nipple discharge.

The risk of cancer varies with age and the characteristic of the discharge. In women younger than 60 years, the risk of cancer is less than 1% for those with non-bloody discharge, increasing to 3% if blood is present. Bloody discharge is associated with a higher risk for cancer in those aged 60 years and older.

The utility of guaiac-based tests (e.g. Haemoccult™) in the detection of blood in nipple discharge and cancer is contentious. Research suggests that a negative guaiac-based test result may be used to rule out the presence of blood, but that a positive finding may need further investigation. The result of guaiac-based tests should not be used to determine the presence of breast cancer and further workup should be performed to reach a definitive diagnosis.

CONCLUSIONS

Additional evidence of the effectiveness of diagnostic procedures published since 1996 has been identified and evaluated. Guides to clinical practice in the investigation of women presenting with breast symptoms should incorporate these findings in order to maximise the detection of breast cancer in these women.
INTRODUCTION

In 1997, the National Breast Cancer Centre (NBCC) produced a resource for general practitioners that detailed the recommended steps to be taken in determining the risk of breast cancer in patients presenting with breast symptoms. The Investigation of a New Breast Symptom: A Guide for General Practitioners (the Guide) was pioneering work on three fronts. First, the Guide was the first attempt to apply evidence-based principles to the diagnosis of breast cancer in the setting of general practice in Australia. The Guide was based on the findings of a Technical Report entitled Evidence Relevant to Guidelines for the Investigation of Breast Symptoms. The Technical Report was developed on the basis of systematic reviews of the evidence at a time when the principles of structured reviewing of the diagnostic test literature were in its infancy. Several techniques used in the visionary approach to the development of the Guide are used and taught today as standard methodological approaches.

Second, the Guide and Technical Report were comprehensive in their scope. Both accomplished four clearly stated objectives:

- searching and appraising guidelines existing at the time of development to establish rigour and acceptability in the Australian context
- evaluating the use of the triple test in diagnosing breast cancer in women presenting with a breast lump or asymmetrical prominence
- evaluating the value of ultrasound in diagnosing breast cancer in women presenting with a breast lump or asymmetrical prominence
- estimating the age-specific probability of breast cancer in women presenting with nipple discharge, as well as the utility and interpretation of clinical and cytological findings.

Third, the Guide was widely distributed and remains one of the most popular resources of the NBCC. Endorsement by the Royal Australian College of General Practitioners (RACGP), as well as targeted implementation strategies used in disseminating the resource, were pivotal in the Guide gaining general acceptance in the community of health professionals.
Subsequent studies on the approach of general practitioners to the investigation of breast symptoms conducted after the release of the Guide showed that there were significant improvements in five of six areas. Compared to practice prior to dissemination of the resource and the institution of implementation strategies (including seminars, practice feedback and the opportunity to earn Practice Assessment Points as part of the RACGP Quality Assurance and Continuing Education program), the following changes were noted:

- findings not consistent with normal breast tissue or hormonal changes that were referred for imaging increased from 69% to 90% (p=0.005)
- indeterminate, suspicious or malignant findings on imaging that prompted surgical referral increased from 64% to 85% (p=0.005)
- imaging studies performed prior to fine needle aspiration (FNA) or fine needle aspiration biopsy (FNAB) increased from 72% to 98% (p=0.02)
- FNAB or core biopsies (or surgical referrals) conducted in the case of normal imaging results not consistent with clinical findings increased from 29% to 51% (p=0.05)
- mammographic studies in the event of suspicious or malignant findings on clinical examination or ultrasound increased from 57% to 83% (p=0.005).

In the seven years since the release of the Guide and the Technical Report, there have been many advances in the understanding of breast cancer that have resulted in improvements in diagnosis, management and supportive care afforded to women.

The present report describes evidence published since the Guide was released in 1997. It builds, replicates and extends the findings of a previous version of the Technical Report. The findings contained here will be used to produce an update of the Guide.

The aim of this report is to evaluate the evidence relevant to critical decisions about appropriate investigations used to detect cancer in women who present to their general practitioners with a breast lump or asymmetrical prominence or with a nipple discharge. This report has the following specific objectives:

1. To assess guidelines produced by international agencies for the quality of their evidence base.
2. To examine the evidence of the accuracy and clinical utility of the triple test in women presenting with a breast lump or asymmetrical prominence.
3. To evaluate the value of core biopsy versus fine needle aspiration cytology as a component of the triple test in women presenting with a breast lump or asymmetrical prominence.
4. To evaluate the value of ultrasound in diagnosing breast cancer in women presenting with a breast lump or asymmetrical prominence.
5. To examine the clinical significance of nipple discharge in terms of the crude and age-specific probabilities of breast cancer and the use and interpretation of clinical features and cytology in the assessment of cancer risk.
METHODS

We applied standard techniques of systematic reviewing to identify, appraise and report on the evidence supporting each of the five objectives enumerated above. Where applicable, we retained as much of the original approach outlined previously.²

SEARCHING THE LITERATURE

For searches of the primary literature, we used the electronic databases listed in Table 1. The first edition searched the primary literature to 1997. To minimise duplication while ensuring some degree of overlap in order to refine search strategies, we identified all peer-reviewed articles published in English from 1996 to 2004.

Table 1. Electronic databases used in the search of the primary literature*

<table>
<thead>
<tr>
<th>Database</th>
<th>Period Searched</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACP Journal Club (Ovid)</td>
<td>1996 to September/October 2004</td>
</tr>
<tr>
<td>Biological Abstracts (Ovid)</td>
<td>1996 to 2001</td>
</tr>
<tr>
<td>Cancer Library (NCI)</td>
<td>1996 to September 2004</td>
</tr>
<tr>
<td>CancerLit (NCI)</td>
<td>1996 to 2002</td>
</tr>
<tr>
<td>CINAHL (Ovid)</td>
<td>1996 to September 2004</td>
</tr>
<tr>
<td>Cochrane Central Register of Controlled Trials†</td>
<td>Issue 2, 2004</td>
</tr>
<tr>
<td>Cochrane Database of Systematic Reviews†</td>
<td>Issue 2, 2004</td>
</tr>
<tr>
<td>Database of Abstracts of Reviews of Effects†</td>
<td>Issue 2, 2004</td>
</tr>
<tr>
<td>EMBASE (Ovid)</td>
<td>December week 2 2004</td>
</tr>
<tr>
<td>Health Technology Assessment Database†</td>
<td>Issue 2, 2004</td>
</tr>
<tr>
<td>Journals@Ovid</td>
<td>1996 to September</td>
</tr>
<tr>
<td>Medline (Ovid)</td>
<td>1996 to December week 2 2004</td>
</tr>
<tr>
<td>NHS Economic Evaluation Database†</td>
<td>Issue 2, 2004</td>
</tr>
<tr>
<td>PubMed</td>
<td>1996 to September 2004</td>
</tr>
</tbody>
</table>

* Abbreviations: CINAHL, Cumulative Index to Nursing and Allied Health Literature; NCI, National Cancer Institute US.
† Available as part of the Cochrane Library.

We used search terms tailored to the particular objective; these are listed in Appendix A. In addition, we searched websites of international organisations and professional bodies for technical reports; these websites are listed in Appendix B.

All materials were collected and appraised using entry criteria listed in Table 2. We applied the eligibility criteria using a two-stage process. Titles, abstracts or executive summaries of all materials were evaluated for general concordance with the topic, population and outcomes of interest. Publications that failed to meet eligibility criteria at this stage were excluded from further consideration. Those that were considered in the second stage, in which the full text was appraised, included publications whose titles,
abstracts or executive summaries suggested that they meet eligibility criteria or publications in which a decision could not be made from reading the summaries.

Studies that were found to meet eligibility criteria after review of the full text provide the basis of this report and are listed in Appendix C. Second stage exclusions are listed in Appendix D.

Table 2. Entry criteria used to assess eligibility of published material

<table>
<thead>
<tr>
<th>Objective</th>
<th>Entry criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>General criteria</td>
<td>Original primary research, excluding duplicate publications, reviews or opinion pieces</td>
</tr>
<tr>
<td>Guideline assessment (Obj 1)</td>
<td>The report or guideline was endorsed by, or prepared under the auspices of, a professional body or similar organisation</td>
</tr>
<tr>
<td>Triple test (Obj 2)</td>
<td>Consecutive patients recruited from a general practice setting</td>
</tr>
<tr>
<td></td>
<td>Quantitative information on the true and false positive rates was available or easily derived</td>
</tr>
<tr>
<td></td>
<td>Results for individual components and combinations of components were reported</td>
</tr>
<tr>
<td>Core biopsy (Obj 3)</td>
<td>Results were reported for a single group of at least 100 patients for core biopsy and FNAC</td>
</tr>
<tr>
<td></td>
<td>Consecutive patients recruited from a clinical setting, excluding such studies conducted in restricted subsets of patients such as those: where the population was limited to women already known to have breast cancer</td>
</tr>
<tr>
<td></td>
<td>where the study population comprised only those patients who had had, or were scheduled for, lump or breast removal surgery because of a high degree of suspicion of breast cancer on the basis of their diagnostic imaging results</td>
</tr>
<tr>
<td></td>
<td>where the study population comprised only those patients who had biopsy because of a high degree of suspicion of breast cancer</td>
</tr>
<tr>
<td></td>
<td>Quantitative information on the true and false positive rates was available or easily derived</td>
</tr>
<tr>
<td>Ultrasound (Obj 4)</td>
<td>Results were reported for a single group of at least 100 patients for at least two of the listed diagnostic modalities in the same patients: mammography, ultrasonography or a combined interpretation of the mammography and ultrasound findings</td>
</tr>
<tr>
<td></td>
<td>Consecutive patients recruited from a clinical setting, excluding such studies conducted in restricted subsets of patients such as those: where the population was limited to women already known to have breast cancer</td>
</tr>
<tr>
<td></td>
<td>where the study population comprised only those patients who had had, or were scheduled for, lump or breast removal surgery because of a high degree of suspicion of breast cancer on the basis of their diagnostic imaging results</td>
</tr>
<tr>
<td></td>
<td>where the study population comprised only those patients who had biopsy because of a high degree of suspicion of breast cancer</td>
</tr>
<tr>
<td></td>
<td>Quantitative information on the true and false positive rates was available or easily derived</td>
</tr>
<tr>
<td>Nipple discharge (Obj 5)</td>
<td>Results were reported for a group of patients presenting with unilateral or bilateral nipple discharge</td>
</tr>
<tr>
<td></td>
<td>Determinations of the presence or absence of cancer utilised accepted procedures</td>
</tr>
</tbody>
</table>
APPRAISING QUALITY AND VALIDITY

GUIDELINES AND ALGORITHMS

We used the Appraisal of Guidelines for Research and Evaluation (AGREE) Instrument\(^4\) to assess the validity and quality of clinical practice guidelines (CPGs) and algorithms. The AGREE Instrument, developed on the basis of previous work in the area of CPG development and implementation,\(^5\)-\(^7\) was created specifically for the purpose of assessing the extent to which potential biases in CPG development had been addressed by guideline developers, the degree to which recommendations were internally and externally valid, and the impact and feasibility of other factors linked to uptake and dissemination.

The AGREE Instrument consists of 23 items under six key domains (Table 3). Each item was rated on a four-point scale from 4 (“Strongly Agree”) and 1 (“Strongly Disagree”). Scores of each domain were standardised to produce a composite score for that domain representing the percentage of the maximum possible score. Given that domains were independent, no single score was calculated. As of the publication date of this report, the AGREE Instrument was available at www.agreecollaboration.org/instrument.\(^4\)

<table>
<thead>
<tr>
<th>Domain</th>
<th>Number of items</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scope and purpose</td>
<td>3</td>
</tr>
<tr>
<td>Stakeholder involvement</td>
<td>4</td>
</tr>
<tr>
<td>Rigour of development</td>
<td>7</td>
</tr>
<tr>
<td>Clarity and presentation</td>
<td>4</td>
</tr>
<tr>
<td>Applicability</td>
<td>3</td>
</tr>
<tr>
<td>Editorial independence</td>
<td>2</td>
</tr>
</tbody>
</table>

* Abbreviations: AGREE, Appraisal of Guidelines for Research and Evaluation

DIAGNOSTIC TEST STUDIES

Methodological information relating to the study population, the nature of the intervention and the definition of outcomes were extracted from the included studies. Particular attention was paid to methodological factors known to influence the quality of diagnostic studies.

A detailed assessment of study quality was undertaken using a modification of the diagnostic-specific checklist published by the Cochrane Screening and Diagnostic Tests Methods Group (Table 4). Quality assessment was undertaken on the basis of the information clearly enunciated in the published paper. No attempt was made to contact authors to seek clarification.
Table 4. Items used to assess the quality of diagnostic test studies

<table>
<thead>
<tr>
<th>Item</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Was the test compared with a valid reference measure?</td>
<td></td>
</tr>
<tr>
<td>Were the test and the reference standard measured independently (blind) of each other?</td>
<td></td>
</tr>
<tr>
<td>Were all patients assessed by the reference standard or was the choice of patients who were assessed by the reference standard independent of the test results (avoidance of verification bias)?</td>
<td></td>
</tr>
<tr>
<td>Was the test performed independently of all other clinical information?</td>
<td></td>
</tr>
<tr>
<td>Was the reference standard measured before any treatment interventions were started with knowledge of test results (avoidance of treatment paradox)?</td>
<td></td>
</tr>
<tr>
<td>Were the mammography results independent of the ultrasound results?</td>
<td></td>
</tr>
<tr>
<td>Were mammography and ultrasound measured at a similar point in time?</td>
<td></td>
</tr>
<tr>
<td>Did the study include a consecutive sample of patients with breast symptoms who were referred for imaging?</td>
<td></td>
</tr>
<tr>
<td>Was a representative spectrum of disease captured?</td>
<td></td>
</tr>
<tr>
<td>Was the disease prevalence indicative of the target population?</td>
<td></td>
</tr>
</tbody>
</table>
OBJECTIVE 1: APPRAISAL OF INTERNATIONAL GUIDELINES

We identified three guidelines for the diagnosis of breast disease published by organisations other than the NBCC.\textsuperscript{8-10} The first set of guidelines\textsuperscript{8} was a revision of the only set of recommendations identified in the first edition of this report. The guidelines were published by NHS Cancer Screening Programmes, Cancer Research UK. Targeted towards general practitioners, they outline the referral pathways for women with breast problems including lumps, pain, nipple discharge, changes to the nipple or skin and family history.

While, on the whole, the overall objectives and target users of the guidelines were specifically described, the set of patients and clinical questions to which it was meant to apply was not. It was not clear whether individuals from relevant professional groups were consulted in its development, whether patients’ views and preferences were sought, or whether the guidelines were piloted in targeted users. The development of the guidelines was likewise poorly described. There was no information about the methods used to search for evidence, the criteria used to select evidence or the methods used to formulate recommendations, making decisions about its external validity difficult. The different options for management were clearly presented, but not supported by tools for its application.

Guidelines published by the Institute for Clinical Systems Improvement (USA)\textsuperscript{9} clearly defined the overall objectives, clinical questions, target users and patients. The development included representatives from a wide range of professional groups, although it was unclear whether patients’ views and preferences were sought. The rigour of development, clarity and presentation and applicability of these guidelines were the most transparent of the three.

The final set of guidelines identified from the electronic search were published by The Brigham and Women’s Hospital (USA),\textsuperscript{10} ostensibly to apply to the clinical services within particular hospitals. The objectives, clinical questions, target users and patients to whom the guidelines were to apply were relatively clearly described, although it was unclear whether the views and preferences of patients were specifically sought and incorporated. The method of development was vague, making the links between specific recommendations and its supporting evidence difficult to assess. Moreover, there was no discussion about the potential organisational barriers, cost implications or audit activities involved in implementing the guidelines.
A summary of the appraisal of the three guidelines against AGREE criteria is presented in Table 5. Among the three guidelines identified, the development, preparation and presentation of guidelines published by the Institute for Clinical Systems Improvement (USA)\(^9\) were the most concordant with AGREE criteria. There was marginal improvement in the quality of the guidelines produced by Cancer Research UK\(^8\) compared to those found previously.\(^2\)

Guideline development groups focusing on the management of women with breast symptoms in the general practice setting should consider these guidelines in the light of local patterns of care, differences in health and financing systems, and end user and patient behaviour.

Table 5. Standardised domain scores (%) for clinical practice guidelines according to AGREE criteria*

<table>
<thead>
<tr>
<th>Guideline</th>
<th>Scope and purpose</th>
<th>Stakeholder involvement</th>
<th>Rigour of development</th>
<th>Clarity and presentation</th>
<th>Applicability</th>
<th>Editorial Independence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cancer Research UK(^8)</td>
<td>44</td>
<td>42</td>
<td>24</td>
<td>42</td>
<td>0</td>
<td>17</td>
</tr>
<tr>
<td>Institute for Clinical Systems Improvement (USA)(^9)</td>
<td>100</td>
<td>58</td>
<td>76</td>
<td>92</td>
<td>67</td>
<td>67</td>
</tr>
<tr>
<td>Brigham and Women’s Hospital (USA)(^10)</td>
<td>78</td>
<td>58</td>
<td>38</td>
<td>58</td>
<td>0</td>
<td>17</td>
</tr>
</tbody>
</table>

* Domain scores represent the proportion of the maximum score of 100% to which the particular guideline was judged to meet
OBJECTIVE 2:
THE ACCURACY AND CLINICAL UTILITY OF THE TRIPLE TEST IN WOMEN PRESENTING WITH A BREAST LUMP OR ASYMMETRICAL PROMINENCE

The triple test traditionally consists of the triad of clinical examination (C), mammography (M) and fine needle aspiration cytology (FNAC). Generally, the triple test is deemed to give a positive result if any of its components is reported as “suspicious” or “malignant”. The use of other diagnostic modalities in place of or in addition to the traditional components has been reported. While we focused on the classical components of the triple test in this report, we also described the effectiveness of core biopsy in place of FNAC. Additional information about the effectiveness of ultrasound is given in the subsequent section.

The initial search identified 1,093 citations. Preliminary application of the inclusion and exclusion criteria reduced the number to 156 requiring assessment of the full text. For the determination of the diagnostic characteristics of the classic triple test, none of the studies met entry criteria. Most were excluded on the basis of failing to provide information about individual components of the triple test or various combinations of the triple test components in a single study population. Two available studies examined the accuracy of core biopsy. Second stage exclusions are listed in Appendix D and article flow is shown in Appendix E.

A systematic review\textsuperscript{11} published in 2002, updating data from the first edition of the Technical Report,\textsuperscript{2} confirms this general lack of published information fulfilling the eligibility criteria (Les Irwig, Personal Communication, 2005). In the absence of additional information, results from the first edition of the Technical Report\textsuperscript{2} are described.
THE ACCURACY OF A POSITIVE AND NEGATIVE TRIPLE TEST

The triple test is considered positive if any one of its component tests results in a positive result (i.e. “suspicious” or “malignant”). Six studies\textsuperscript{12-17} provided sufficient information to estimate the true positive rate (TPR) and false positive rate (FPR) of the triple test (Table 6).

Table 6. True positive rate (TPR) and false positive rate (FPR) for the triple test and each component

<table>
<thead>
<tr>
<th></th>
<th>Triple test</th>
<th>Clinical examination</th>
<th>Mammography</th>
<th>Fine needle aspiration cytology</th>
</tr>
</thead>
<tbody>
<tr>
<td>TPR, %*</td>
<td>99.6</td>
<td>85</td>
<td>90</td>
<td>91</td>
</tr>
<tr>
<td>FPR, %†</td>
<td>38</td>
<td>20</td>
<td>27</td>
<td>7</td>
</tr>
<tr>
<td>Specificity, %†</td>
<td>62</td>
<td>80</td>
<td>73</td>
<td>93</td>
</tr>
</tbody>
</table>

* Based on 735 cancers † Based on 1128 non-cancers

The triple test result may significantly modify the pre-test probability of a patient (Figure 1). A negative triple test (C-M-F-) suggests that the probability of breast cancer is less than 1%, unless the pre-test probability is greater than 60%. A positive finding on one of the components of the triple test will steadily increase the probability of cancer across the entire range of pre-test probabilities. For instance, a patient with a pre-test probability of 5% will, on receipt of a positive triple test, have a post-test probability of 12%.

**Figure 1. The probability of breast cancer based on results of the triple test.**

Abbreviations: -, a negative test result (i.e. “benign”); +, a positive test result (i.e. “suspicious” or “malignant”); C, clinical examination; F, fine needle aspiration cytology; M, mammography
THE ACCURACY OF COMBINATIONS OF COMPONENTS OF THE TRIPLE TEST

The components of the triple test may be found to be positive or negative in a correlated manner. Empirically, the results of the six studies providing information on all combinations of the three components are summarised in Table 7.

<table>
<thead>
<tr>
<th>Component and result</th>
<th>Combined data (%)†</th>
<th>Likelihood ratio (+)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Cancer</td>
<td>No cancer</td>
</tr>
<tr>
<td>C M F</td>
<td></td>
<td></td>
</tr>
<tr>
<td>+ + +</td>
<td>73.3</td>
<td>1.6</td>
</tr>
<tr>
<td>+ + -</td>
<td>5.3</td>
<td>8.5</td>
</tr>
<tr>
<td>+ - +</td>
<td>5.7</td>
<td>1.0</td>
</tr>
<tr>
<td>- + +</td>
<td>8.7</td>
<td>2.1</td>
</tr>
<tr>
<td>+ - -</td>
<td>1.1</td>
<td>8.5</td>
</tr>
<tr>
<td>- + -</td>
<td>2.4</td>
<td>14.3</td>
</tr>
<tr>
<td>- - +</td>
<td>3.0</td>
<td>2.1</td>
</tr>
<tr>
<td>- - -</td>
<td>0.4</td>
<td>61.9</td>
</tr>
</tbody>
</table>

* Abbreviations: -, a negative test result (i.e. “benign”); +, a positive test result (i.e. “suspicious” or “malignant”); C, clinical examination; F, fine needle aspiration cytology; M, mammography
† Results are pooled from six studies providing information on all three tests (individually and in combination) in a single study population

The key item, likelihood ratio (+), expresses the extent to which the pre-test odds of disease would be modified by the test result. As expected, the biggest increase in pre-test odds occurs when all three components record a positive result while the biggest decrease occurs when all three components are negative. In the presence of positive FNAC, a single additional clinical or mammographic finding increases the odds of breast cancer by approximately equivalent amounts (5.7 and 4.1 times the pre-test level, respectively). Similarly, isolated clinical or mammographic findings in the presence of a negative FNAC result will decrease the odds of breast cancer by equivalent amounts (10% and 20% of the pre-test level, respectively). There is very little change in the post-test odds if clinical and mammographic findings agree but both disagree with FNAC results (i.e. C+M+F- or C-M+F+). A visual representation of the effect is given in Figure 2 for a range of pre-test probabilities.
Figure 2. The probabilities of breast cancer based on results of individual components of the triple test.

Abbreviations: -, a negative test result (i.e. "benign"); +, a positive test result (i.e. "suspicious" or "malignant"); C, clinical examination; F, fine needle aspiration cytology; M, mammography

The accuracy of combined results for mammography and fine needle aspiration cytology

If the results of the clinical examination and medical history were incorporated into the pre-test probabilities prior to the application of imaging modalities and invasive procedures, all information would serve as a new pre-test probability that is modified by the results of mammography and fine needle aspiration cytology. Such results are presented in Table 8 and Figure 3.

Table 8. Distribution of results for mammography and fine needle aspiration of women with and without cancer*

<table>
<thead>
<tr>
<th>Component and result</th>
<th>Combined data (%)†</th>
<th>Likelihood ratio (+)</th>
</tr>
</thead>
<tbody>
<tr>
<td>M</td>
<td>F</td>
<td>Cancer</td>
</tr>
<tr>
<td>+</td>
<td>+</td>
<td>82.0</td>
</tr>
<tr>
<td>+</td>
<td>-</td>
<td>7.8</td>
</tr>
<tr>
<td>-</td>
<td>+</td>
<td>8.7</td>
</tr>
<tr>
<td>-</td>
<td>-</td>
<td>1.5</td>
</tr>
</tbody>
</table>

* Abbreviations: -, a negative test result (i.e. "benign"); +, a positive test result (i.e. "suspicious" or "malignant"); F, fine needle aspiration cytology; M, mammography
† Results are pooled from six studies12-17 providing information on all three tests (individually and in combination) in a single study population.
Figure 3. The probabilities of breast cancer based on results of mammography and fine needle aspiration cytology.

Abbreviations: -, a negative test result (i.e. “benign”); +, a positive test result (i.e. “suspicious” or “malignant”); F, fine needle aspiration cytology; M, mammography.
OBJECTIVE 3:

THE VALUE OF CORE BIOPSY VERSUS FINE NEEDLE ASPIRATION CYTOLOGY AS A COMPONENT OF THE TRIPLE TEST IN DIAGNOSING BREAST CANCER IN WOMEN PRESENTING WITH A BREAST LUMP OR ASYMMETRICAL PROMINENCE

The research question addressed was whether the use of core biopsy in place of FNAC as a component of the triple test improves the diagnosis of breast cancer presenting with breast symptoms. Sensitivity, specificity and positive predictive value (PPV) of the imaging data were extracted.

Of a total 49 studies identified, two\textsuperscript{18,19} were available to determine the diagnostic characteristics of core biopsy in place of FNAC. All were conducted in patients who were referred for breast symptoms or triple test assessment. Exclusions are listed in Appendix D.

Shannon and colleagues\textsuperscript{18} presented the details of 1768 consecutive patients referred for triple test assessment, 822 (46.5\%) of whom presented because of a breast symptom. Core biopsies or FNAC were performed on the basis of the clinician’s best judgment. All test results, including imaging and clinical findings, were discussed in multidisciplinary meetings. Westenend and colleagues\textsuperscript{19} described the results of FNAC and core biopsy performed on 286 breast lesions, 81.1\% of which were palpable on presentation. The authors did not distinguish between symptomatic and screen-detected lesions.
DIAGNOSTIC PERFORMANCE

Table 9 summarises the diagnostic characteristics of FNAC and core biopsy reported in the two studies. Shannon and colleagues reported that core biopsy produced statistically significant improvements in TPR compared to FNAC but Westenend and colleagues failed to detect such a difference.

Table 9. Reported diagnostic performance of studies examining the accuracy of core biopsy versus FNAC*

<table>
<thead>
<tr>
<th>Study and test</th>
<th>N</th>
<th>Prevalence (%)</th>
<th>TPR</th>
<th>FPR</th>
<th>PPV (M)</th>
<th>PPV (S)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shannon 200118</td>
<td>822</td>
<td>19.8†</td>
<td>99.5</td>
<td>14.5</td>
<td>99.5</td>
<td>---‡</td>
</tr>
<tr>
<td>CB</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FNAC</td>
<td></td>
<td></td>
<td>88.8</td>
<td>52.4</td>
<td>100.0</td>
<td>84.0</td>
</tr>
<tr>
<td>Westenend 200119</td>
<td>286</td>
<td>27.6</td>
<td>88.0</td>
<td>10.0</td>
<td>99.0</td>
<td>100.0</td>
</tr>
<tr>
<td>CB</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FNAC</td>
<td></td>
<td></td>
<td>92.0</td>
<td>18.0</td>
<td>100.0</td>
<td>78.0</td>
</tr>
</tbody>
</table>

* Abbreviations: CB, core biopsy; FNAC, fine needle aspiration cytology; FPR, false positive rate; N, total sample size; PPV (M), positive predictive value for malignant result; PPV (S), positive predictive value for suspicious result  † Limited to the subset of patients attending for breast symptoms ‡ Data not provided

These results show that core biopsy is at least as sensitive and specific as fine needle aspiration cytology in the diagnostic setting.

CHOOSING BETWEEN CORE BIOPSY AND FINE NEEDLE ASPIRATION CYTOLOGY20

Both FNA cytology and core biopsy can be performed as outpatient procedures. Hormone receptor studies can be performed on specimens derived from both techniques. While core biopsy and FNAC are complementary procedures, one or another may be more appropriate in a particular clinical situation.

The decision to use FNA cytology, core biopsy or both will be influenced by various factors including the size and clinical characteristics of the lesion, the characteristics of the lesion identified on imaging, the likelihood of achieving a definitive diagnosis, the woman’s ability to tolerate more than one procedure, the expertise of the clinician performing the procedure, the preference of the managing clinician, the availability of pathologists with experience in cytology, the need for hormone receptor assay or tumour marker studies in inoperable tumours, consideration of subsequent surgical management, and the need for a rapid result.

Fine needle aspiration cytology is quicker and less expensive to perform than core biopsy. It is generally less traumatic than core biopsy, is associated with a low complication rate and, in most instances, does not require local anaesthesia. The results for FNAC are available relatively rapidly (within a few hours in some centres) and the presence of a cytopathologist may facilitate an immediate result. It is generally more
appropriate for women taking anticoagulant medication but is inappropriate for the assessment of microcalcifications.

However, FNAC has its disadvantages. Proper training is required in the preparation of quality smears and considerable expertise in cytology is required to interpret specimens. Fine needle aspiration cytology does not enable the pathologist to distinguish between ductal carcinoma in situ (DCIS) and invasive cancer. In some cases, a definitive diagnosis may be difficult to reach in some lesions. Fine needle aspiration cytology may not be the sampling technique of choice for lesions that are relatively hypocellular and yield scanty epithelial material.

Core biopsy yields tissue fragments useful in the determination of the presence of DCIS or invasive carcinoma; tissue may also be available for adjunctive tests. It is useful in the evaluation of lesions likely to be of low histological grade and in those presenting as architectural distortions for which FNAC may fail or has failed to provide a diagnosis. More generally, core biopsy can be used when FNAC fails to correlate with clinical findings or imaging studies.

Core biopsy is the investigation of choice in the evaluation of microcalcifications. It may be preferred when appropriate cytological expertise is not available. Core biopsies are associated with an increased risk of complications, including haematoma and haemorrhage. A lesion previously identified on mammography may not be identified in subsequent open biopsy due to complete removal of the lesion or in the presence of inflammation and fibrosis due to biopsy. In addition, core biopsy may interfere with the interpretation of the subsequent excision biopsy, particularly with grading and the estimation of the size of the lesion. This is particularly relevant in the case of small lesions.

The reliability of FNAC and core biopsy as diagnostic procedures generally depends on the skill of the operator. It is not always possible to immediately assess the adequacy of core biopsy performed for a mass lesion or architectural distortion. Core biopsy specimens require adequate fixation and processing, usually taking a minimum of one working day before results can be available. Core biopsies require local anaesthesia and are generally more expensive than FNAC.

There are no absolute rules determining whether FNAC or core biopsy is the more appropriate investigation. General guidance is provided in Breast Fine Needle Aspiration Cytology and Core Biopsy: A Guide for Practice by the NBCC to inform clinical decisions in the context of symptomatic breast lesions.20
INDICATIONS FOR FINE NEEDLE ASPIRATION CYTOLOGY

- investigation of palpable masses, regardless of whether they are considered benign or malignant
- investigation of impalpable image-detected masses that are considered likely to be benign or with typically malignant features
- evaluation of cystic lesions with atypical imaging features
- confirmation of a diagnosis of breast cancer when core biopsy is not available, not possible, or contraindicated.

INDICATIONS FOR CORE BIOPSY

- investigation of lesions with suspicious features identified on imaging that cannot be identified on ultrasound
- further evaluation of a benign cytological pattern in the presence of a suspicious lesion on imaging
- further evaluation of a lesion for which cytology results are atypical or suspicious
- when a single surgical procedure is the desired outcome (for example, wide excision and axillary dissection). It must be noted that a core biopsy showing only DCIS does not exclude the possibility of the presence of invasive carcinoma in the lesion
- evaluation of microcalcifications that are radiologically indeterminate, suspicious or typically malignant. In such cases, core biopsies should be radiographed prior to histological processing to confirm adequate sampling of the lesion.
OBJECTIVE 4:
THE VALUE OF ULTRASOUND IN DIAGNOSING BREAST CANCER IN WOMEN PRESENTING WITH A BREAST LUMP OR ASYMMETRICAL PROMINENCE

The research addressed the following questions:

- Does the use of ultrasound in addition to mammography improve the diagnosis of breast cancer in women presenting with breast symptoms?
- Does the use of ultrasound instead of mammography improve the diagnosis of breast cancer in women presenting with breast symptoms?

In addition, the impact of age upon the relative diagnostic performance of ultrasound and mammography was of particular interest. The review did not focus on the role of mammography and ultrasound in screening or in the monitoring of women who had been diagnosed with breast cancer.

Sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), diagnostic accuracy and diagnostic odds ratio based upon dichotomous interpretation of the imaging data were extracted. If results were reported in the publication using alternative cut-points, these were also extracted.

After the elimination of duplicate citations and the addition of further citations identified from the reference lists of key publications, a total of 613 unique citations remained. Application of the inclusion and exclusion criteria reduced the number of articles to four.21-24 Several studies were excluded because they reported results only for a highly selected subgroup of the patient cohort in question. The patients included in such studies were not representative of the cohort referred for the initial investigation of a breast symptom, which would have distorted the diagnostic results if applied to the broader group of patients referred for diagnostic mammography and ultrasound. Confirmation that these studies represented the wrong patient population is evident from the higher prevalence of breast cancer reported in these studies compared to those conducted in consecutive referred populations.25,26 Second-stage exclusions are listed in Appendix D.
STUDY QUALITY

In general, the design of the included studies addressed the research questions appropriately. In all studies, the population consisted of adult women referred for diagnostic imaging on the basis of a breast lump or other symptom. In one study the contralateral breast was also imaged. As this breast was asymptomatic, the results contributed by the second breast should be considered as screening.

In several of the studies, the patients undergoing both mammography and ultrasound represent a subset of all referred patients. As this subgroup was selected on the basis of demographic or clinical characteristics, the results may not be generalisable to the entire population of referred patients (Table A5-1 and Table A5-2).

Several of the studies used a five-point scoring system for level of certainty of diagnosis, however this was ultimately dichotomised to calculated diagnostic performance.

The studies generally were of adequate quality to address the research questions (Table A5-2). They represented an improvement in quality from many of the earlier studies. In particular, all studies attempt to include patients with negative imaging results and determine their outcomes. In three of the four studies, the ultrasound result was determined with knowledge of the mammography result.

DIAGNOSTIC PERFORMANCE

Table 10 presents the calculated diagnostic results. Houssami and colleagues designed their study in such a way as to match disease-positive patients with controls, resulting in a prevalence of disease of 50.7%, when the actual disease prevalence in the referred population was 1.9%. Similarly, the inclusion of the imaging results from the contralateral breast in the study by Flobbe and colleagues underestimated the calculated prevalence of disease due to dilution by results from paired-organ data. When calculated on the basis of individual patients, the prevalence of breast cancer was 6.3% and is likely to be a truer reflection of disease prevalence among the referred, symptomatic population. The prevalence of breast cancer in these studies ranged from 1.9% to 24.7%, with an overall prevalence of 3.9% (671 cases in 17,117 referrals).

ULTRASOUND INSTEAD OF MAMMOGRAPHY

Both studies reporting results for mammography alone versus ultrasound alone found that ultrasound gave a greater TPR and diagnostic odds ratio (DOR) and smaller FPR compared to mammography. While the earlier edition of this report did not find any consistency in TPR and FPR between ultrasound and mammography in six studies, five of the six studies reported that the DOR of ultrasound was greater than that of mammography.
ULTRASOUND IN ADDITION TO MAMMOGRAPHY

Three studies compared the incremental value of ultrasound in this setting.\textsuperscript{21,23,24} Two studies compared ultrasound plus mammography to mammography alone,\textsuperscript{23,24} while the other study reported results for clinical examination and mammography with or without ultrasound.\textsuperscript{21} All three studies reported increases in TPR and DOR and decreases in FPR when ultrasound was combined with mammography or ultrasound combined with clinical examination and mammography. Some increases in DOR were substantial: from 72.1 to 565.7 in the report by Flobbe and colleagues\textsuperscript{21}, and from 50.7 to 223.5 in the study by Zonderland and colleagues.\textsuperscript{24} While there were modest increases in the likelihood ratios positive, all three studies reported likelihood ratios negative of 0.05 or less.

Table 10. Reported diagnostic performance of mammography and ultrasonography in the diagnosis of breast cancer\textsuperscript{*}

<table>
<thead>
<tr>
<th>Study and basis of test result</th>
<th>N</th>
<th>Prevalence (%)</th>
<th>TPR</th>
<th>FPR</th>
<th>PPV</th>
<th>NPV</th>
<th>DOR</th>
<th>LR(+)</th>
<th>LR(-)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flobbe 2003\textsuperscript{21}</td>
<td>3,835</td>
<td>3.4†</td>
<td>82.9</td>
<td>8.1</td>
<td>26.2</td>
<td>99.4</td>
<td>55.0</td>
<td>10.21</td>
<td>0.19</td>
</tr>
<tr>
<td>M only</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>U only</td>
<td></td>
<td></td>
<td>87.6</td>
<td>4.5</td>
<td>40.4</td>
<td>99.6</td>
<td>150.4</td>
<td>19.53</td>
<td>0.13</td>
</tr>
<tr>
<td>C+M</td>
<td></td>
<td></td>
<td>91.5</td>
<td>13.0</td>
<td>19.7</td>
<td>99.7</td>
<td>72.1</td>
<td>7.06</td>
<td>0.10</td>
</tr>
<tr>
<td>C+M+U</td>
<td></td>
<td></td>
<td>96.9</td>
<td>5.2</td>
<td>39.2</td>
<td>99.9</td>
<td>565.7</td>
<td>18.51</td>
<td>0.03</td>
</tr>
<tr>
<td>Houssami 2003\textsuperscript{22}</td>
<td>473</td>
<td>50.7 (1.9)§</td>
<td>75.8</td>
<td>12.4</td>
<td>86.3</td>
<td>77.9</td>
<td>22.1</td>
<td>6.09</td>
<td>0.28</td>
</tr>
<tr>
<td>M only</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>U only</td>
<td></td>
<td></td>
<td>81.7</td>
<td>12.0</td>
<td>87.5</td>
<td>82.3</td>
<td>32.6</td>
<td>6.80</td>
<td>0.21</td>
</tr>
<tr>
<td>Shetty 2003\textsuperscript{23}</td>
<td>411</td>
<td>3.4</td>
<td>85.7</td>
<td>32.2</td>
<td>8.6</td>
<td>99.3</td>
<td>12.6</td>
<td>2.66</td>
<td>0.21</td>
</tr>
<tr>
<td>M only</td>
<td></td>
<td></td>
<td>100.0</td>
<td>11.6</td>
<td>23.3</td>
<td>100.0</td>
<td>N/A</td>
<td>8.63</td>
<td>0.00</td>
</tr>
<tr>
<td>M+U</td>
<td></td>
<td></td>
<td>86.0</td>
<td>10.8</td>
<td>72.2</td>
<td>95.1</td>
<td>50.7</td>
<td>7.94</td>
<td>0.16</td>
</tr>
<tr>
<td>Zonderland 1999\textsuperscript{24}</td>
<td>1,103</td>
<td>24.7</td>
<td>95.2</td>
<td>8.2</td>
<td>79.2</td>
<td>98.3</td>
<td>223.5</td>
<td>11.64</td>
<td>0.05</td>
</tr>
<tr>
<td>M only</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M+U</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Abbreviations: C, clinical examination; DOR, diagnostic odds ratio; FPR, false positive rate; LR(-), likelihood ratio negative; LR(+), likelihood ratio positive; M, mammography; N, total sample size; NPV, negative predictive value; PPV, positive predictive value; U, ultrasound
† Prevalence was 3.4% when asymptomatic contralateral breasts were included. Prevalence was 6.3% (127/2020) on a per patient basis, which is likely to be a truer reflection of prevalence in the symptomatic population
‡ Data for the same patients were reported in an earlier publication\textsuperscript{27}
§ The prevalence of breast cancer is artificial, as matched sampling of controls was conducted
|| Value is undefined
AGE-SPECIFIC RESULTS

Through the use of receiver operator characteristic (ROC) curves to compare clinical examination and mammography versus clinical examination, mammography and ultrasound in two distinct age groups —those 50 years and younger, and those older than 50 years—Flobbe and colleagues\textsuperscript{21} found that the addition of ultrasound increased the area under the curve (AUC) for both age groups (0.87 to 0.96 for patients 50 years and younger; 0.95 to 0.99 for patients older than 50 years; p=0.03 for both groups) in patients referred specifically for a breast lump. The improvements were largely due to ultrasound correctly ‘downgrading’ positive results found on clinical examination and mammography. The magnitude of the improvement was most pronounced in the younger group. These findings contrast with those found in the group of patients referred due to other breast symptoms such as pain, nipple discharge and skin abnormalities. The authors found no statistically significant difference in the AUC with or without ultrasound.

Houssami and colleagues\textsuperscript{22} also compared the TPR and FPR of mammography and ultrasound by age category. The TPR of ultrasound was better than mammography in all age groups up to 45 yrs, but in none of the \textit{a priori} defined age groups did the difference reach statistical significance. In both the 36–40 and 41–45 age groups, the lower 95% CI was close to zero, and the magnitude of the absolute difference in TPR was sizeable (15.4% and 14.3%, respectively). The large number of age categories (and, therefore, smaller number of patients in each category) may have prohibited these differences reaching statistical significance. When the results were calculated in a \textit{post hoc} fashion for women 45 years or younger, the TPR of ultrasound was significantly higher than mammography (84.9% versus 71.7%, p=0.035).

There was no obvious relationship between age and FPR for either mammography or ultrasound in the study by Houssami and colleagues.\textsuperscript{22} The investigators found minimal differences between the specificities of mammography and ultrasound across the age range, except for those between the ages of 36 and 40 years. In this age group, the specificity was 8.1% better with ultrasound (91.9% vs 83.8%, 95% CI for difference 0.7%, 16.9%).

Zonderland and colleagues\textsuperscript{24} provided a comprehensive report of the diagnostic characteristics of mammography alone and mammography combined with ultrasound across a range of age groups (Table 11). The addition of ultrasound to mammography improved the true positive rate of breast cancer diagnosis, particularly among women below the age of 50 years. In contrast, the addition of ultrasound had minimal impact upon the false positive rate.
### Table 11. True and false positive rates for mammography with or without ultrasonography by age group from Zonderland and colleagues*24

<table>
<thead>
<tr>
<th>Characteristic and test</th>
<th>Age group</th>
<th>30–39</th>
<th>40–49</th>
<th>50–59</th>
<th>60–69</th>
<th>≥70</th>
</tr>
</thead>
<tbody>
<tr>
<td>True positive rate (%)</td>
<td>M only</td>
<td>72</td>
<td>78</td>
<td>82</td>
<td>94</td>
<td>92</td>
</tr>
<tr>
<td></td>
<td>M and U</td>
<td>94</td>
<td>96</td>
<td>89</td>
<td>99</td>
<td>97</td>
</tr>
<tr>
<td>False positive rate (%)</td>
<td>M only</td>
<td>4</td>
<td>6</td>
<td>19</td>
<td>23</td>
<td>29</td>
</tr>
<tr>
<td></td>
<td>M and U</td>
<td>6</td>
<td>4</td>
<td>12</td>
<td>18</td>
<td>17</td>
</tr>
</tbody>
</table>

* Abbreviations: M, mammography; U, ultrasound

It is unlikely that the addition of ultrasound will appreciably add to the detection of breast cancer in older women who have had a thorough clinical examination and mammography. In younger women, ultrasound is the preferred imaging technique due to the increased true positive rate and similar (or reduced) false positive rate in comparison to mammography. The optimum age cut-point is still uncertain.
OBJECTIVE 5:

THE CLINICAL SIGNIFICANCE OF NIPPLE DISCHARGE
AND THE UTILITY OF CYTOLOGY IN DIAGNOSING
BREAST CANCER

THE CLINICAL SIGNIFICANCE OF NIPPLE DISCHARGE

From a total of 151 citations, we identified six studies\textsuperscript{28-33} that examined the relationship between nipple discharge and the probability of breast cancer (Table 12). As none provided the extensive cross-tabulations of findings found in Ciatto and colleagues’ study\textsuperscript{34} (identified in the first edition of this report), or its sample size or patient spectrum, some relevant results are presented here.

Table 12. Studies examining the association between nipple discharge and breast cancer*

<table>
<thead>
<tr>
<th>Study</th>
<th>N</th>
<th>Setting</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cabioglu 2003\textsuperscript{28}</td>
<td>146</td>
<td>Women with nipple discharge presenting to a university clinic over seven years</td>
</tr>
<tr>
<td>Florio 1999\textsuperscript{29}</td>
<td>1,251</td>
<td>Women with nipple discharge presenting to a university clinic over ten years</td>
</tr>
<tr>
<td>Gupta 2004\textsuperscript{30}</td>
<td>1,530\textsuperscript{†}</td>
<td>Presentations to breast clinics for nipple discharge</td>
</tr>
<tr>
<td>Hou 2002\textsuperscript{31}</td>
<td>487</td>
<td>Women with unilateral nipple discharge without palpable breast masses</td>
</tr>
<tr>
<td>Pritt 2004\textsuperscript{32}</td>
<td>395\textsuperscript{‡}</td>
<td>Presentations to a single breast clinic for nipple discharge</td>
</tr>
<tr>
<td>Sauter 2004\textsuperscript{33}</td>
<td>175</td>
<td>Women scheduled for diagnostic breast surgery</td>
</tr>
</tbody>
</table>

* Abbreviations: N, total sample size  † Includes 36 males  ‡ Includes 3 males

The most common presentations for nipple discharge are those with serous or milky characteristics occurring in women younger than 60 years. Bloody nipple discharge accounts for approximately one-tenth of all nipple discharge (Table 13).
Table 13. Distributions of nipple discharge characteristics as percentages of all discharges and of those cases found to have cancer

<table>
<thead>
<tr>
<th>Characteristic of nipple discharge</th>
<th>Age</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>All discharges</td>
<td>&lt; 40 years</td>
<td>40 to 59 years</td>
<td>≥ 60 years</td>
<td>All Ages</td>
</tr>
<tr>
<td>Serous</td>
<td>25</td>
<td>25</td>
<td>2</td>
<td>52</td>
</tr>
<tr>
<td>Milky</td>
<td>20</td>
<td>11</td>
<td>0</td>
<td>31</td>
</tr>
<tr>
<td>Purulent</td>
<td>2.5</td>
<td>4</td>
<td>&lt;1</td>
<td>7</td>
</tr>
<tr>
<td>Bloody</td>
<td>3</td>
<td>6</td>
<td>2</td>
<td>11</td>
</tr>
<tr>
<td>Total</td>
<td>51</td>
<td>47</td>
<td>4</td>
<td>100</td>
</tr>
<tr>
<td>All cancers</td>
<td>0</td>
<td>6</td>
<td>6</td>
<td>12</td>
</tr>
<tr>
<td>Serous</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>Milky</td>
<td>0</td>
<td>3</td>
<td>6</td>
<td>9</td>
</tr>
<tr>
<td>Purulent</td>
<td>13</td>
<td>29</td>
<td>32</td>
<td>74</td>
</tr>
<tr>
<td>Total</td>
<td>16</td>
<td>39</td>
<td>44</td>
<td>100</td>
</tr>
</tbody>
</table>

In those younger than 60 years, the risk of cancer is less than one percent for those with non-bloody discharge. This increases to three percent if blood is present. Bloody discharge is also associated with a higher risk for cancer in those aged 60 years and older, with the probability approaching 10 percent (Table 14).

Table 14. Probability (%) of cancer by age and characteristic of discharge

<table>
<thead>
<tr>
<th>Discharge</th>
<th>Age</th>
<th>&lt; 60</th>
<th>≥ 60</th>
</tr>
</thead>
<tbody>
<tr>
<td>Other</td>
<td></td>
<td>&lt; 1</td>
<td>*</td>
</tr>
<tr>
<td>Serous</td>
<td>&lt;1</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Bloody</td>
<td>3</td>
<td>9</td>
<td></td>
</tr>
</tbody>
</table>

* Few cases available to allow precise estimation

The utility of Hemoccult™ in the detection of blood in nipple discharge and cancer is contentious. Only three available studies provided evidence. An early study by Salmon and colleagues showed good correlation between Hemoccult™ results and the presence of blood in discharge specimens as determined by cytology. The TPR and FPR were 100% and 12.5%, respectively. This suggests that a negative Hemoccult™ result may be used to rule out the presence of blood, but that a positive finding may need further investigation.
The results from two studies suggest that Hemoccult™ findings from discharge specimens should not be used to determine the presence of breast cancer and that further workup should be performed to reach a definitive diagnosis. Simmons and colleagues\textsuperscript{36} reported the following diagnostic characteristics for Hemoccult™: TPR, 50%; FPR, 100%; PPV, 20%; and NPV, 0%. Ambrogetti and colleagues\textsuperscript{37} reported that the PPV for cancer with a positive Hemoccult™ result for nipple discharge was less than 10%.

THE DIAGNOSTIC ACCURACY OF CYTOLOGY OF DISCHARGE SMEARS IN THE DIAGNOSIS OF BREAST CANCER

Previous reports have estimated that cytology of discharge smears for breast cancer had a sensitivity of between 40% and 50% and a specificity of over 99\%.\textsuperscript{2} More recent results are inconsistent (Table 15), with one study showing that a negative test result reduces pre-test odds of disease by a sizeable amount,\textsuperscript{28} while the other showed that the positive test result is more clinically meaningful in adjusting pre-test probabilities of disease.\textsuperscript{32}

<table>
<thead>
<tr>
<th>Study</th>
<th>N</th>
<th>TPR</th>
<th>FPR</th>
<th>PPV</th>
<th>NPV</th>
<th>LR(+)</th>
<th>LR(-)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cabioglu 2003\textsuperscript{28}</td>
<td>94</td>
<td>26.7</td>
<td>18.9</td>
<td>28.6</td>
<td>79.6</td>
<td>1.41</td>
<td>0.001</td>
</tr>
<tr>
<td>Pritt 2004\textsuperscript{32}</td>
<td>46</td>
<td>84.6</td>
<td>3.0</td>
<td>91.2</td>
<td>94.1</td>
<td>28.2</td>
<td>0.16</td>
</tr>
</tbody>
</table>

* Abbreviations: FPR, false positive rate; LR(-), likelihood ratio negative; LR(+), likelihood ratio positive; N, total sample size; NPV, negative predictive value; PPV, positive predictive value
CONCLUSIONS

Additional evidence of the effectiveness of diagnostic procedures published since the release of *The Investigation of a New Breast Symptom: A Guide for General Practitioners* has been identified and evaluated. The implications of these findings for clinical practice should be considered in updates to the Guide with the goal of maximising the detection of breast cancer in women presenting with breast symptoms.


### LIST OF ABBREVIATIONS

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>AGREE</td>
<td>Appraisal of Guidelines for Research and Evaluation</td>
</tr>
<tr>
<td>AUC</td>
<td>area under the curve</td>
</tr>
<tr>
<td>C</td>
<td>clinical examination (in the context of the triple test)</td>
</tr>
<tr>
<td>CB</td>
<td>core biopsy</td>
</tr>
<tr>
<td>CINAHL</td>
<td>Cumulative Index to Nursing and Allied Health Literature</td>
</tr>
<tr>
<td>CPG</td>
<td>clinical practice guideline</td>
</tr>
<tr>
<td>DCIS</td>
<td>ductal carcinoma <em>in situ</em></td>
</tr>
<tr>
<td>DOR</td>
<td>diagnostic odds ratio</td>
</tr>
<tr>
<td>FNA</td>
<td>fine needle aspiration</td>
</tr>
<tr>
<td>FNAC</td>
<td>fine needle aspiration cytology (F, in the context of the triple test)</td>
</tr>
<tr>
<td>FPR</td>
<td>false positive rate</td>
</tr>
<tr>
<td>LR</td>
<td>likelihood ratio</td>
</tr>
<tr>
<td>M</td>
<td>mammography (in the context of the triple test)</td>
</tr>
<tr>
<td>N</td>
<td>total sample size</td>
</tr>
<tr>
<td>NBCC</td>
<td>National Breast Cancer Centre</td>
</tr>
<tr>
<td>NCI</td>
<td>National Cancer Institute (US)</td>
</tr>
<tr>
<td>NPV</td>
<td>negative predictive value</td>
</tr>
<tr>
<td>PPV</td>
<td>positive predictive value</td>
</tr>
<tr>
<td>RACGP</td>
<td>Royal Australian College of General Practitioners</td>
</tr>
<tr>
<td>ROC</td>
<td>receiver operator characteristic</td>
</tr>
<tr>
<td>TPR</td>
<td>true positive rate</td>
</tr>
<tr>
<td>U</td>
<td>ultrasound</td>
</tr>
</tbody>
</table>
APPENDIX A: SEARCH STRATEGIES

The following search strategies were constructed for the PubMed search interface. They were changed to suit the particular search engine (Table 1).

TRIPLE TEST


CORE BIOPSY VERSUS FINE NEEDLE ASPIRATION CYTOLOGY

("Breast Neoplasms"[MeSH] AND ((fine AND needle AND (aspiration or cytology)) OR "aspiration cytology" OR FNA OR FNAC) AND ((core AND biops*) OR CB OR CNB)) AND (compar* OR "triple test" OR palpable OR suspicious)) AND (specificity[Title/Abstract])

ULTRASOUND (USING EMBASE.COM)


NIPPLE DISCHARGE

((nipple* OR breast*) AND discharg*) AND ("Breast Neoplasms"[MeSH]) AND ((relative[Title/Abstract] AND risk*[Title/Abstract]) OR (relative risk[Text Word]) OR risks[Text Word] OR cohort studies[MeSH:noexp] OR (cohort[Title/Abstract] AND stud*[Title/Abstract])))
APPENDIX B:

WEBSITES

AUSTRALIAN SITES

Australia and New Zealand Horizon Scanning Network - http://www.horizonscanning.gov.au


INTERNATIONAL SITES


Alberta Heritage Foundation for Medical Research - http://www.ahfmr.ab.ca

Canadian Coordinating Office Health Technology Assessment - http://www.ccohta.ca

Centre for Reviews and Dissemination - http://www.york.ac.uk/inst/crd

ECRI - http://www.ecri.org/

Finnish Office for Health Care Technology Assessment - http://www.stakes.fi/finohta

Institute for Clinical Systems Improvement - http://www.icsi.org

National Coordinating Centre for health technology Assessment - http://www.hta.nhsweb.nhs.uk


National Horizon Scanning Centre - http://www.publichealth.bham.ac.uk/horizon

National Institute for Clinical Excellence - http://www.nice.org.uk


New Zealand Health Technology Assessment - http://nzhta.chmeds.ac.nz

NHS Quality Improvement Scotland - http://www.nhshealthquality.org/nhsqis

Oregon Health Policy and Research - http://www.ohprr.state.or.us

Swedish Council on Technology Assessment in Health Care - http://www.sbu.se

Veterans Affairs Technology Assessment Program - http://www.va.gov/vatap/
APPENDIX C:

LIST OF INCLUDED STUDIES


APPENDIX D:
SECOND-STAGE EXCLUSIONS

TRIPLE TEST


Evidence relevant to GUIDELINES FOR THE INVESTIGATION OF BREAST SYMPTOMS
CORE BIOPSY VERSUS FINE NEEDLE ASPIRATION BIOPSY


ULTRASOUND


Evidence relevant to GUIDELINES FOR THE INVESTIGATION OF BREAST SYMPTOMS


## APPENDIX E:

### TABLES FOR ULTRASOUND SECTION

Table E1. Characteristics of included studies.

<table>
<thead>
<tr>
<th>Study</th>
<th>Country</th>
<th>Setting</th>
<th>Patient Cohort</th>
<th>Mammography and ultrasound results reported for same patients?</th>
<th>Approach for combined result</th>
<th>Nature of mammography Categorisation of positive and negative Treatment of equivocal results</th>
<th>Nature of ultrasound Categorisation of positive and negative Treatment of equivocal results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flobbe 2003²¹</td>
<td>The Netherlands</td>
<td>Hospital radiology department; referred from GP or specialist</td>
<td>All patients referred for breast imaging from October 1999 to August 2000. Imaging results the contralateral breast were considered as screening. Consecutive; prospective; 2,020 patients (3,835 breasts, including 20 men); 16–90 years.</td>
<td>Yes. All patients referred within this period got U in addition to M. Results are reported for C + M, C + M + U. Combined results available for C + M (highest of both scores); C + M + U (complex definition, refer paper), but radiologist made overall interpretation when different.</td>
<td>Two view using Siemens Mammomat. Reported by radiologist assigning a score on five-point scale: 1=normal, 2=benign, 3=probably benign, 4=suggestive malignant, 5=malignant. Negative result = 1, 2; positive result = 3, 4, 5. Indeterminate M results recorded as 'probably benign' (score 3) for analysis.</td>
<td>Whole breast using Philips ATL5000, 12.5 MHz transducer. Performed by radiologist assigning a score on five-point scale: 1=normal, 2=benign, 3=probably benign, 4=suggestive malignant, 5=malignant. Negative result = 1, 2; positive result = 3, 4, 5. Indeterminate M results recorded as 'probably benign' (score 3) for analysis.</td>
<td></td>
</tr>
<tr>
<td>Houssami 2003†²²</td>
<td>Australia</td>
<td>One-stop breast clinic for symptomatic women mainly referred from GP</td>
<td>Diagnostic case-control study. Patients selected from all patients referred to clinic from 1994 to 1996 from 27 to 55 years. Cases: all women within cohort that were proven to have cancer (n=240). Controls: randomly-selected sample of women within the cohort who did not have cancer, age-matched to the cancer cases. n=240</td>
<td>Majority. M used in women ≥30 years. U used in all women &lt;30 years and those women &gt;30 years with an abnormality noted on breast examination. Combined result was not systematically reported in publication.‡</td>
<td>Standard two view, with additional views if necessary using GE Senograph 600T FD HF &amp; 600T HF Reported by radiologist assigning a score on five-point scale: 1=no significant abnormality, 2=benign, 3=indeterminate, 4=suspicious, 5=malignant. Negative result = 1, 2; positive result = 3, 4, 5.</td>
<td>Targeted using Aloka with 7.5 MHz transducer; Apogee 800 with 11.5 MHz; SSA 250 with 7.5 MHz. Reported by radiologist assigning a score on five-point scale: 1=no significant abnormality, 2=benign, 3=indeterminate, 4=suspicious, 5=malignant. Negative result = 1, 2; positive result = 3, 4, 5.</td>
<td></td>
</tr>
<tr>
<td>Study</td>
<td>Country</td>
<td>Setting</td>
<td>Patient Cohort</td>
<td>Mammography and ultrasound results reported for same patients?</td>
<td>Approach for combined result</td>
<td>Nature of mammography</td>
<td>Categorisation of positive and negative treatment of equivocal results</td>
</tr>
<tr>
<td>------------------</td>
<td>---------------</td>
<td>----------------------------------</td>
<td>-------------------------------------------------------------------------------</td>
<td>----------------------------------------------------------------</td>
<td>-----------------------------</td>
<td>------------------------</td>
<td>-----------------------------------------------------------------------</td>
</tr>
<tr>
<td>Shetty 2003²³</td>
<td>United States</td>
<td>Two breast clinics within hospital. Patients were referred mainly from gynaecologists.</td>
<td>Consecutive patients with palpable abnormalities who underwent both M and U between September 1999 and May 2000. Retrospective study enrolling 388 patients aged between 19 and 87 years.</td>
<td>Yes. Method not reported for combined result.</td>
<td>Straight mediolateral views plus standard mediolateral-oblique/intermedial-oblique and craniocaudal views using Mammoplus; Senographe DMR; Performa; Contour. Additional views if required. Reported by radiologist assigning a result of negative, benign or suspicious. Negative result = negative or benign; positive result = suspicious.</td>
<td>Targeted using GE LOGIQ 700MR, Acuson 128XP, Aloka SSD-1700 (all with a 7 MHz or higher transducer). Reported by physician assigning a result of negative, benign or suspicious. Negative result = negative or benign; positive result = suspicious.</td>
<td></td>
</tr>
<tr>
<td>Zonderland 1999²⁴</td>
<td>The Netherlands</td>
<td>Outpatient clinic</td>
<td>Consecutive patients referred to clinic who underwent M (n=4,728) of which a subset went on to have U (n=1,103) Prospective study enrolling women between the ages of 30 and 93 years.</td>
<td>Yes. Combined result was &quot;based on the results of ultrasound in addition to and in accordance with findings already obtained at mammography&quot;</td>
<td>Standard craniocaudal and mediolateral oblique views using Bennett Contour, Mamomat 3000 with additional views if required. Reported by radiologist assigning a score on a five-point scale: 1=benign, 2=probably benign, 3=equivocal, 4=probably malignant, 5=malignant. Negative result = 1, 2; positive result = 3, 4, 5.</td>
<td>Implies that targeted US of lesions undertaken, but not clearly reported. Used Aloka 650 CL with 10 MHz transducer; Aloka 280 with 7.5 MHz transducer. Reported by radiologist assigning a score on a five-point scale: 1=benign, 2=probably benign, 3=equivocal, 4=probably malignant, 5=malignant. Negative result = 1, 2; positive result = 3, 4, 5.</td>
<td></td>
</tr>
</tbody>
</table>

* Abbreviations: C, clinical examination; GP, general practitioner; M, mammography; U, ultrasound. † Data for the same patients were reported in an earlier paper.²⁷ ‡ Limited combined results were reported in the text. Table E2 Study quality for included studies.*
<table>
<thead>
<tr>
<th>Study</th>
<th>Reference standard adequate?</th>
<th>Test and reference standard measured independent of each other?</th>
<th>Were all patients assessed by the reference standard?</th>
<th>Test diagnosis made independent of other clinical information?</th>
<th>Reference standard measured before treatment started?</th>
<th>M result determined independent of U result?</th>
<th>U result determined independent of M result?</th>
<th>Were M and U measured at similar point in time?</th>
<th>Consecutive patients with breast symptoms presenting for imaging?</th>
<th>Was a representative spectrum of disease captured? Were patients with more advanced or recurrent disease included? Was the disease prevalence indicative of the target population?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flobbe 2003†</td>
<td>Adequate. Histology result or active 14-month follow-up of national pathology database that captures all patients.</td>
<td>Yes, but not vice versa.</td>
<td>Yes, through either histology or follow-up.</td>
<td>No</td>
<td>No</td>
<td>Probably</td>
<td>Yes</td>
<td>No. M was always performed first and result known at time of U</td>
<td>Yes, same visit</td>
<td>Yes</td>
</tr>
<tr>
<td>Houssami 2003†</td>
<td>Adequate. Cases: histology; controls: active 24-month follow-up of state cancer database.</td>
<td>Yes, but not vice versa.</td>
<td>Yes, through either histology or follow-up.</td>
<td>NR</td>
<td>NR</td>
<td>Probably</td>
<td>Yes</td>
<td>Yes. All imaging re-assessed in a blinded fashion.</td>
<td>Yes, same visit</td>
<td>Cases: yes; controls: no, a random sample of 240 from consecutive cohort of 13,583.</td>
</tr>
</tbody>
</table>
Table E2  Study quality for included studies. (continued)*

<table>
<thead>
<tr>
<th>Study</th>
<th>Reference standard adequate?</th>
<th>Test and reference standard measured independent of each other?</th>
<th>Were all patients assessed by the reference standard?</th>
<th>Test diagnosis made independent of other clinical information?</th>
<th>Reference standard measured before treatment started?</th>
<th>M result determined independent of U result?</th>
<th>U result determined independent of M result?</th>
<th>Were M and U measured at similar point in time?</th>
<th>Consecutive patients with breast symptoms presenting for imaging?</th>
<th>Was a representative spectrum of disease captured? Were patients with more advanced or recurrent disease included? Was the disease prevalence indicative of the target population?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shetty 200323</td>
<td>Not reported, although an earlier paper states that patients with a negative combined diagnosis were followed for 24 months.</td>
<td>Yes, but not vice versa.</td>
<td>Yes, but not vice versa.</td>
<td>NR</td>
<td>NR</td>
<td>NR</td>
<td>Probably</td>
<td>Yes</td>
<td>No</td>
<td>Yes, same visit</td>
</tr>
<tr>
<td>Zonderland 199924</td>
<td>Adequate. Histology, follow-up M after 12 months or active 18 month follow-up of national pathology database that captures all patients.</td>
<td>Yes, but not vice versa.</td>
<td>Yes, but not vice versa.</td>
<td>Yes, through either histology or follow-up.</td>
<td>NR</td>
<td>NR</td>
<td>Probably</td>
<td>Yes</td>
<td>No</td>
<td>No, performed after M by same radiologist</td>
</tr>
</tbody>
</table>

* Abbreviations: C, clinical examination; GP, general practitioner; M, mammography; NR, not reported; U, ultrasound.
† Data for the same patients were reported in an earlier paper.27