Comparative Effectiveness Review

Number 2

Effectiveness of Noninvasive Diagnostic Tests for Breast Abnormalities

AHRQ
Agency for Healthcare Research and Quality
Advancing Excellence in Health Care • www.ahrq.gov

This report is based on research conducted by the ECRI Evidence-based Practice Center (EPC) under contract to the Agency for Healthcare Research and Quality (AHRQ), Rockville, MD (Contract No. 290-02-0019). The findings and conclusions in this document are those of the authors, who are responsible for its contents; the findings and conclusions do not necessarily represent the views of AHRQ. Therefore, no statement in this report should be construed as an official position of AHRQ or the U.S. Department of Health and Human Services.

Comparative Effectiveness Review

Number 2

Effectiveness of Noninvasive Diagnostic Tests for Breast Abnormalities

Prepared for:
Agency for Healthcare Research and Quality
U.S. Department of Health and Human Services
540 Gaither Road
Rockville, MD 20850
www.ahrq.gov

Contract No. 290-02-0019

Prepared by:
ECRI Evidence-based Practice Center
Plymouth Meeting, Pennsylvania

Investigators
Wendy Bruening, PhD
Jason Launders, MSc
Nathan Pinkney, BS, RDMS
Harvey Kostinsky, MS
Karen Schoelles, MD, SM
Charles Turkelson, PhD

AHRQ Publication No. 06-EHC005-EF
February 2006

Suggested citation:
Bruening W, Launders J, Pinkney N, Kostinsky H, Schoelles K, Turkelson C. Effectiveness of Noninvasive Diagnostic Tests for Breast Abnormalities. Comparative Effectiveness Review No. 2. (Prepared by ECRI Evidence-based Practice Center under Contract No. 290-02-0019.) Rockville, MD: Agency for Healthcare Research and Quality. February 2006. Available at: www.effectivehealthcare.ahrq.gov/reports/final.cfm.

Preface

The Agency for Healthcare Research and Quality (AHRQ) conducts the Effective Health Care Program as part of its mission to organize knowledge and make it available to inform decisions about health care. As part of the Medicare Prescription Drug, Improvement, and Modernization Act of 2003, Congress directed AHRQ to conduct and support research on the comparative outcomes, clinical effectiveness, and appropriateness of pharmaceuticals, devices, and health care services to meet the needs of Medicare, Medicaid, and the State Children's Health Insurance Program (SCHIP).

AHRQ has an already-established network of Evidence-based Practice Centers (EPCs) that produce Evidence Reports/Technology Assessments to assist public- and private-sector organizations in their efforts to improve the quality of health care. The EPCs now lend their expertise to the Effective Health Care Program by conducting Comparative Effectiveness Reviews of medications, devices, and other relevant interventions, including strategies for how these items and services can best be organized, managed, and delivered.

Systematic reviews are the building blocks underlying evidence-based practice; they focus attention on the strength and limits of evidence from research studies about the effectiveness and safety of a clinical intervention. In the context of developing recommendations for practice, systematic reviews are useful because they define the strengths and limits of the evidence, clarifying whether assertions about the value of the intervention are based on strong evidence from clinical studies. For more information about systematic reviews, see www.effectivehealthcare.ahrq.gov/purpose.

AHRQ expects that Comparative Effectiveness Reviews will be helpful not only to government programs but also to individual health plans, providers, and purchasers, and to the health care system as a whole. In addition, AHRQ is committed to presenting information in different formats so that the greatest range of decisionmakers possible (and that includes consumers who make decisions about their own and their family's health) can benefit from the evidence.

Work under this program is transparent and user driven. Please visit the Web site (www.effectivehealthcare.ahrq.gov) to see draft research questions and reports or to join an e-mail list to learn about new program products and opportunities for input.

Acknowledgments

We would like to acknowledge with appreciation the following members of the Technical Expert Panel for their advice and consultation to the Evidence-based Practice Center during preparation of this report.

Technical Expert Panel

Robyn L. Birdwell, MD
Director, Breast Imaging Section
Brigham and Women's Hospital
Department of Radiology
Boston, MA

Robert D. Rosenberg, MD
Professor of Radiology
Chief, Diagnostic Imaging
University of New Mexico/
Albuquerque, NM

M. Carolina Hinestrosa, MA, MPH
Executive Vice President for Programs and Planning
National Breast Cancer Coalition
Washington, DC

EPC Program Director

We would also like to acknowledge with appreciation the EPC Program Director during the conduct of this project.

Kenneth S. Fink, MD, MGA, MPH
Director
Evidence-based Practice Center Program
Center for Outcomes and Evidence
Agency for Healthcare Research and Quality
Rockville, MD

AHRQ Contacts

Beth A. Collins Sharp, PhD, RN
Acting Director
Evidence-based Practice Center Program
Center for Outcomes and Evidence
Agency for Healthcare Research and Quality
Rockville, MD

William Lawrence, MD, MS
Task Order Officer
Evidence-based Practice Center Program
Center for Outcomes and Evidence
Agency for Healthcare Research and Quality
Rockville, MD

Contents

Tables

Figures

Executive Summary

Background

Breast cancer is the second most common malignancy in women. The American Cancer Society estimated that in the United States in 2005, 212,930 women would be newly diagnosed as having breast cancer and there would be 40,870 deaths due to this disease. Because early breast cancer is asymptomatic, the only way to detect it is through screening. Mammography is a widely accepted method for breast cancer screening. As a screening test, mammography is used to rule out cancer by missing very few cases of cancer--i.e., by having a low false negative rate. As a result, most women who have an abnormal mammogram do not have cancer.

Because an abnormal screening mammogram requires a diagnostic test to confirm whether cancer is present, many women who do not have cancer will undergo diagnostic tests. Typically, suspicious lesions are evaluated with tissue biopsy, either by excision or by needle sampling. If a noninvasive diagnostic test were available that could accurately exclude malignancy, many women with an abnormal mammogram who do not have cancer could avoid biopsy. However, such a test must be sufficiently accurate not to miss cancer in those women who have it. Positron emission tomography (PET) scanning, scintimammography, magnetic resonance imaging (MRI), and ultrasonography (US) have been proposed for this purpose, yet the accuracy of these noninvasive diagnostic technologies in excluding breast cancer in women at average risk remains unclear.

An ideal diagnostic test to evaluate breast abnormalities found by mammography or breast examination would distinguish women who need to have a biopsy from those who can safely avoid one. A woman who has a negative test result should be very confident that she does not have breast cancer before deciding to forgo a biopsy. To help patients, policymakers, and clinicians determine whether these noninvasive tests are sufficiently accurate to be appropriate for evaluation of women with an abnormal mammogram or exam finding, this report summarizes available data on the performance of these tests in the evaluation of women presenting with breast abnormalities that suggest the possibility of breast cancer. The report addresses the following questions:

1. What are the sensitivity and specificity of the tests for diagnosis of breast cancer in women presenting with an abnormal mammogram or a palpable breast abnormality?

2. For women with relevant demographic risk factors (e.g., age, family history) and clinical risk factors (e.g., Breast Imaging Reporting and Data System [BIRADS] status or morphologic characteristics of the lesion), what are the positive and negative predictive values of the above diagnostic tests?

3. Are there other factors that affect the accuracy or acceptability of the tests considered in Questions 1 and 2?

Conclusions

- A total of 81 studies met inclusion criteria to evaluate the accuracy of MRI, PET, scintimammography, or US for the diagnosis of breast cancer in women. The findings of accuracy for these tests, summarized as sensitivity, specificity, and negative likelihood ratios, are summarized in Table A. Although all of the technologies evaluated could reduce the need for biopsy in women with an abnormal mammogram who do not have cancer, each would miss some cancers.

- To place the tests' accuracy information into perspective, an average woman in the United States who has an abnormal mammogram requiring a biopsy for evaluation has approximately a 20-percent risk of cancer. For women at this *average* level of risk of cancer after an abnormal mammogram, based upon the tests' negative likelihood ratios:

    - For every 1,000 women who had a negative PET scan, about 924 women would have avoided an unnecessary biopsy, but 76 women would have missed cancers.

    - For every 1,000 women who had a negative scintimammogram, about 907 women would have avoided an unnecessary biopsy, but 93 women would have missed cancers. (These numbers are for nonpalpable lesions only; numbers could not be calculated for all lesions.)

    - For every 1,000 women who had a negative MRI, about 962 women would have avoided an unnecessary biopsy, but 38 women would have missed cancers.

    - For every 1,000 women who had a negative US, about 950 women would have avoided an unnecessary biopsy, but 50 women would have missed cancers.

- An individual woman's risk of breast cancer in the face of a suspicious finding on mammogram or clinical examination may vary widely from the average; the woman and her health care provider should discuss the extent of cancer risk. In general, the higher a woman's risk of cancer is before undergoing a noninvasive test, the higher is the risk that she has cancer even if the test is negative.

- If a less than 2-percent risk of having breast cancer with a negative diagnostic test is considered an acceptable level of risk for a diagnostic test to reliably preclude biopsy, none of these tests was sufficiently accurate to replace biopsy for women at average risk of breast cancer.

- Based on results for only nonpalpable lesions (usually detected by mammography), data were insufficient to estimate the accuracy of PET scanning, MRI, or ultrasound. Scintimammography was not sufficiently accurate to avoid biopsy in women at average

risk as judged by the acceptability standard of less than a 2-percent risk of breast cancer with a negative diagnostic test.

- Based on results for only palpable lesions, data were insufficient to estimate the accuracy of PET scanning, MRI, ultrasound, and scintimammography.

- The evidence supporting our conclusions permits us to be moderately confident that publication of future studies will not overturn our findings. Flaws in the evidence base include incomplete reporting of study design and patient characteristics, and insufficient numbers of studies reporting data for particular subgroups of patients; these flaws are responsible for ranking our confidence in our overall conclusion as moderate rather than strong.

Table A. Summary of Major Findings on Noninvasive Diagnostic Tests for Breast Abnormalities

Category	Test characteristic	PET scanning	Scinti-mammography	MRI	Ultrasound
Suspicious lesions in general	Sensitivity	82.2%	NC	92.5%	86.1%
	Specificity	78.3%	NC	72.4%	66.4%
	Negative likelihood ratio	0.33	NC	0.16	0.21
	Negative predictive value at 20% prevalence	92.4%	NC	96.2%	95.0%
	Stability of estimates	Low	Unexplained heterogeneity	Moderate	Moderate
Nonpalpable lesions	Sensitivity	NC	68.7%	NC	NC
	Specificity	NC	84.8%	NC	NC
	Negative likelihood ratio	NC	0.41	NC	NC
	Negative predictive value at 20% prevalence	NC	90.7%	NC	NC
	Stability of estimates	Insufficient evidence	Moderate	Insufficient evidence	Insufficient evidence
Palpable lesions	Sensitivity	NC	NC	NC	NC
	Specificity	NC	NC	NC	NC
	Negative likelihood ratio	NC	NC	NC	NC
	Negative predictive value at 20% prevalence	NC	NC	NC	NC
	Stability of estimates	Insufficient evidence	Insufficient evidence	Insufficient evidence	Unexplained heterogeneity

Abbreviations: PET = positron emission tomography. MRI = magnetic resonance imaging. NC= not calculated.
Notes: Sensitivity is the probability that a test is positive in those with the disease. Specificity is the probability that a test is negative in those without the disease. In this table, sensitivity is the mean threshold sensitivity reported in the included studies. The sensitivity threshold is the degree of abnormality that prompts a recommendation for biopsy. The corresponding specificity was determined using the Summary Receiver Operating Characteristic (SROC) curve. The SROC curve (in which the true positive rate is given on the y-axis and the false positive rate on the x-axis) depicts the relationship between sensitivity and specificity, illustrating that a change in the sensitivity threshold of a test inevitably affects the specificity of the test. Negative likelihood ratio is the ratio of the probability of a negative test in women with cancer to the probability of a negative test in women without cancer; based on the risk of having the disease prior to the test, it is used to calculate the risk of having the disease despite a negative test result. The negative likelihood ratios in this table were calculated using fixed-effects meta-analytic pooling. The negative likelihood ratio that can be calculated from the sensitivity and specificity reported in the table differs slightly from the summary negative likelihood ratio obtained by meta-analytic pooling; however, these values are not statistically different. Negative predictive value is the probability of not having cancer in women with a negative test result. Negative predictive value was calculated using the summary negative likelihood ratio.

Remaining Issues

- There was insufficient evidence to estimate how accurate these tests are in women whose mammogram indicates a lesion that probably is benign. Because these noninvasive tests are most likely to be used to evaluate such women, this is a major shortcoming of the current literature.

- A limitation of the available studies is the extremely high prevalence of breast cancer in the patients enrolled in them. This limitation reduces confidence that the results apply to all women who have suspicious mammograms.

- Analyses of benefits and harms using data from studies that enroll women more representative of the population of women who have suspicious mammograms would directly address the question of whether women benefit overall from being evaluated by noninvasive imaging methods.

Introduction

Since the advent of mammography and clinical breast examination, many asymptomatic women have had an abnormal finding. Typically, suspicious lesions are evaluated with tissue biopsy, either by excision or by needle sampling. However, only a low percentage of women undergoing biopsy actually have cancer, suggesting that many of them could avoid biopsy if a non-invasive diagnostic test were available that could, with high sensitivity, rule out malignancy. Several technologies have been proposed for this purpose, yet the outcomes of using these non-invasive diagnostic technologies remains unclear. The ultimate purpose of this evaluation is to help patients, policymakers, and clinicians determine when it is appropriate to use these non-invasive technologies.

Background

Breast cancer is the second most common malignancy of women.[1] The American Cancer Society estimates that in the U.S. in 2005, 212,930 women will be newly diagnosed as having breast cancer, and there will be 40,870 deaths due to this disease.[2] White women are more likely to develop breast cancer than women of other ethnic/racial groups, but black women are more likely to die from breast cancer.[3] Asian and third-world countries have much lower prevalences of breast cancer than does the U.S., U.K., and western Europe.[4] Breast cancer incidence was stable in the United States through the 1990s, but deaths due to breast cancer have been declining since 1995.[5] Forty to sixty percent of patients diagnosed with breast cancer will ultimately die from the disease.[6]

The majority of breast cancers develop from the epithelial lining of the milk ducts or lobules. Initial detection of breast cancer is usually the result of lumps noticed upon physical examination or areas of abnormal density identified by x-ray screening mammography. Survival rates depend on the stage of disease at diagnosis. At stage 0 (carcinoma *in situ*) the five-year survival rate is 100%. Five-year survival rates for women with stage IV (cancer has spread beyond the breast) are only 16%.[7] These observations suggest that breast cancer mortality rates can be significantly reduced by identifying cancers at earlier stages. Because early breast cancer is asymptomatic, the only way to detect it is through population-wide screening. Mammography is a widely accepted method for breast cancer screening.[8,9]

Mammography uses x-rays to examine the breast for calcifications, masses, or other abnormal structures. Currently most professional organizations recommend that women older than fifty years of age receive a yearly mammogram.[10] Some professional organizations recommend that routine breast cancer screening begin earlier, though mammography screening is less effective in younger women.[8]

The American College of Radiology has created a standardized system for reporting the results of mammography, the Breast Imaging Reporting and Data System (BIRADS).[11-13] There are seven categories of assessment and recommendation:

0. Assessment is incomplete, and additional imaging evaluation is needed.

1. Negative. There is no appreciable abnormality to report.

2. Benign finding. Benign finding such as benign calcifications, intramammary lymph nodes and calcified fibroadenomas.

3. Probably benign finding. An abnormality that has a high probability of being benign.

4. Suspicious abnormality. Biopsy should be considered.

5. Highly suggestive of malignancy. Biopsy is very strongly recommended.

6. Confirmed diagnosis of malignancy.

Often, a woman receives a biopsy after discovery of a suspicious lesion by mammography or physical examination (BIRADS score 4 or 5). However, only a low percentage of women undergoing biopsy actually have cancer-- the positive predictive value of mammography has been estimated to be less than 30%, possibly as low as 15%.[14,15] If a non-invasive diagnostic test were available that could accurately rule out malignancy, many women could safely avoid biopsy. Several technologies—magnetic resonance imaging (MRI), ultrasonography (US), scintimammography, positron emission tomography (PET) scanning—have been proposed for this purpose.

Mammography often leads to identification of a "probably benign" lesion (BIRADS score 3), or an uninformative mammography (BIRADS score 0). Clinicians may be reluctant to refer such a patient for a biopsy; they may also be reluctant to do nothing. Often such patients are referred for frequent repeat mammography examinations. If an accurate non-invasive diagnostic test were available to examine these women, many women could avoid these repeat mammography exams, with their attendant discomfort, inconvenience, x-ray exposure, and emotional distress.

The ultimate goal of this evaluation is to determine when it is clinically appropriate to use these non-invasive technologies to evaluate breast abnormalities. In order for clinicians to decide upon "clinical appropriateness" for any particular patient, an accurate estimate of the negative and positive predictive values of the non-invasive tests for women with a variety of demographic and clinical risk factors must be available. Because women with a previous history of breast cancer and women known to carry BRCA1 and BRCA2 mutations have a very different risk profile than the rest of the population, we will not evaluate the use of non-invasive technologies for such women in this report. Instead, we will focus on the use of non-invasive imaging technology for women considered to be at "normal" risk of breast cancer who present with an abnormal finding by mammography or physical examination. We will also examine the influence of age; the morphological characteristics of the lesion; BIRADS status; the presence of calcifications, and other key clinical risk factors.

Methods

Scope and Key Questions

This report addresses three questions. We address the first two employing a systematic review, and the third as a narrative overview. The questions we address are:

Key Question 1. For the following diagnostic tests as applied to the breast (positron emission tomography (PET) scanning, scintimammography (SC), magnetic resonance imaging (MRI), and ultrasonography (US)) what are the sensitivity and specificity of the tests for diagnosis of breast cancer in women presenting with:

a) An abnormal mammogram, overall and by BIRADS classification or other relevant clinical classification (e.g., presence or absence of calcification, well circumscribed lesions, etc.)

b) A palpable breast abnormality

c) What percentage of women in the studies in this question were age 65 or older, and do sensitivity and specificity vary by older vs. younger than age 65?

Key Question 2. For women with relevant demographic risk factors (e.g., age, family history) and clinical risk factors (e.g., BIRADS status or morphologic characteristics of the lesion), what are the positive and negative predictive values of the above diagnostic tests?

Key Question 3. Are there other factors that affect the accuracy or acceptability of the tests considered in Questions 1 and 2? Our answer to this question is a narrative overview of device- and operator- specific factors that may affect diagnostic accuracy, such as timing of administration of contrast agents and scan, interpretation of images, and settings of the devices. We also review potentials for harm from the test, including issues such as radiation exposure or excessive discomfort associated with the test. To address this question, we used information from review articles, experts, and other sources.

Other diagnostic and imaging technologies are outside the scope of this report, as are any other issues, outcomes, patient categories, or questions about breast cancer diagnosis not explicitly mentioned in the three Key Questions. Evaluations of women known to have susceptibility mutations in either the BRCA1 or BRCA2 genes are outside the scope of this report, as are women known to have a prior history of breast cancer. This report focuses on specific technologies, specific uses of these technologies, and specific patient populations. The focus of this report does not imply that other uses of these technologies for other purposes and other patient populations are not important or valid.

Outcomes

The outcomes of interest in this report are diagnostic test characteristics. Diagnostic test characteristics are measures of how well the diagnostic test performs. In published studies,

the results of the relevant tests are determined to be correct or incorrect as measured by histopathological examination of a biopsy specimen. In other words, published studies use biopsy as a "gold standard" to determine whether the patient truly has or does not have breast cancer, and the performance of the test of interest is compared to that of the gold standard. Four important measures of a diagnostic test's performance can be calculated. These are the test's sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV). These diagnostic test characteristics are briefly explained below. For further details see Appendix E.

Sensitivity and specificity are properties of a test that are useful when deciding whether to use a test. Sensitivity is the proportion of people with the disease who have a positive test for the disease. A test with high sensitivity will rarely misclassify people with the disease as not having the disease (the test has a low rate of false-negatives). Specificity is the proportion of people without the disease who have a negative test. A test with high specificity will rarely misclassify people without the disease as diseased (a low rate of false-positives).

There is a tradeoff between the sensitivity and specificity of a test. The test threshold, above which patients are sent for biopsy, is essentially an arbitrary decision (in that it is not biologically fixed, but can assume any value that meets the needs of patients and clinicians). As a consequence, for any given image of a breast lesion, sensitivity can only be increased by decreasing specificity. For example, if a clinician decides that only patients with extremely abnormal images will be sent for biopsy, the test will become extremely specific but not very sensitive. One consequence of this is that the clinician will falsely diagnose many patients as *not* having breast cancer. On the other hand, a clinician may not want to miss any cases of breast cancer. In this case, the clinician decides that patients with borderline abnormal images will be sent for biopsy, so the test will become more sensitive but less specific. One consequence of this is that the clinician will send many patients who do not have breast cancer for an unnecessary biopsy. A graphical method of diagnostic test analysis that illustrates the tradeoff between sensitivity and specificity at each threshold is called a receiver operating characteristic curve (ROC curve).

Two other important measures of diagnostic test performance are the positive and negative predictive values. The positive predictive value of a test is the probability of a patient having the disease following a positive test result. The negative predictive value is the probability of a patient not having the disease following a negative test result. Unlike sensitivity and specificity, predictive values are influenced by the prevalence of the disease in the population of patients being tested.

Because sensitivity and specificity are difficult to directly apply to clinical situations, and predictive values vary markedly as a function of disease prevalence (i.e., may be different for each patient subpopulation) a combined measure of diagnostic performance, the likelihood ratio, is a clinically useful diagnostic test performance measure.[16,17] Negative likelihood ratios measure the ability of the test to accurately "rule out" disease, and positive likelihood ratios measure the ability of the test to accurately detect disease. Likelihood ratios can be directly used in Bayes' theorem to calculate posttest odds of having a disease from the pretest suspicion of the patient's odds of having that disease. In general, a positive likelihood ratio greater than 10, or a negative likelihood ratio less than 0.1, may have a very substantial impact on clinical decision-making through meaningful revision of disease probability (commonly referred to as 'clinically useful'). A positive likelihood ratio of less than 2 or a negative likelihood ratio of greater than 0.5, has a clinically insignificant impact on decision-making (commonly referred to as 'not clinically

useful'). Likelihood ratios in between these two extremes may or may not have an impact on clinical decision-making, depending upon the context; likelihood ratios in this in-between category are commonly referred to as 'possibly clinically useful'.[18]

Determining whether a 'possibly clinically useful' likelihood ratio has a significant impact on decision-making requires consideration of each clinical context. For example, a patient presents with symptoms that lead the clinician to believe there is a 50% chance of a bacterial infection of the heart valves. The clinician orders a diagnostic test that has a negative likelihood ratio of 0.2. The test comes back negative for bacterial infection of the heart valves. Applying Bayes' theorem, the physician now believes there is a 17% chance of a bacterial infection of the heart valves. However, the physician may decide to treat with antibiotics anyway, because there is little chance of side effects from treatment, and very unpleasant possibilities if an infection of the heart valves is left untreated. However, what if the suspected disease was a brain tumor? It is unlikely the physician would proceed with surgery on only a 17% suspicion. In complicated clinical situations, a full cost-benefit decision analysis may be necessary to decide whether a diagnostic test has sufficient accuracy (estimated through likelihood ratios) to guide clinical decision-making.[19]

To help readers put the findings of this report into perspective, we have listed some published negative and positive likelihood ratios for common diagnostic tests relevant to breast cancer in Table 1.

Table 1. Published likelihood ratios for common diagnostic tests relevant to breast cancer

Diagnostic test	Positive likelihood ratio	Negative likelihood ratio
Sentinel lymph node biopsy to rule out metastatic disease in the axilla		0.086[20]
Large core needle biopsy to rule out or rule in breast cancer in a suspicious lesion	16.2[21]	0.03[21]
Fine needle aspiration, performed by an experienced cytologist, to rule out breast cancer in a suspicious lesion		0.02 to 0.11[22]
Diagnostic mammography of a palpable lump, to rule out or rule in breast cancer	5.6[22]	0.15[22]
Screening mammography of asymptomatic women to detect breast cancer	2.2 to 2200[22,23]	

Evidence Base

Literature Searches

Details of our literature searches, which included searches of 11 electronic databases, hand searches of the bibliographies of all retrieved articles, and searches of the gray literature, are presented in Appendix A. Literature Search Strategies.

Inclusion/Exclusion Criteria for Questions 1 and 2

We used the following criteria to determine which studies would be included in our analysis:

1. Study must evaluate the effectiveness of at least one of the non-invasive diagnostic technologies that are the focus of this report (PET scanning, scintimammography, MRI, ultrasound).
   Other diagnostic technologies are outside the scope of this report.

2. Used current generation scanners and protocols.
   Studies of outdated technology (defined as studies published before 1991) and experimental technology are not relevant to current clinical practice.

3. For studies of PET scanning, the study must have used 18-fluorodeoxyglucose as the tracer.
   18-fluorodeoxyglucose is the standard tracer used in clinical practice. Studies of other tracers are outside the scope of this report.

4. For studies of scintimammography, the study must have used ^{99m}Tc-sestamibi as the tracer.
   This is the standard tracer used in clinical practice. Studies of other tracers are outside the scope of this report.

5. For studies of MRI scanning, the study must have used a dedicated breast coil and a gadolinium-based contrast agent.
   Use of a dedicated breast coil and a contrast agent are standard clinical practice. Studies of other types of contrast agents are outside the scope of this report.

6. For studies of ultrasound, the study must have used gray-scale B-mode ultrasound with a transducer of 7 MHz or higher resolution.
   Studies of tissue harmonics, color Doppler scanning, power Doppler scanning, or other methods are not standard clinical practice, and are outside the scope of this report.

7. Study must report test sensitivity, specificity, negative or positive predictive values, or sufficient data to calculate these measures of diagnostic test performance
   Other outcomes are beyond the scope of this report.

8. Study must be published in English.
   Translation costs prohibit the inclusion of studies published in other languages.

9. Study must be published as a peer-reviewed full article. Meeting abstracts will not be included.
   Published meeting abstracts have not been peer-reviewed and often do not include sufficient details about experimental methods to permit one to verify that the study was well designed.[24,25] In addition, it is not uncommon for abstracts that are published as part of conference proceedings to describe studies that are never published as full articles.[26-29]

10. Enrolled human subjects.
    Animal studies or studies of "phantoms" are outside the scope of the report.

11. The study must have enrolled 10 or more individuals.
    The results of small studies are typically more variable and less generalizable than those of larger studies.[30-32]

12. Study must not enroll individuals known to have mutations in BRCA1 or BRCA2, or individuals known to have a prior history of breast cancer, unless data from these populations are reported separately.
    Women with BRCA1 or BRCA2 mutations, and women with a prior history of breast cancer, are outside the scope of this report. If more than 4% of the study population consists of such individuals, and the data for each population are not reported separately, the study is excluded.

13. Study must enroll individuals found to have breast abnormalities by mammography or physical examination.
    Primary screening studies are outside the scope of this report.

14. Study must be prospective in design.

15. Study must be either a randomized controlled trial or a prospective diagnostic cohort study.
    Case-control studies, non-randomized controlled studies, and case reports were excluded.

16. In diagnostic cohort studies, 85% or more of the patients must have been evaluated with both the diagnostic test of interest and biopsy.
    Open surgical biopsy or core needle biopsy are both acceptable reference standards.[21] Fine needle aspiration followed by cytology is acceptable for evaluation of cystic lesions. For the purposes of this report, fine needle aspiration of solid lesions is not an acceptable reference standard.[33-36]

17. When several sequential reports from the same study center are available, only outcome data from the largest and most recent report were included. However, we used relevant data from earlier and smaller reports if the report presented pertinent data not presented in the larger, more recent report.

Data Abstraction

The following data were abstracted from the included trials: study design, details of imaging procedures, population characteristics (including sex, age, ethnicity), eligibility and exclusion criteria, information on thresholds used, and results for each outcome. Information abstracted from the included studies is presented in Appendix E. Evidence Tables.

Analysis

Strength and Stability of Evidence Assessment

We rated the strength and stability of the evidence using an algorithm that we developed. This algorithm provides systematic, reproducible, transparent, and *a priori* decision rules for rating the strength of the evidence. It extends the recommendations made in the AHRQ report, "Systems to Rate the Strength of Scientific Evidence," which concluded that the strength of evidence depends on the quality, quantity, and consistency of the available data.[37] The algorithm distinguishes between qualitative questions (Does it work?) and quantitative questions (How well does it work?) and, as shown in Table 2, assigns a separate rating of the evidence for these two kinds of questions. Evidence underpinning the answers to qualitative questions is rated according to its strength, and evidence underpinning the answers to quantitative questions is rated according to the stability of the evidence. Qualitative conclusions backed by strong evidence are less likely than weaker conclusions to be overturned by new evidence, and the quantitative estimates provided by answers to quantitative questions that are backed by stable estimates are less likely to exhibit much change upon the publication of new estimates. These definitions are similar to those proposed by the GRADE working group.[38] The algorithm itself is shown in Appendix B. Study Quality and Strength of Evidence Evaluation.

Table 2. Definitions of Strength and Stability of Evidence

Strength of Evidence Rating	Interpretation
Qualitative Conclusion	
Strong Evidence	Evidence supporting the qualitative conclusion is convincing. It is highly unlikely that new evidence will lead to a change in this conclusion.
Moderate Evidence	Evidence supporting the qualitative conclusion is somewhat convincing. There is a small chance that new evidence will overturn or strengthen our conclusion. Regular monitoring of the relevant literature is recommended at this time.
Weak Evidence	Although some evidence supports the qualitative conclusion, this evidence is tentative and perishable. There is a reasonable chance that new evidence will overturn or strengthen our conclusions. Frequent monitoring of the relevant literature is recommended at this time.
Unacceptably Weak Evidence	The available evidence that exists is not of sufficient strength to warrant drawing an evidence-based conclusion from it. Frequent monitoring of the relevant literature is recommended at this time.
Quantitative Conclusion	
High Stability	The estimate of diagnostic test performance in the conclusion is stable. It is highly unlikely that the magnitude of this estimate will substantially change as a result of the publication of new evidence.
Moderate Stability	The estimate of diagnostic test performance in the conclusion is somewhat stable. There is a small chance that the magnitude of this estimate will substantially change as a result of the publication of new evidence. Regular monitoring of the relevant literature is recommended at this time.
Low Stability	The estimate of diagnostic test performance in the conclusion is likely to be unstable. There is a reasonable chance that the magnitude of this estimate will substantially change as a result of the publication of new evidence. Frequent monitoring of the relevant literature is recommended at this time.
Unacceptably Low Stability	Estimates of the diagnostic test performance are too unstable to allow a quantitative conclusion to be drawn at this time. Frequent monitoring of the relevant literature is recommended.

In the context of diagnostics, answers to the qualitative questions of whether the diagnostic test works (i.e., whether it performs better than chance), are often not of particular interest. Answers to quantitative questions are typically of greater interest, and are provided by summary estimates of test performance. Where possible, we addressed semi-qualitative questions of whether the diagnostic test works sufficiently well to be clinically useful by direct examination of summary estimates of test performance; however, such questions can only be fully addressed by decision analyses, which are outside the scope of this report.

To aid in assessing the quality of each of the studies included in this assessment, we used a quality assessment instrument developed by ECRI, shown in Appendix B. Study Quality and Strength of Evidence Evaluation. This instrument examines different factors of diagnostic study design that have the potential to reduce the validity of the conclusions that can be drawn from a trial. In brief, the tool was designed so that a study attribute that, in theory, protects a study from bias receives a “Yes” response. If the study clearly does not contain that attribute it receives a “No” response. If poor reporting precludes assigning a “Yes” or “No” response for an attribute, then “NR” is recorded (NR = not reported).

To estimate the quality of an individual study, we computed a normalized score so that a perfect study received a score of 10, a study for which the answers to all items was "No" received a score of 0, and a study for which the answers to all questions was "NR" was 2.5.

We then classified the overall quality of the evidence base by taking the median quality score. Quality scores were converted to categories as shown below. The definitions for what constitutes low, moderate, or high quality evidence were determined *a priori* by a committee of four ECRI methodologists.

less than 5.0: Unacceptable (exclude study or entire evidence base)

less than 7.0: Low quality

less than 9.0: Medium quality

9.0 +: High quality

The consistency of the strength of evidence was addressed differently for quantitative and qualitative conclusions. The consistency of the evidence base for quantitative conclusions was measured with statistical tests of heterogeneity. We used the heterogeneity statistic I^2. Typically we use a threshold for I^2 of 0.5 because, according to Higgins and Thompson, this value represents moderate heterogeneity.[39,40] I^2 tends to be a more useful test than the Q statistic in most situations because of the low statistical power of Q. If the data were found to be consistent, a summary effect size was calculated; if not, meta-regression (described further below) was used to explore the reasons for heterogeneity.

We tested the consistency of the evidence base for semi-qualitative conclusions drawn from likelihood ratios by an informal vote-counting method. If the 95% confidence intervals of 80% or more of the effect sizes were within the "possibly clinically useful" range of likelihood ratios (2 or greater for positive likelihood ratios, 0.5 or less for negative likelihood ratios), then the conclusion of "clinically useful" was defined to be robust and consistent.

The quantity of the overall strength of evidence rating was addressed by measuring the stability of summary estimates. A stable summary estimate indicates that the accumulated body of evidence is large enough to have accurately measured the "true" effect size. The stability of summary estimates was tested with cumulative meta-analysis.[41,42] Studies were sequentially added in order of decreasing size into a meta-analysis, and summary estimates calculated for each step. We defined the summary estimate as stable if the 95% confidence intervals of the last three summary estimates calculated by cumulative meta-analysis were within 5% of the final summary estimate. The stability of summary estimates was only measured if the data were found to not be heterogeneous (consistent). If the data are not consistent (hetereogeneous), no summary estimate was calculated.

Data Synthesis

If more than two studies reported relevant data, we calculated and tested the diagnostic odds ratios from these studies for heterogeneity with the I^2 test described above.[39,40] If the data were

found to not be heterogeneous, we pooled the data. If data from five or more studies were available, meta-analysis of diagnostic test performance was performed by constructing summary receiver operating characteristic (SROC) curves that depict the true positive rate of the test on the y-axis and the false positive rate on the x-axis. We used the method of Moses and Littenberg to construct these curves.[43] The actual curves are shown in the body of the report. To summarize the information in a format that is commonly used in the literature, two points on each curve were selected and used to describe the SROC curves in tables and text; these two points are the sensitivity and specificity on the SROC curve at the mean reported sensitivity for the articles included in the meta-analysis, and at 95% sensitivity. The latter point was used as a standard reporting convention for SROC curves and does not imply that this is the actual sensitivity of the tests. If data from fewer than five studies were available, we did not construct a SROC, and instead directly pooled the sensitivity and specificity reported by the studies.[44] Directly pooling the diagnostic test characteristics in this fashion ignores any variability of the diagnostic threshold and may lead to an underestimation of diagnostic performance.[44] However, a SROC derived from fewer than five studies is unlikely to be an accurate representation of the test's performance.[43]

To aid in clinical interpretation, we also analyzed the data using fixed-effects meta-analytical pooling of likelihood ratios, as described by Stengel et al.[45] The summary negative likelihood ratios and Bayes theorem were used to calculate the post-test probability of disease and estimate the number of missed cases of cancer for women at the average risk of cancer of those being referred to a biopsy in the U.S. These results were included in the discussion and summary results to place the diagnostic accuracy of these diagnostic tests into a meaningful clinical context. We chose to use the summary negative likelihood ratios for these calculations rather than negative predictive values estimated from the SROC results because they were more directly applicable; the likelihood ratio produces a single result, already at the threshold most likely to be used in clinical settings. Conclusions drawn from the SROC results are essentially identical to the results drawn from the summary negative likelihood ratios; any apparent differences are within the range of error.

We used the odds ratio form of Bayes theorem to estimate the number of cancer cases missed if a diagnostic test is negative. This form of Bayes theorem, for the probability of cancer given that the test is negative, is expressed by the formula:

$$\text{Post} - \text{Test Odds} = \text{Pre} - \text{Test Odds} \times \text{LR} -$$

Where:

$$\text{Pre} - \text{Test Odds} = \frac{\text{Prevalence}}{1 - \text{Prevalence}}$$

and:

$$\textit{Posterior probability} = \frac{\textit{Post} - \textit{Test Odds}}{1 + \textit{Post} - \textit{Test Odds}}$$

In this case, the posterior probability represents the probability that a woman has cancer given that the diagnostic test was negative. We used a prevalence of cancer of 20% to represent the average probability of cancer in women undergoing a decision of biopsy or non-invasive test, based on Banks et al. reported prevalence of breast cancer of 11.6% to 24.4%, depending on age, for more than 100,000 unselected patients with positive mammograms.[46] The results have been

expressed in two different ways in the report: as an individual woman's chance of having disease after a negative test result (simply the post-test probability), and as a the number of missed cases of cancer. If 1,000 women test negative for cancer, and each woman's chance of having cancer is the post-test probability, the number of missed cases of cancer in this cohort of 1,000 women can be estimated by multiplying the post-test probability by 1,000. The posterior probability can readily be converted to a negative predictive value (NPV) by the formula:

$$NPV = 1 - Posterior\ \ probability.$$

We used meta-regression to attempt to explain heterogeneity when it was detected. To avoid an ecological fallacy, we included only study-level covariates in the meta-regression. Such covariates may have included (but not be limited to) study quality items from the study quality scale, and reported variations in imaging methods. All meta-regressions were performed using the permutation method by Higgins and Thompson.[47] This method counters the well-known Type I error inflation rate associated with the covariates entered in standard meta-regressions. Meta-regressions were performed using restricted maximum likelihood (REML) as computed by the Stata statistical software package.[48]

The third preliminary Key Question was addressed using a narrative review. This review was conducted by ECRI's Health Devices Group, which has conducted laboratory and hospital-based testing of medical devices for over 30 years. The health devices group consists of experts with experience in using and evaluating most of the technology used to diagnose and treat patients, including diagnostic imaging. This requires an understanding of the physical principles that govern a device, and how the device is used clinically. For most diagnostic imaging equipment, this is achieved by observing the equipment in use and interviewing routine users, including radiologists and technologists.

Peer Review

We requested peer review of the draft of this report from seven content or methodology experts and one patient advocacy organization. Their comments were reviewed, and where possible, incorporated into the final document. See Appendix D. Peer Reviewers for a list of the reviewers.

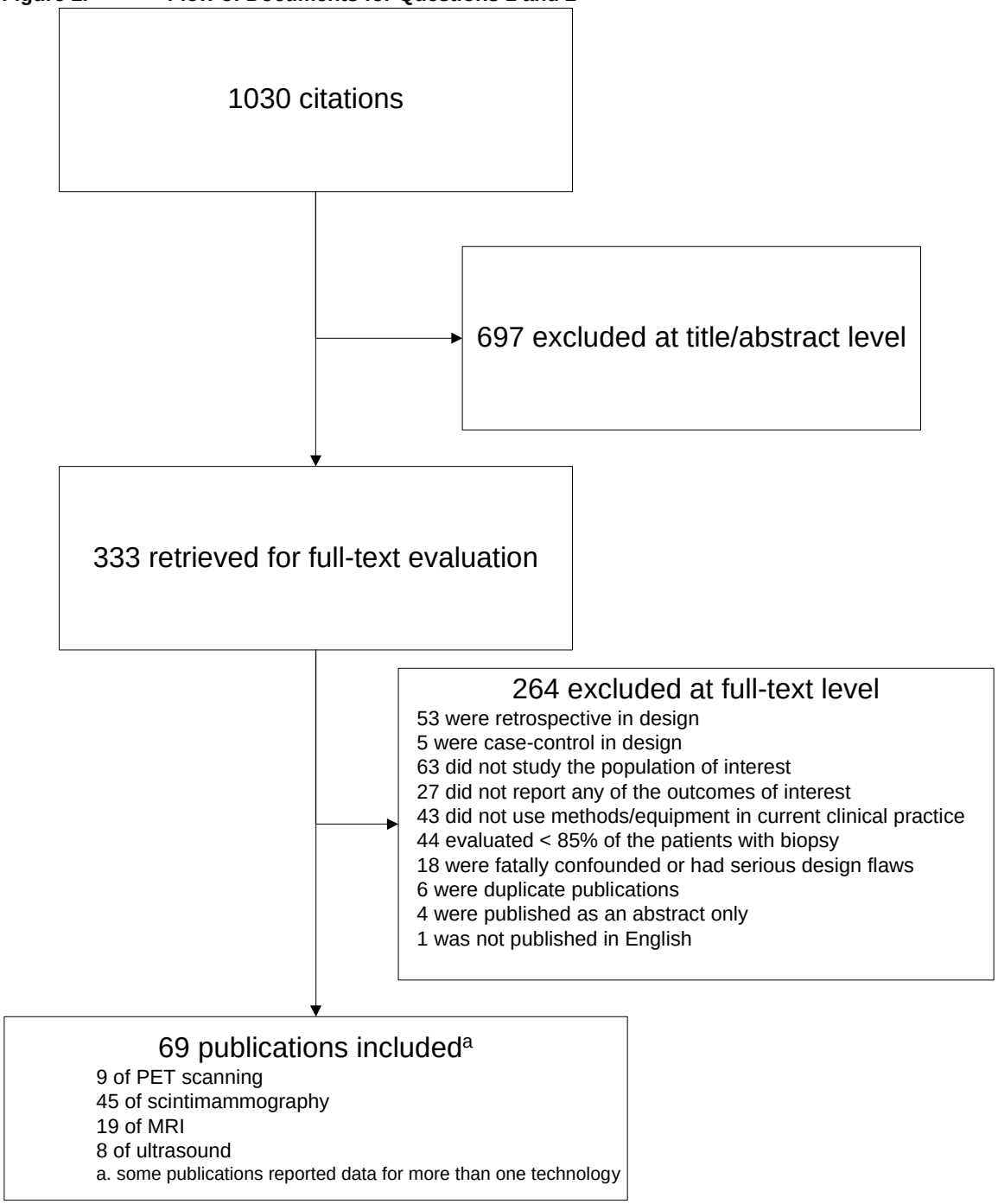
Results

Overview of Evidence Base for Questions 1 and 2

Our literature searches identified 1,030 potentially relevant articles. The processing of these documents is summarized in Figure 1. The titles and abstracts of these articles were examined. All studies (333) that potentially met the inclusion criteria for Questions 1 and 2 were retrieved. The most common reason for not retrieving an article was because it clearly did not address either of the questions of interest. After retrieval, the full text of the articles was examined. Two hundred sixty four of the 333 retrieved articles did not meet our inclusion criteria. The primary reason for exclusion (63 articles) was not studying women who presented with an abnormal finding by mammography or physical examination. The remaining articles were excluded for a variety of other reasons, shown in Figure 1. These articles and the reason(s) for their exclusion are listed in Appendix C. List of Studies Excluded from Questions 1 and 2.

Having excluded 264 articles, 69 articles remained. These articles are listed in Appendix E. Evidence Tables.

Figure 1. Flow of Documents for Questions 1 and 2



Summary of Results

Positron Emission Tomography Scanning

In positron emission tomography (PET) scanning, a small amount of radioactive glucose (18-fluoro-2-deoxyglucose: FDG) is injected into the bloodstream, and a gamma camera,

dedicated breast scanner, or whole-body scanner is used to generate images that highlight areas of high tracer uptake. Glucose accumulates in tissues of high growth rate or rapid metabolism (i.e., tumors). Whole-body scanners have a ring of detectors that surround the patient and image the entire body. Gamma cameras have only two detectors, one at each side of the patient, and image only a restricted portion of the body. Dedicated breast scanners have two detectors designed to image only the breasts.

Summary

Our analysis found that whole-body PET scanning was more accurate than gamma camera PET imaging for ruling out breast cancer. The strength of the evidence supporting this conclusion was weak, indicating a reasonable chance that publication of new evidence could overturn this conclusion. No studies of dedicated breast PET scanners met the inclusion criteria. At a fixed sensitivity of 95%, the specificity of whole-body PET scanning was only 46.7%. At the mean threshold of the included studies, the sensitivity of PET scanning was 82.2%, the specificity was 78.3%, and the negative predictive value (for a population with a prevalence of 70%) was 64.8%. The stability of the summary diagnostic odds ratio used to obtain these summary diagnostic test characteristics was low, indicating a reasonable chance that publication of new evidence could substantially change these estimates.

We found that the negative likelihood ratio for whole-body PET scanning of women referred for further evaluation of the breast was 0.33. This result means that if a woman with a suspicious lesion is diagnosed as cancer-free by PET scanning, her actual chance of having breast cancer drops from 20%[1] to 7.6%. There were no or insufficient data to come to any conclusions about the use of PET to evaluate any sub-populations of patients.

The authors of one systematic review published in 2001 concluded that the negative predictive value of PET was too low for routine use in ruling out breast cancer after detection of an abnormality. The results of our analysis concur with this conclusion.

The studies, patients, and analyses used to reach these conclusions are described in the following section. The performance of PET in evaluating women with suspicious breast lesions is summarized in Table 3. The estimates of summary diagnostic test characteristics are of low stability, indicating that there is a reasonable chance that the magnitude of the estimate could substantially change as a result of the publication of new evidence.

[1] A woman in the general screening population who has a positive finding on mammography and/or physical examination has an approximately 20% chance of having breast cancer.

Table 3. Summary Test Performance of PET

Patient subgroup	N studies	N lesions	Sensitivity	Specificity	Prevalence (Range)	PPV	NPV	+ LHR (95% CI)	- LHR (95% CI)
Suspicious breast lesion, whole-body scanning	6	186	At mean threshold 82.2% Stability of estimate: Low	At mean threshold 78.3% At 95% sensitivity 46.7% Stability of estimate: Low	70.4% (45.2% to 95.0%)	At mean threshold 90.0% At 95% sensitivity 80.9% Stability of estimate: Low	At mean threshold 64.8% At 95% sensitivity 79.7% Stability of estimate: Low	Evidence is unacceptably weak to support a conclusion	0.33 (95% CI 0.24 to 0.46) Stability of estimate: Low
Suspicious breast lesion, whole-body scanning vs. gamma camera imaging	1	30	No summary diagnostic test characteristics were calculated because the evidence base was unacceptably weak to support a quantitative conclusion.		86.7%	Whole-body PET scanning is more accurate for ruling out breast cancer than is gamma camera imaging. Strength of evidence: weak.			
Suspicious breast lesion, gamma camera	2	58	The body of evidence is unacceptably weak for evaluation of test performance.						
Palpable lesion	2	66	The body of evidence is unacceptably weak for evaluation of test performance.						
Palpable lesion, ≥1.0 cm	1	24	The body of evidence is unacceptably weak for evaluation of test performance.						
BIRADS 5	2	49	The body of evidence is unacceptably weak for evaluation of test performance.						
BIRADS 3, ≥1.0 cm	1	18	The body of evidence is unacceptably weak for evaluation of test performance.						
BIRADS 4-5, ≥1.0 cm	1	14	The body of evidence is unacceptably weak for evaluation of test performance.						
Lesions ≥1.0 cm	2	50	The body of evidence is unacceptably weak for evaluation of test performance.						
Lesions ≥1.5 cm	1	29	The body of evidence is unacceptably weak for evaluation of test performance.						
Patients <65 yrs of age: whole-body	2	58	The body of evidence is unacceptably weak for evaluation of test performance.						

Patient subgroup	N studies	N lesions	Sensitivity	Specificity	Prevalence (Range)	PPV	NPV	+ LHR (95% CI)	- LHR (95% CI)
Patients <65 yrs of age: gamma camera	2	55	The body of evidence is unacceptably weak for evaluation of test performance.						
Patients ≥65 years of age	1	19	The body of evidence is unacceptably weak for evaluation of test performance.						

PPV = positive predictive value
NPV = negative predictive value
LHR = likelihood ratio

Analysis and Results for Key Questions 1. and 2. Diagnostic Test Characteristics, Predictive Values, and Likelihood Ratios of PET

Other Published Technology Assessments

We identified one systematic review of PET use after mammography. This review, published by Sampson et al. in 2002, assessed the performance of PET in the differential diagnosis of benign and malignant lesions among patients with abnormal mammograms or a palpable breast mass.[49,50] The review included 13 articles published before March 2001.

Sampson et al. performed a meta-analysis using a random-effects model, and selected a point on the SROC that reflected test performance, with a sensitivity of 89% and a specificity of 80%. When the prevalence of malignancy was 50%, 40% of all patients would benefit by avoiding the harm of a biopsy with negative biopsy results. However, the negative predictive value was found to only be 88%. For a patient with a negative PET scan, the authors concluded that a 12% chance of a missed or delayed diagnosis of breast cancer is too high to make it worth the 88% chance of avoiding biopsy of a benign lesion.

Included Studies

Nine diagnostic cohort studies of 18-fluorodeoxyglucose PET scanning met our inclusion criteria. Characteristics of the studies and included patients are summarized in Table 28, in Appendix E. Evidence Tables. All but one of the studies used a whole-body PET scanner. The other (Holle et al. 1996[51]) used a gamma camera. The quality of all of the studies was moderate (median score of 7.9, range 7.4 to 8.8). The most common flaw in these studies was not reporting whether readers of tests were blinded to patient information or the results of other tests.

The included patients were incompletely described, with few details about demographics reported. The patients were primarily women with suspicious lesions, detected by physical exam, mammography, or ultrasound, who had been scheduled for biopsy, who presented at a time when the PET scanner was available. The number of patients enrolled in the studies was restricted due to scanner availability. Two of the studies excluded patients with lesions smaller than 1.0 cm (Brix et al.[52] and Crowe et al.[53]). Patients ranged in age from 20 to 86, and reported mean ages ranged from 48.3 to 58.4, suggesting that the patient populations studied are younger than the typical breast cancer population. Only three of the nine studies (Holle et al. 1996[51], Yutani et al. 1999[54], Yutani et al. 2000[55]) reported the percentage of patients 65 years of age or older: 31 out of 120 patients, 25.8%. Information about the percentage of patients 65 years of age or older was not reported for the remaining 156 patients (see Table 30 in Appendix E. Evidence Tables).

Patients Referred for Evaluation of Suspicious Breast Lesions

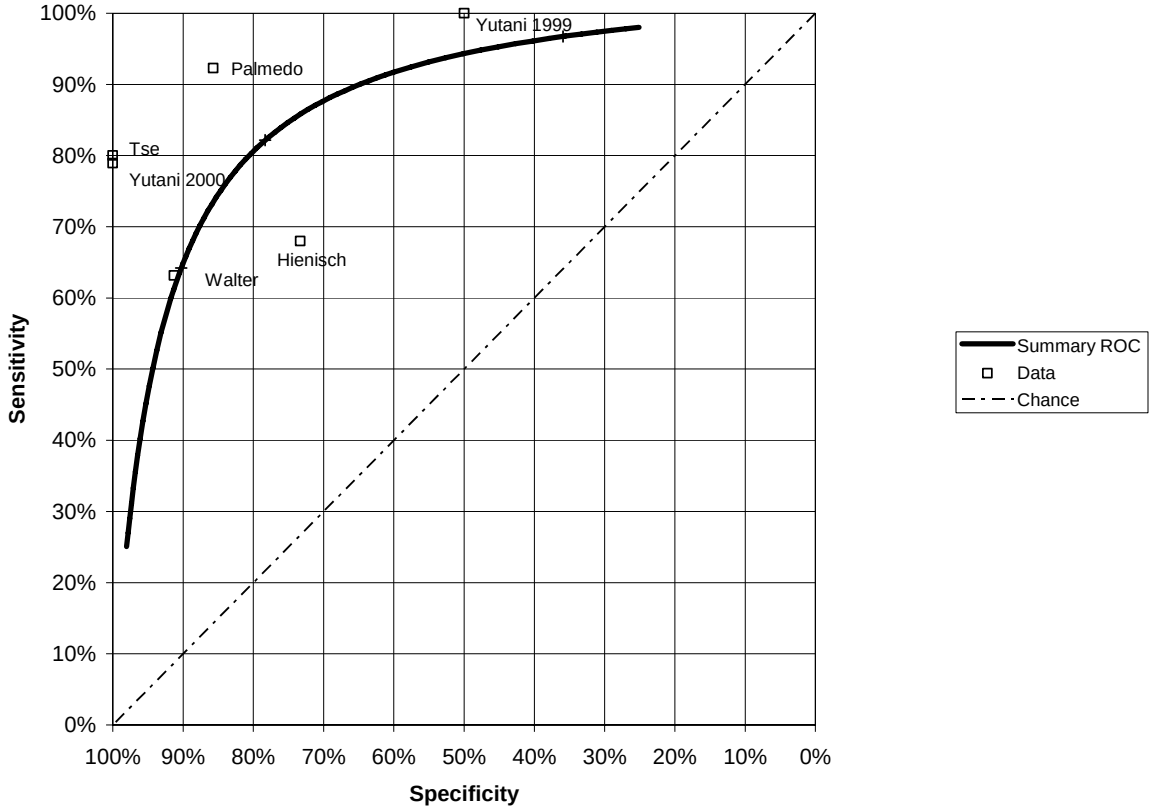
Six studies reported results for 186 lesions in patients referred for further evaluation by whole-body PET scanning for suspicious breast lesions (abnormal mammogram and/or physical examination and/or ultrasound examination), listed in Table 4. The data were not heterogeneous ($I^2 = 0.0\%$), indicating no substantial variability among their results (see Table 31, in Appendix E. Evidence Tables). We therefore combined the data to produce a SROC curve (Figure 2). At the mean threshold of these studies, the sensitivity was 82.2%, and the specificity was 78.3%. At a fixed sensitivity of 95%, the specificity was 46.7%, the positive predictive value

was 95.7%, and the negative predictive value was 79.7% (the prevalence of breast cancer in this population was 70.4%). This evidence base is of moderate quality; however, cumulative meta-analysis shows that the stability of the diagnostic odds ratio (and, hence, the stability of the sensitivity, specificity, and positive and negative predictive values) is low.

Table 4. Studies of PET for Suspicious Breast Lesions

Study	N lesions	Quality score	Sensitivity (95% CI)	Specificity (95% CI)	Positive predictive value (95% CI)	Negative predictive value (95% CI)	Positive likelihood ratio (95% CI)	Negative likelihood ratio (95% CI)
Hienisch et al. 2003[56]	40	7.4	68.0% (48.4 to 82.7)	73.3% (48.0 to 88.9)	81.0% (59.9 to 92.1)	57.9% (36.3 to 76.7)	2.55 (1.95 to 3.34)	0.44 (0.23 to 0.83)
Walter et al. 2003[57]	42	7.9	63.2% (41.0 to 80.7)	91.3% (73.0 to 97.4)	85.7% (59.8 to 95.7)	75.0% (56.6 to 87.2)	7.26 5.15 to 10.24)	0.40 (0.22 to 0.74)
Yutani et al. 2000[55]	40	7.9	78.9% (63.6 to 88.8)	100.0% (34.0 to 99.3)	100.0% (88.4 to 99.9)	20.0% (6.0 to 51.1)	4.69 (3.98 to 5.54)	0.26 (0.12 to 0.57)
Yutani et al. 1999[54]	30	8.8	100.0% (86.8 to 99.9)	50.0% (15.4 to 84.6)	92.9% (77.1 to 97.9)	100.0% (34.0 to 99.3)	1.96 (1.86 to 2.07)	0.04 (0.00 to 0.66)
Palmedo et al. 1997[58]	20	7.9	92.3% (66.4 to 98.3)	85.7% (48.4 to 97.0)	92.3% (66.4 to 98.3)	85.7% (48.4 to 97.0)	6.46 (5.52 to 7.56)	0.09 (0.01 to 0.60)
Tse et al. 1992[59]	14	7.9	80.0% (48.9 to 94.0)	100.0% (50.5 to 99.5)	100.0% (67.0 to 99.7)	66.7% (30.1 to 89.9)	7.73 (5.61 to 10.65)	0.25 (0.08 to 0.78)
6 studies	186 lesions	median 7.9 Moderate	At mean threshold 82.2%	At mean threshold 78.3% At 95% sensitivity 46.7%	At mean threshold 90.0% At 95% sensitivity 80.9%	At mean threshold 64.8% At 95% sensitivity 79.7%	Heterogeneous, no summary estimate calculated	0.33 (0.24 to 0.46)

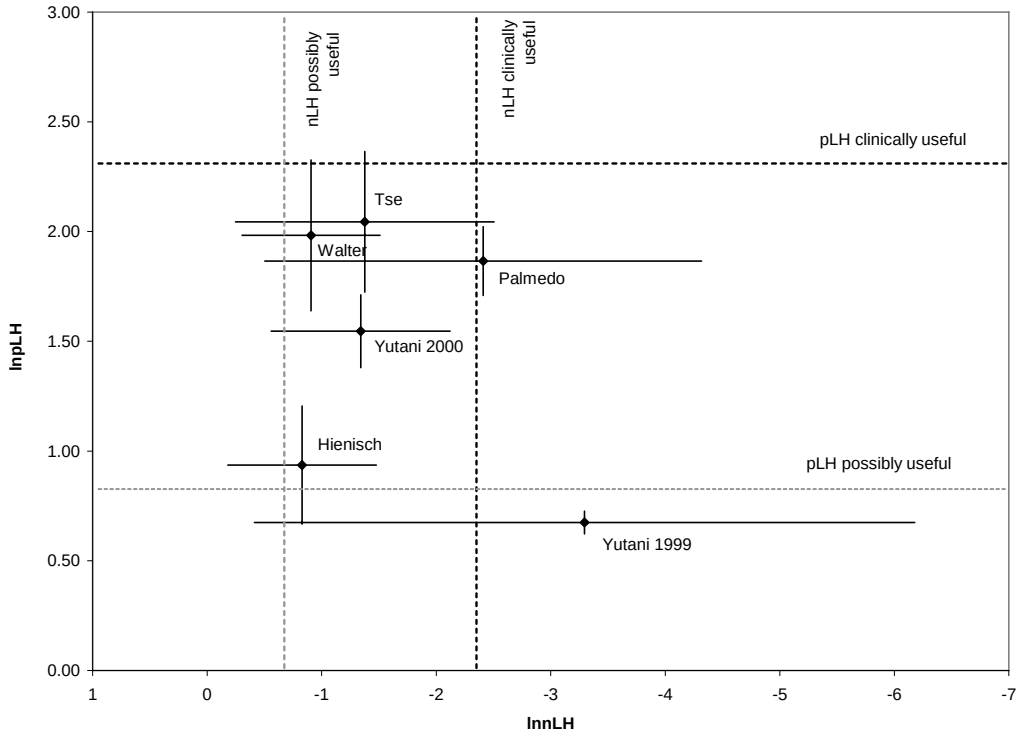
Figure 2. SROC of Whole-Body PET Scanning of Patients Referred for Evaluation of Suspicious Breast Lesions



We calculated the summary likelihood ratios from the six studies of patients referred for further evaluation. The data for the negative likelihood ratio were not heterogeneous (I^2 = 12.3%), so we meta-analytically combined them. The summary negative likelihood ratio was 0.33 (95% CI 0.24 to 0.46), indicating that the test is possibly useful in the clinic. The evidence base is moderate in quality; however, cumulative meta-analysis showed that the summary negative likelihood ratio was of low stability.

The data for the positive likelihood ratio were heterogeneous (I^2 = 73.1%) and therefore not combined. These likelihood ratios were within the possibly clinically useful range (2 or greater, or in natural logs, as used in Figure 3, 0.69 or greater) for only four out of the six studies (67%). Therefore the data are not qualitatively robust, and cannot be used to support a qualitative conclusion. Because positive likelihood ratios are not directly relevant to the purpose of the PET scan for this report, we did not attempt to explain the heterogeneity further.

Figure 3. Likelihood Ratios of PET for Evaluating Suspicious Breast Lesions



nLh = negative likelihood ratio
pLH = positive likelihood ratio
lnpLH = natural log of positive likelihood ratio
lnnLH = natural log of negative likelihood ratio

Other Patient Groups

Only one or two studies reported results for several different patient subgroups (see Table 3, and Table 30 in Appendix E. Evidence Tables). For all of these subgroups, the evidence bases were found to be unacceptably weak for evaluation of test performance due to their small sizes.

Gamma Camera vs. Whole-Body Scanning

One study evaluated the same 30 patients, referred for evaluation of suspicious breast lesions, by both whole-body PET scanning and by gamma camera. Although this evidence base is unacceptably weak for estimation of summary diagnostic test characteristics due to its small size, it is sufficient to support a weak qualitative conclusion because of the extremely large magnitude of the reported difference in diagnostic accuracy between the two methods of imaging. The negative predictive value of whole-body scanning was reported to be 100%, while the negative predictive value of gamma camera imaging was reported to only be 33.3% (Table 5). The negative likelihood ratio of whole-body scanning was 0.04, indicating a clinically useful test, while the negative likelihood ratio of gamma camera imaging was only 0.31. The study is of moderate quality (8.8), and the magnitude of effect is extremely large; therefore, the strength of the evidence is weak.

Table 5. Whole-Body PET Scanning vs. Gamma Camera PET Scanning[a]

Type of scan	Sensitivity (95% CI)	Specificity (95% CI)	Positive predictive value (95% CI)	Negative predictive value (95% CI)	Positive likelihood ratio (95% CI)	Negative likelihood ratio (95% CI)
Whole-body	100.0% (86.8 to 99.9)	50.0% (15.4 to 84.6)	92.9% (77.1 to 97.9)	100.0% (34.0 to 99.3)	1.96 (1.86 to 2.07)	0.04 (0.00 to 0.66)
Gamma camera	84.6% (66.3 to 93.7)	50.0% (15.4 to 84.6)	91.7% (73.9 to 97.5)	33.3% (10.1 to 69.9)	1.69 (1.44 to 1.99)	0.31 (0.08 to 1.17)

a. Data from Yutani et al. 1999[54], for 30 lesions.

Key Question 3. Are there other factors which affect the accuracy or acceptability of PET scanning?

The following narrative review identifies and discusses key factors that experts in the field believe may affect the performance of PET imaging of the breast.

Equipment Differences

Camera Design

Three types of detectors are used for PET imaging of the breast: whole-body PET cameras, standard gamma cameras modified for coincidence detection, and dedicated breast PET cameras. Whole-body PET cameras consist of a ring of detectors that surround the patient. The ring of detectors maximizes the geometric detection efficiency of gamma photons emitted from the patient. Standard gamma cameras can be used if detector heads are placed on opposing sides of the patient and modifications of the electronics (to enable detection of coincident events) are used. The geometry of this arrangement is, however, not optimal for breast imaging since a significant proportion of photons will not be detectable. Also, gamma cameras are designed to detect gamma rays of lower energy compared to those emitted in PET studies. Dedicated mammography PET scanners consist of two small detector heads similar to a standard gamma camera. However, since they are used close to the patient's breast, the geometrical losses are significantly reduced.

There are variations in PET camera performance.[60] Since PET camera performance is governed by the statistic of photon counting, it is theoretically possible to compensate for inferior counting performance with longer image acquisition times or higher doses of FDG. However, factors such as image reconstruction algorithm cannot be corrected for.[60] In a general comparison of PET cameras the investigators concluded that the difference in contrast leads to a systematic difference in measured standardized uptake value (SUV; see below for further explanation), which may affect diagnostic accuracy.[60] It is not clear how clinically relevant these differences are with respect to breast imaging.

Scan Time

Longer image acquisition times will improve the count statistics and, therefore, image quality of any PET scan. However, other factors such as patient movement, comfort, and workflow

mean that acquisition times should be kept as short as possible. The optimum exam time will depend on the characteristics of the detector.[60] Users of dedicated breast PET cameras report acquisition times of four to five minutes.[61] Direct comparisons of scan times between dedicated breast and whole-body PET cameras are difficult because a much larger area is included in whole-body scans and the exam time usually includes an attenuation correction step. Typically the exam takes 45 to 60 minutes.[62] The acquisition time for a single table position is seven to fifteen minutes.[60,63] Lesion detection is compromised if the acquisition time is too short.[64]

Attenuation Correction and SUV Calculation

In whole-body PET studies it is standard practice to acquire a second set of images so that the reconstructed images can be corrected to account for differences in the attenuation of the gamma photons in different areas of the body. In breast imaging some users believe that attenuation correction is essential for tumor localization and quantification of uptake.[63] In contrast, one study found that lesion detectability was impaired when attenuation correction was used.[65]

An attenuation correction image is not acquired when using dedicated mammography PET cameras. Instead, a simple geometric attenuation correction is effective since the attenuation properties of the breast are relatively uniform compared to other body parts.[66]

Regions with higher FDG uptake will appear as foci on the reconstructed images. The SUV is the mean activity detected normalized for the injected dose and body weight. Regions with higher FDG uptake will have a higher SUV. A threshold value is used to identify possible malignancies.[61] However, the SUV is dependent on the image reconstruction algorithm to the extent that diagnostic accuracy may be affected.[60] The reconstruction algorithm is manufacturer dependent. In addition, the SUV depends on several study-specific factors such as FDG uptake time, patient motion, size of the lesion, histology of lesion, patient motion, patient weight, blood glucose level, patient position, and spatial resolution.[62,63,67] Therefore, it is reasonable to conclude that the diagnostic performance of breast PET imaging will vary among manufacturers and studies.

Effect of Operator and Image Interpreter on PET Imaging

The interpretation of breast PET images involves the visual identification of foci and analysis of the computed SUV. Judgment is required to identify foci from the sometimes patchy background.[68] Breast implants can cause photopenic artifacts that underestimate the SUVs.[69] However, automatic SUV calculation is a standard feature on all PET systems. A threshold value is used to determine malignancy.[63] Therefore, intra-observer differences should be insignificant.

Patient Glucose Levels

Since FDG is a glucose analog, the performance of FDG imaging is affected by the patient's glucose levels. It is common practice to instruct patients to fast for four to six hours before any PET study and to monitor glucose levels during imaging. Because muscle activity can affect blood glucose levels, it may be necessary to use a muscle relaxant to reduce muscle activity in anxious patients.[62] Glucose levels must be low. Diabetic patients must have controlled glucose levels.[62,63]

Administered Dose of Tracer

Since FDG is a non-specific tracer, the administered dose does not depend on the anatomy of interest. Therefore, standard doses of FDG are administered for breast imaging using whole-body scanners (300 to 400 MBq, 10 mCi).[62,63] Dedicated mammography PET cameras use significantly lower doses of FDG (~75 MBq, 2 mCi) due to improved geometrical detection efficacy.[68,70] Higher doses may result in a higher tumor detection rate.[63] The relatively short half-life of FDG compared to other radiopharmaceuticals means that the exact dose administered will depend on the availability of FDG and skill of the radiopharmacist. To avoid the effects of FDG contamination at the injection site, the injection site should be contra-lateral to the breast with the suspected lesion.[63]

Radioisotope Uptake Time

FDG is gradually taken up in areas of increased metabolism. Concurrently the ^{18}F is decaying. These two processes counteract each other. Therefore, there is an optimum uptake time. The standard uptake time used in routine PET imaging is about an hour.[62,63] Longer uptake times may improve diagnostic performance.[71] However, other factors must be considered when deciding upon an uptake time, such as workflow and patient well-being (during uptake the patient must rest in isolation, which can be stressful for some patients).

Patient Position

When imaging the breast in a whole-body PET camera, the optimum position for the patient is to lie prone with the breasts hanging free.[63] A study comparing the prone and supine positioning found that SUVs were significantly higher when the patient was prone and padding was used to enable the breasts to hang free.[67]

When using dedicated breast PET cameras, the positioning is similar to x-ray mammography with the breast compressed between the two detector heads.[61,68,70] The compression helps reduce patient movement and the patient may be seated or prone during the acquisition.

Lesion Size

The detectability of lesions that are smaller than the spatial resolution of the imaging system is compromised by the partial volume effect.[63] Spatial resolution is important in breast imaging since the lesions are small. Spatial resolution in PET imaging is influenced by the camera design and the physics of positron annihilation. Whole-body cameras have spatial resolution

specifications in the range 6 to 8 mm. The spatial resolution of dedicated mammography PET cameras is reported to be 2 to 3 mm.[61,70] It is possible to increase sensitivity by mathematically correcting for partial volume contrast losses. However, as sensitivity is increased the specificity decreases.[72]

The sensitivity of PET for breast cancer detection is believed to be highly dependent on the diameter of the lesions.[63,64] A number of studies have reported that the sensitivity of whole-body PET decreases for lesions less than 20 mm in diameter.[54,72,73] Detection of lesions with a modified gamma camera has only been demonstrated for lesions greater than 20 mm in diameter.[54,73] Dedicated mammography PET cameras are reported to be able to detect lesions less than 10 mm in size.[61,68]

Patient Safety and Comfort

Two issues should be considered for patient safety: the radiation dose and pharmaceutical-related adverse reactions. Patient comfort depends on the type of PET camera being used. These issues are discussed below:

Radiation Dose

Using a typical dose for a whole-body scan, the effective radiation dose delivered is 19 μSv/MBq (the value depends on the how often the patient voids). This translates to 7.6 mSv for a typical 400 MBq whole-body PET exam or 1.4 mSv in dedicated mammography PET. The whole-body effective dose is similar to other nuclear medicine studies and equivalent to about two and a half years of background radiation. The dedicated mammography PET effective dose is similar to some x-ray examinations or about 6 months of background radiation. Therefore, radiation dose is not a significant concern, particularly if a dedicated mammography PET camera is used.

Following the exam, the short half-life of ^{18}F means that additional precautions, such as avoiding public transportation, are not necessary.[74]

Pharmaceutical Safety

The intravenous administration of any pharmaceutical could lead to an adverse reaction. In a retrospective analysis of 81,801 administrations of PET radiopharmaceutical, the number of serious adverse reactions reported was zero.[75] Therefore, PET radiopharmaceuticals can be considered safe.

Patient Comfort

All PET studies require the patient to relax for about an hour before image acquisition begins. In a whole-body PET camera, the patient must lie prone for fifteen minutes to an hour, depending on the coverage of the study. No compression is used and the gantry is not as confining as an MR unit. No significant patient comfort issues have been reported.
When using a dedicated mammography PET camera, mild compression is used for about five minutes. The level of compression is less than that used during x-ray mammography.[68] Therefore, patient comfort is not a significant concern.

Accreditation Factors

PET systems are cleared for marketing through the U.S. Food and Drug Administration (FDA) 510(k) process. This clearance is for general-purpose whole-body imaging and is not limited to specific indications. In August 2003, the FDA cleared the first small field-of-view positron emission tomography system (i.e., a dedicated mammography PET camera). Naviscan PET Systems, Inc., formerly PEM Technologies (Rockville, MD, USA), received 510(k) clearance for its PEM 2400 PET system for general imaging of the distribution of injected PET radiopharmaceuticals to determine various metabolic and physiologic functions.

The Intersocietal Commission for the Accreditation of Nuclear Medicine Laboratories (ICANL) offers voluntary accreditation to facilities based on a peer review of their staff's qualifications, education, equipment, quality control, and volume of clinical procedures.[76]

All medical and technical staff are required to meet specific minimum experience and education requirements in order for their facility to be accredited by ICANL. Options available to a facility's medical staff range from board certification in nuclear medicine to board certification in a specialty area with a minimum number of years practice and volume of studies interpreted.

The accreditation program requires the technical director and all technologists working in the facility to hold the RT(N) credential from the American Registry of Radiologic Technologists (ARRT) or the CNMT credential from the Nuclear Medicine Technology Certification Board (NMTCB). In all situations, the physician is ultimately responsible to see that the appropriate images are obtained.

Scintimammography

Scintimammography is similar to PET scanning in that it detects tissues that accumulate higher levels of a radioactive tracer. The tracer most commonly used for breast examination is ^{99m}Tc-sestamibi. Gamma cameras used for scintimammography are designed to perform either planar (two-dimensional) imaging of the three-dimensional radiopharmaceutical distribution or single photon emission tomography (SPECT). SPECT is a technique that may use one, two, or three heads to create a three-dimensional representation of the administered radiopharmaceutical.

Summary

Our analysis found that for non-palpable lesions, at a fixed 95% sensitivity, the specificity of scintimammography was only 39.2%. At the mean threshold of the included studies, the sensitivity was 68.7% and the specificity was 84.8%. The stability of the summary diagnostic odds ratio used to obtain these summary estimates was moderate, indicating a small chance that publication of new evidence could substantially change these estimates.

For palpable lesions and suspicious breast lesions in general, there was unexplained heterogeneity in the data, and therefore summary diagnostic test characteristics could not be calculated (summarized in Table 6).

Our analysis found that, when used to evaluate women with non-palpable lesions, the

negative likelihood ratio of scintimammography was 0.41. This finding means that if a woman with a non-palpable lesion is diagnosed as cancer-free by scintimammography, her chance of actually having breast cancer drops from 20%[2] to 9.3%. Similarly, for a woman referred for further evaluation of non-palpable lesions with microcalcifications who is diagnosed as cancer-free by scintimammography, her chance of actually having breast cancer is still 11.1%, and for a woman referred for further evaluation of lesions larger than 1 cm who is diagnosed as cancer-free by scintimammography, her chance of actually having breast cancer is only 1.7%. However, the stability of our estimates of diagnostic test performance is low and, therefore, there is a reasonable chance that the magnitude of the estimates could substantially change upon publication of new information.

The authors of four other technology assessments concluded that scintimammography was effective and cost-saving for identifying women who could safely avoid biopsy, but at an acceptable cost of some (28 cases/1,000 patients) missed cases of cancer and a potential loss of life. We, however, found a higher rate of missed cancers (93 cases/1,000 patients). The difference between our results and these previous findings may be attributable to the fact that we included studies published after these technology assessments were prepared.

[2] A woman in the general screening population who has a positive finding on mammography and/or physical examination has an approximately 20% chance of having breast cancer.

Table 6. Summary Test Performance of Scintimammography

Patient subgroup	N studies	N lesions	Sensitivity	Specificity	Prevalence (Range)	PPV	NPV	+ LHR (95% CI)	- LHR (95% CI)
Suspicious breast lesion	22	3793	Unexplained heterogeneity. No summary diagnostic test characteristics could be calculated.		45.7% (16.2% to 86.1%)	Unexplained heterogeneity. No summary diagnostic test characteristics could be calculated.		Possibly clinically useful. Strength of evidence: Moderate	Possibly clinically useful in ruling out cancer. Strength of evidence: Moderate
Palpable lesions	11	1012	Unexplained heterogeneity. No summary diagnostic test characteristics could be calculated.		72.6% (27.1% to 85.9%)	Unexplained heterogeneity. No summary diagnostic test characteristics could be calculated.		Possibly clinically useful. Strength of evidence: Moderate	The evidence base is unacceptably weak for evaluation of test performance
Non-palpable lesions	10	509	At mean threshold 68.7% Stability of estimate: Moderate	At mean threshold 84.8% At 95% sensitivity 39.2% Stability of estimate: Moderate	57.0% (17.8% to 78.3%)	At mean threshold 85.7% At 95% sensitivity 67.4% Stabiility of estimate: Moderate	At mean threshold 67.2% At 95% sensitivity 85.5% Stability of estimate: Moderate	Possibly clinically useful. Strength of evidence: Moderate	0.41 (0.34 to 0.49) Stability of estimate: Low
Non-palpable lesions with microcalcifications	3	79	58.1% (43.4% to 72.9%) Stability of estimate: Low	86.1% (74.8% to 97.4%) Stability of estimate: Low	54.4% (46.2% to 61.9%)	83.3% (72.2% to 94.5%) Stability of estimate: Low	63.3% (47.5% to 79.0%) Stability of estimate: Low	4.27 (3.47 to 5.26) Stability of estimate: Low	0.50 (0.32 to 0.78) Stabiliity of estimate: Low
Lesions >1.0 cm	3	306	95.1% (92.2% to 97.9%) Stability of estimate: Low	77.8% (68.7% to 86.8%) Stability of estimate: Low	73.5% (61.5% to 87.0%)	92.2% (88.7% to 95.7%) Stability of estimate: Low	85.1% (77.4% to 92.9%) Stability of estimate: Low	The evidence base is unacceptably weak for evaluation of test performance	0.07 (0.05 to 0.10) Stability of estimate: Low
Lesions ≤1.0 cm	2	66	The evidence base is unacceptably weak for evaluation of test performance.						
Lesions 1.0 to 1.5 cm	1	31	The evidence base is unacceptably weak for evaluation of test performance.						

Patient subgroup	N studies	N lesions	Sensitivity	Specificity	Prevalence (Range)	PPV	NPV	+ LHR (95% CI)	- LHR (95% CI)
Lesions >1.5 cm, planar imaging	1	58	The evidence base is unacceptably weak for evaluation of test performance.						
Lesions ≥1.5 cm, SPECT imaging	1	29	The evidence base is unacceptably weak for evaluation of test performance.						
Younger than 65 years of age, planar imaging	2	47	The evidence base is unacceptably weak for evaluation of test performance.						
Younger than 65 years of age, SPECT imaging	1	34	The evidence base is unacceptably weak for evaluation of test performance.						
65 years of age or older	1	18	The evidence base is unacceptably weak for evaluation of test performance.						
BIRADS 1-2	1	696	The evidence base is unacceptably weak for evaluation of test performance.						
BIRADS 3-4	1	348	The evidence base is unacceptably weak for evaluation of test performance.						
BIRADS 3-5	1	78	The evidence base is unacceptably weak for evaluation of test performance.						
BIRADS 4-5	1	91	The evidence base is unacceptably weak for evaluation of test performance.						
BIRADS 3	1	20	The evidence base is unacceptably weak for evaluation of test performance.						
BIRADS 4	1	30	The evidence base is unacceptably weak for evaluation of test performance.						
BIRADS 5, planar imaging	2	227	The evidence base is unacceptably weak for evaluation of test performance.						
BIRADS 5, SPECT imaging	1	30	The evidence base is unacceptably weak for evaluation of test performance.						
Indeterminate findings on mammogram	2	100	The evidence base is unacceptably weak for evaluation of test performance.						
Positive by fine needle aspiration	1	111	The evidence base is unacceptably weak for evaluation of test performance.						
Dense breast tissue	1	276	The evidence base is unacceptably weak for evaluation of test performance.						
Fatty breast tissue	1	304	The evidence base is unacceptably weak for evaluation of test performance.						
Premenopausal patients	1	117	The evidence base is unacceptably weak for evaluation of test performance.						
Post-menopausal patients	1	140	The evidence base is unacceptably weak for evaluation of test performance.						
SPECT vs. planar imaging	4	243	The evidence base is unacceptably weak to support a conclusion as to whether SPECT or planar is less likely to miss cases of cancer.						

PPV = positive predictive value
NPV = negative predictive value
LHR = likelihood ratio

Analysis and Results for Key Questions 1. and 2. Diagnostic Test Characteristics, Predictive Values, and Likelihood Ratios of Scintimammography

Other Published Technology Assessments

We identified two decision/cost effectiveness analyses and two systematic reviews of the use of scintimammography to evaluate women after a positive mammography exam. These reports were published prior to publication of some of the studies included in the present report. Regardless, the findings of these four reports are briefly summarized in Table 7. All of the authors of these analyses concluded that scintimammography was effective and cost-saving for identifying women who could safely avoid biopsy. However, the authors of both cost-effectiveness analyses concluded that there was a cost of some missed cancers, and a potential loss of life, associated with the routine use of scintimammography.

The estimates of scintimammography test performance used in the two decision analyses were different than the estimates we derived from our analysis of the current literature. For non-palpable lesions, at mean threshold we found a sensitivity of 68.7% and a specificity of 84.8%, and we estimated the routine use of scintimammography would lead to 93 missed cancers per 1,000 women. Hillner[77] used a sensitivity of 85% and a specificity of 90%, data taken from a preliminary report from a single study (published only in abstract form at the time Hillner's report was being prepared in 1996). At a specificity of 85%, our analysis found scintimammography had a sensitivity of only 65%. Using higher estimates of test performance led Hillner et al. to conclude that scintimammography would miss fewer cancers (only 28 missed per 1,000 women) than our estimate of 93 missed per 1,000 women.

Table 7. Other Published Technology Assessments of Scintimammography

Study	Methods	Conclusions
Hillner 1997[77]	A decision analysis model comparing scintimammography to core biopsy and open surgical biopsy for hypothetical cohorts of women with nonpalpable breast lesions detected by mammography. The performances of scintimammography and biopsy were estimated from the general literature.	The model predicted that per 1,000 women, core biopsy would miss seven invasive and 10 *in situ* cancers, as compared to open surgery. Scintimammography would miss an additional 16 invasive cancers and 12 *in situ* cancers, as compared to core biopsy. However, most missed cancers would be detected if all women with negative findings received a 6-month followup mammography, and 65% of women undergoing scintimammography would be able to avoid any type of biopsy. Compared to undergoing immediate surgery, costs would be reduced by 20% with core biopsy, and by 39% with scintimammography. For each cancer diagnosis that was delayed by six months, the authors concluded that scintimammography would save $77,500.
Allen et al. 2000[78]	A decision tree sensitivity analysis comparing three patient management strategies: core needle biopsy after indeterminate or positive mammograms; core needle biopsy after positive mammograms, but patients with indeterminate mammograms were examined by scintimammography, and sent for core biopsy only if positive by scintimammography; all patients with indeterminate or positive mammograms were examined by scintimammography, and sent for core biopsy only if positive by scintimammography. Values used in the analysis were derived from the general literature.	The model predicted that the use of scintimammography would save money by reducing the number of biopsies, but at a cost of lost life expectancy. The use of scintimammography after indeterminate mammograms would save $189 million per year (assuming 21 million women undergo mammographic screening per year) at a cost of a loss of 0.000178 years of mean life expectancy. The use of scintimammography after positive and indeterminate mammograms would save $420 million per year, at a cost of a loss of 0.000222 years of life expectancy.
Liberman et al. 2003[79]	A systematic review of the literature on the accuracy of scintimammography in the diagnosis of breast cancer. The review included 64 papers published between January 1967 and December 1999. The diagnostic test characteristics were individually combined meta-analytically in a fixed-effects model. Quality of the studies was formally assessed and used to weight the studies in the meta-analysis.	The aggregated summary test characteristics for scintimammography were 85.2% sensitivity and 86.6% specificity. For patients with a palpable mass, sensitivity was 87.8% and specificity was 87.5%. For patients without a palpable mass, lesions detected by mammography, sensitivity was 66.8% and specificity was 86.9%. The authors of the review concluded that scintimammography may be used effectively as an adjunct to mammography and physical examination in the diagnosis of breast cancer.
Medical Advisory Secretariat, Ontario Ministry of Health 2003[80]	A systematic review of the literature on the effectiveness of scintimammography in breast cancer detection. Studies published between 1992 and 2002 were eligible for inclusion. Seven studies directly comparing ultrasound to scintimammography, and 49 studies assessing the accuracy of scintimammography, were included. The data from the included studies was combined meta-analytically using the method of Littenburg and Moses.[81]	The authors concluded that scintimammography is an effective imaging technique that can improve the ability to classify patients correctly. Summary receiver operating curves were shown, but no summary test characteristics were derived.

Allen et al.[78] obtained estimates of test performance from the literature available at the time, and the numbers used, 52% sensitivity, 93.3% specificity, are similar to our findings, but at a higher threshold than that used in the majority of recently published studies. The threshold chosen by Allen et al. would send very few women for biopsy, resulting in substantial cost savings in exchange for the missed cases of cancer. If Allen et al. had used the lower threshold in common use in the current literature (around 70% sensitivity), many more biopsies would have been performed per missed case of cancer, and Allen et al. may have reached a different conclusion as to the clinical utility of scintimammography.

The authors of the two systematic reviews both concluded that scintimammography was an effective imaging tool for aiding in the diagnosis of breast cancer. Both systematic reviews reported the sensitivity and specificity of scintimammography. They did not discuss the possible clinical implications of these estimates, nor did they report likelihood ratios or predictive values. Therefore, the authors' rationale for directly concluding from simple estimates of sensitivity and specificity that scintimammography is an effective imaging tool is unclear.

Included Studies

Forty-four diagnostic cohort studies published in 45 manuscripts met our inclusion criteria. Characteristics of the studies and included patients are summarized in Table 32, in Appendix E. Evidence Tables. The quality of all of the studies was moderate (median score of 7.9, range 6.0 to 8.8). A common shortcoming in these studies was not reporting whether readers of tests were blinded to patient information or the results of other tests.

The patients enrolled in these studies were incompletely described. Most of the studies enrolled any patient referred for biopsy due to a suspicious lesion discovered on physical exam or mammography. A few studies had more specific criteria, such as the presence of microcalcifications, specific BIRADS categories after mammography, or only non-palpable lesions. The patients ranged in age from 17 to 94. Reported mean ages were around 50, suggesting that the patient populations studied are younger than the typical breast cancer population. Information about the percentage of patients 65 years of age or older was only reported for 815 of the 5,644 patients. Of these 815 patients, 163, or 20%, were 65 years of age or older.

The diagnostic test characteristics that we computed from data reported by each of the studies are shown in Table 34, in Appendix E. Evidence Tables.

Patients Referred for Further Evaluation of Suspicious Breast Lesions

Twenty-two studies reported results for 3793 lesions in patients referred for further evaluation of suspicious breast lesions (abnormal mammogram and/or abnormal physical examination), listed in Table 8. The data were heterogenous (I^2 = 68%) and therefore we did not combine them (Table 35, in Appendix E. Evidence Tables). We performed meta-regression to identify possible causes of the heterogeneity (the results of this analysis are summarized in Table 36, in Appendix E. Evidence Tables). None of the variables we examined had a statistically significant correlation with the diagnostic odds ratio. Because the heterogeneity could not be explained, we did not compute a summary estimate of diagnostic test performance. The individual study estimates of diagnostic test performance are shown in Figure 4. As can be seen, most of the estimates are tightly clustered in the 80 to 90% sensitivity/specificity region.

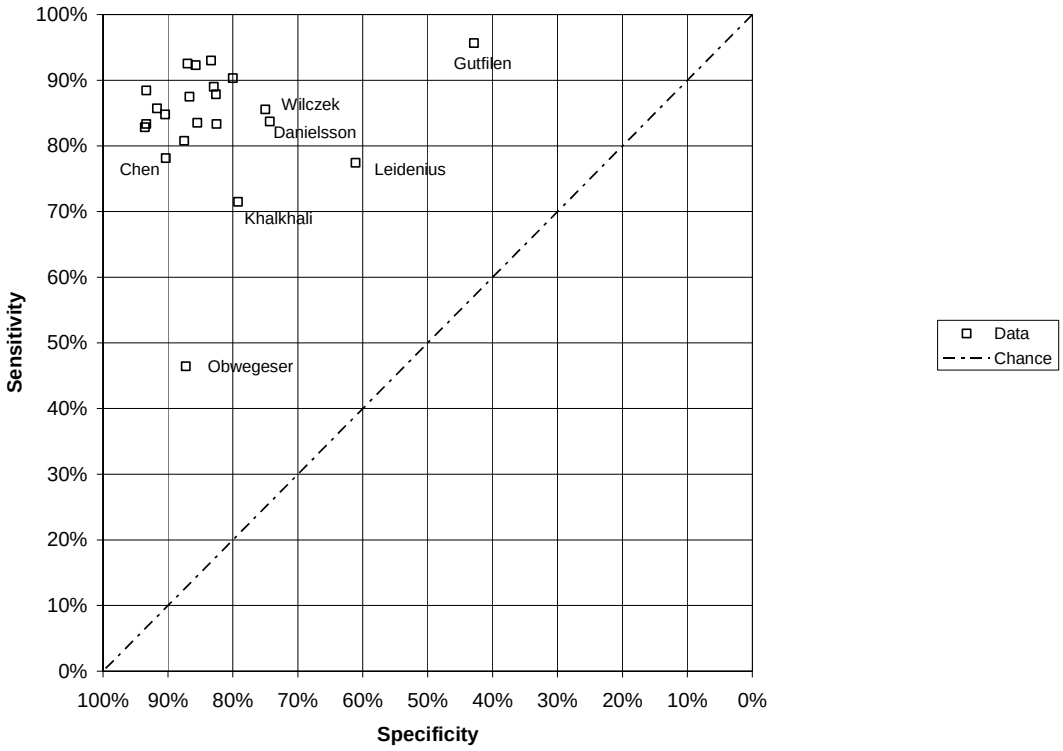
For clarity, in the figure the clustered estimates have not been labeled with study names; only the outliers have been labeled.

Table 8. Studies of Scintimammography for Evaluation of Suspicious Breast Lesions

Study	N lesions	Quality score	Sensitivity (95% CI)	Specificity (95% CI)	Positive predictive value (95% CI)	Negative predictive value (95% CI)	Positive likelihood ratio (95% CI)	Negative likelihood ratio (95% CI)
Sampalis et al. 2003[82]	1243	6	92.5% (88.0% to 95.4%)	86.9% (84.8% to 88.9%)	57.8% (52.3% to 63.0%)	98.4% (97.3% to 99.0%)	7.09 6.82 to 7.37)	0.09 (0.05 to 0.14)
Khalkhali et al. 2002[83]	580	8.3	71.5% (65.1% to 77.1%)	79.2% (74.4% to 83.3%)	69.5% (63.2% to 75.2%)	80.7% (76.0% to 84.7%)	3.44 (3.16 to 3.74)	0.36 (0.29 to 0.45)
Scopinaro et al. 1997[84]	449	8.3	84.8% (80.7% to 88.1%)	90.4% (82.7% to 94.8%)	97.1% (94.5% to 98.4%)	61.2% (52.8% to 68.8%)	8.86 (8.47 to 9.25)	0.17 (0.13 to 0.22)
Tofani et al. 1999[85]	300	8.3	89.0% (84.1% to 92.5%)	82.9% (73.3% to 89.5%)	93.3% (89.0% to 95.9%)	73.9% (64.1% to 81.8%)	5.21 (4.97 to 5.46)	0.13 (0.09 to 0.20)
Mekhmandarov et al. 1998[86]	140	8.8	83.5% (74.2% to 89.9%)	85.5% (73.8% to 92.4%)	89.9% (81.2% to 94.7%)	77.0% (65.0% to 85.7%)	5.74 (5.23 to 6.31)	0.19 (0.12 to 0.31)
Danielsson et al. 1999[87]	121	8.6	83.7% (74.5% to 90.0%)	74.3% (57.9% to 85.7%)	88.9% (80.1% to 94.0%)	65.0% (49.5% to 77.8%)	3.26 (2.97 to 3.57)	0.22 0.13 to 0.37)
Wilczek et al. 2003[88]	119	8.6	85.5% (76.3% to 91.5%)	75.0% (58.9% to 86.1%)	88.8% (79.9% to 93.9%)	69.2% (53.5% to 81.3%)	3.42 (3.13 to 3.74)	0.19 (0.11 to 0.34)
Krishnaiah et al. 2003[89]	104	7.4	83.3% (64.0% to 93.1%)	82.5% (72.7% to 89.2%)	58.8% (42.2% to 73.6%)	94.3% (86.1% to 97.7%)	4.76 (3.98 to 5.69)	0.20 (0.08 to 0.50)
Obwegeser et al. 1999[90]	103	7.4	46.4% (34.0% to 59.3%)	87.2% (74.7% to 93.9%)	81.3% (64.6% to 91.0%)	57.7% (46.1% to 68.5%)	3.64 (2.75 to 4.82)	0.61 (0.47 to 0.80)
Koukouraki et al. 2001[91]	86	7.5	93.0% (85.5% to 96.7%)	83.3% (66.3% to 92.5%)	94.1% (86.9% to 97.4%)	80.6% (63.6% to 90.7%)	5.58 (5.27 to 5.91)	0.08 (0.04 to 0.18)
Schillaci et al. 1997[92]	66	8.8	85.7% (72.0% to 93.2%)	91.7% (73.9% to 97.5%)	94.7% (82.5% to 98.4%)	78.6% (60.4% to 89.6%)	10.29 (9.09 to 11.64)	0.16 (0.07 to 0.33)
Villanueva-Meyer et al. 1996[93]	66	8.6	82.9% (67.2% to 91.8%)	93.5% (79.1% to 98.1%)	93.5% (79.1% to 98.1%)	82.9% (67.2% to 91.8%)	12.84 11.05 to 14.93)	0.18 (0.09 to 0.38)
Chen et al. 1997[94]	63	8.1	78.1% (61.2% to 88.8%)	90.3% (74.9% to 96.5%)	89.3% (72.6% to 96.1%)	80.0% (64.0% to 89.8%)	8.07 (6.72 to 9.70)	0.24 (0.12 to 0.47)

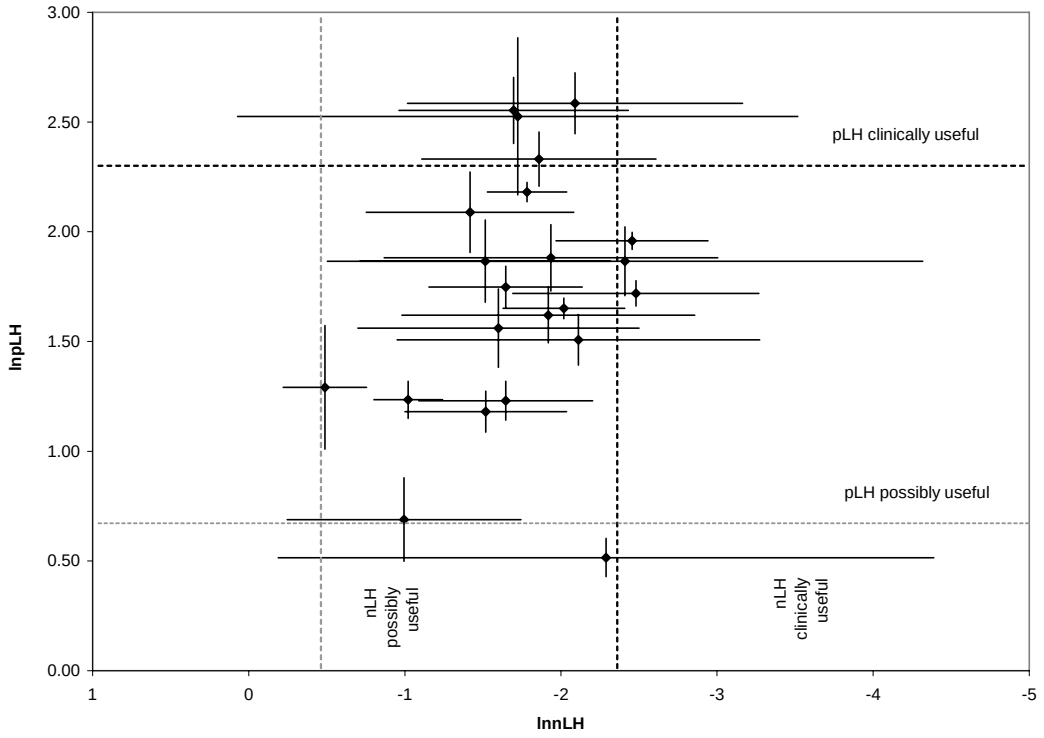
Study	N lesions	Quality score	Sensitivity (95% CI)	Specificity (95% CI)	Positive predictive value (95% CI)	Negative predictive value (95% CI)	Positive likelihood ratio (95% CI)	Negative likelihood ratio (95% CI)
Tiling et al. 1997[95]	56	8.3	87.9% (72.5% to 95.0%)	82.6% (62.7% to 92.8%)	87.9% (72.5% to 95.0%)	82.6% (62.7% to 92.8%)	5.05 (4.45 to 5.74)	0.15 (0.06 to 0.37)
Palmedo et al. 1996[96]	54	7.9	87.5% (68.8% to 95.5%)	86.7% (70.2% to 94.5%)	84.0% (65.2% to 93.4%)	89.7% (73.4% to 96.3%)	6.56 (5.64 to 7.63)	0.14 (0.05 to 0.42)
Leidenius et al. 2002[97]	49	7.9	77.4% (60.1% to 88.5%)	61.1% (38.6% to 79.6%)	77.4% (60.1% to 88.5%)	61.1% (38.6% to 79.6%)	1.99 (1.65 to 2.41)	0.37 (0.17 to 0.78)
Imbriaco et al. 2001[98]	49	7.6	80.8% (62.0% to 91.3%)	87.5% (68.8% to 95.5%)	87.5% (68.8% to 95.5%)	80.8% (62.0% to 91.3%)	6.46 (5.36 to 7.79)	0.22 (0.10 to 0.49)
Papantoniou et al. 2001[99]	41	7.9	88.5% (70.8% to 95.8%)	93.3% (69.9% to 98.6%)	95.8% (79.5% to 99.1%)	82.4% (58.8% to 93.6%)	13.27 (11.55 to 15.25)	0.12 (0.04 to 0.36)
Sanidas et al. 2003[100]	33	8.3	90.3% (74.9% to 96.5%)	80.0% (37.4% to 95.9%)	96.6% (82.6% to 99.2%)	57.1% (25.2% to 83.9%)	4.52 (4.02 to 5.07)	0.12 (0.04 to 0.39)
Gutfilen et al. 2001[101]	30	6.9	95.7% (78.7% to 99.0%)	42.9% (16.1% to 74.8%)	84.6% (66.3% to 93.7%)	75.0% (30.1% to 94.9%)	1.67 (1.53 to 1.83)	0.10 (0.01 to 0.83)
Yuen-Green et al. 1996[102]	21	7.9	83.3% (43.5% to 96.5%)	93.3% (69.9% to 98.6%)	83.3% (43.5% to 96.5%)	93.3% (69.9% to 98.6%)	12.50 (8.74 to 17.88)	0.18 (0.03 to 1.07)
Palmedo et al. 1997[58]	20	7.9	92.3% (66.4% to 98.3%)	85.7% (48.4% to 97.0%)	92.3% (66.4% to 98.3%)	85.7% (48.4% to 97.0%)	6.46 (5.52 to 7.56)	0.09 (0.01 to 0.60)
22 studies	3,793 lesions	median 8.0 Moderate	Heterogeneous, no summary estimates calculated.					

Figure 4. Diagnostic Test Characteristics of Scintimammography for Evaluating Suspicious Breast Lesions



The natural log of the likelihood ratios calculated from data reported by each study are shown graphically in Figure 5. The majority (18 out of 22; 82%) of the negative likelihood ratios are within the possibly/clinically useful range (less than 0.5, or for natural logs as shown in the graph, less than -0.69). Therefore the data are qualitatively consistent, and we can conclude that scintimammography may be clinically useful in ruling out breast cancer, supported by evidence of moderate strength. Similarly, because all but two of the positive likelihood ratios are within the possibly/clinically useful range, we can conclude that scintimammography may be clinically useful for ruling in breast cancer, supported by evidence of moderate strength.

Figure 5. Likelihood Ratios of Scintimammography for Evaluating Suspicious Breast Lesions



nLh = negative likelihood ratio
pLH = positive likelihood ratio
lnpLH = natural log of positive likelihood ratio
lnnLH = natural log of negative likelihood ratio

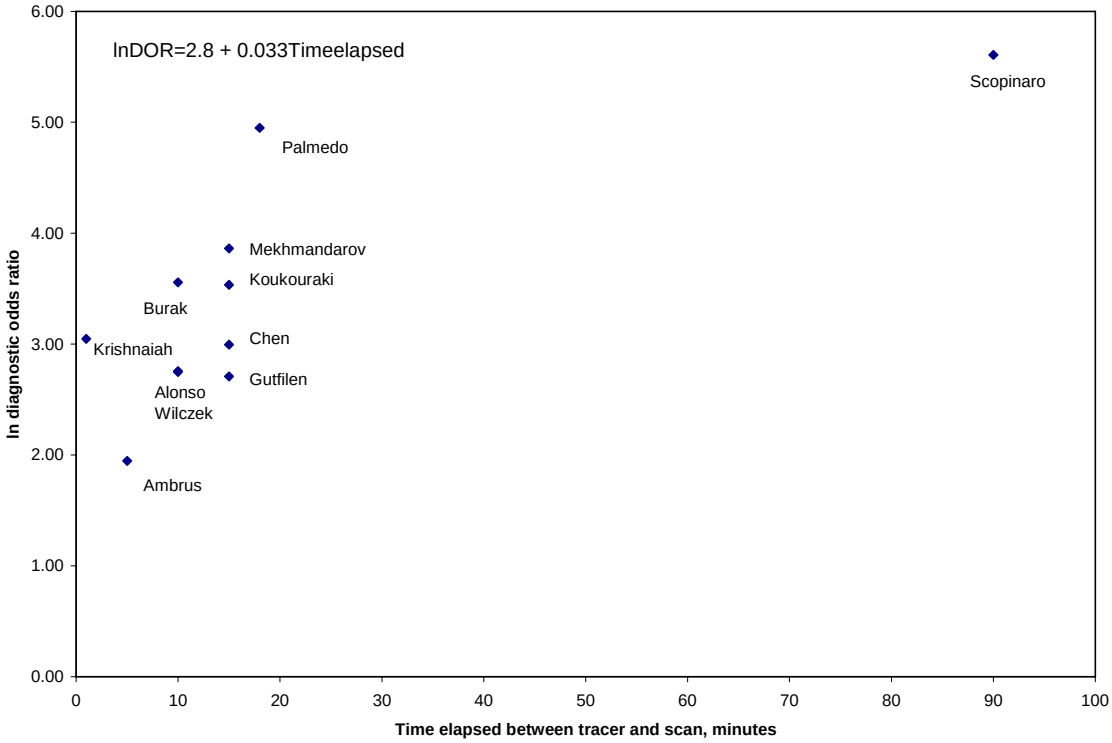
Palpable Lesions

Eleven studies reported results for 1,012 lesions in patients referred for further examination after detection of a palpable lesion (listed in Table 9). The data were heterogeneous ($I^2 = 57\%$) and, therefore, we did not directly combine them (see Table 37 in Appendix E. Evidence Tables). Instead, we performed a meta-regression (summarized in Table 38, in Appendix E. Evidence Tables) to identify possible causes of the heterogeneity. "Time elapsed between tracer injection and scan" significantly correlated with the diagnostic odds ratio ($p = 0.0040$), but explained only 47% of the between-studies variation. None of the other variables we examined had a statistically significant correlation with the diagnostic odds ratio. The final model fitted is shown in Figure 6. All of the studies but one (Scopinaro et al.) allowed fewer than 20 minutes to elapse between injection of the tracer and the scan. When the data from Scopinaro et al. are removed from the analysis, the correlation between time elapsed and diagnostic odds ratio is no longer statistically significant ($p = 0.0650$). Because the statistically significant results are due to the results of one outlying study, this model is not robust enough to permit conclusions.

Table 9. Studies of Scintimammography for Palpable Lesions

Study	N lesions	Quality score	Sensitivity (95% CI)	Specificity (95% CI)	Positive predictive value (95% CI)	Negative predictive value (95% CI)	Positive likelihood ratio (95% CI)	Negative likelihood ratio (95% CI)
Scopinaro et al. 1997[84]	283	8.3	97.3% (94.3% to 98.7%)	89.7% (79.1% to 95.1%)	97.3% (94.3% to 98.7%)	89.7% (79.1% to 95.1%)	9.41 (9.21 to 9.61)	0.03 (0.01 to 0.07)
Alonso et al. 2001[103]	245	7.9	83.1% (77.1% to 87.7%)	76.8% (64.2% to 85.8%)	92.4% (87.3% to 95.4%)	57.3% (46.0% to 67.9%)	3.58 (3.36 to 3.82)	0.22 (0.16 to 0.31)
Mekhmandarov et al. 1998[86]	85	8.8	95.1% (86.4% to 98.2%)	75.0% (55.0% to 87.8%)	90.6% (80.9% to 95.5%)	85.7% (65.2% to 94.8%)	3.80 (3.59 to 4.03)	0.07 (0.02 to 0.20)
Koukouraki et al. 2001[91]	78	7.5	94.0% (85.5% to 97.6%)	72.7% (43.4% to 90.0%)	95.5% (87.4% to 98.4%)	66.7% (39.0% to 86.0%)	3.45 (3.25 to 3.66)	0.08 (0.03 to 0.23)
Wilczek et al. 2003[88]	65	8.6	91.3% (79.5% to 96.5%)	63.2% (41.0% to 80.7%)	85.7% (73.2% to 92.8%)	75.0% (50.4% to 89.6%)	2.48 (2.27 to 2.71)	0.14 (0.05 to 0.37)
Krishnaiah et al. 2003[89]	59	7.4	87.5% (63.7% to 96.3%)	79.1% (64.7% to 88.5%)	60.9% (40.8% to 77.7%)	94.4% (81.7% to 98.3%)	4.18 (3.47 to 5.03)	0.16 (0.04 to 0.58)
Ambrus et al. 1997[104]	51	7.5	50.0% (35.2% to 64.8%)	90.9% (61.9% to 98.1%)	95.2% (77.0% to 99.0%)	33.3% (19.3% to 51.3%)	5.50 4.03 to 7.50)	0.55 (0.38 to 0.79)
Burak et al. 1994[105]	41	7.9	88.9% (71.8% to 96.0%)	85.7% (59.8% to 95.7%)	92.3% (75.6% to 97.7%)	80.0% (54.7% to 92.7%)	6.22 (5.45 to 7.11)	0.13 (0.04 to 0.38)
Palmedo et al. 1996[106,107]	40	7.9	100.0% (83.5% to 99.8%)	78.9% (56.5% to 91.3%)	83.3% (64.0% to 93.1%)	100.0% (79.2% to 99.8%)	4.34 (4.06 to 4.64)	0.03 (0.00 to 0.48)
Chen et al. 2000[108]	38	8.8	77.8% (54.7% to 90.8%)	88.2% (65.4% to 96.5%)	87.5% (63.7% to 96.3%)	78.9% (56.5% to 91.3%)	6.61 (5.16 to 8.46)	0.25 (0.10 to 0.61)
Gutfilen et al. 2001[101]	27	6.9	95.7% (78.7% to 99.0%)	50.0% (15.4% to 84.6%)	91.7% (73.9% to 97.5%)	66.7% (21.0% to 93.3%)	1.91 (1.75 to 2.09)	0.09 (0.01 to 0.75)
11 studies	1,012 lesions	median 7.9 Moderate	Heterogeneous, no summary estimates calculated.					

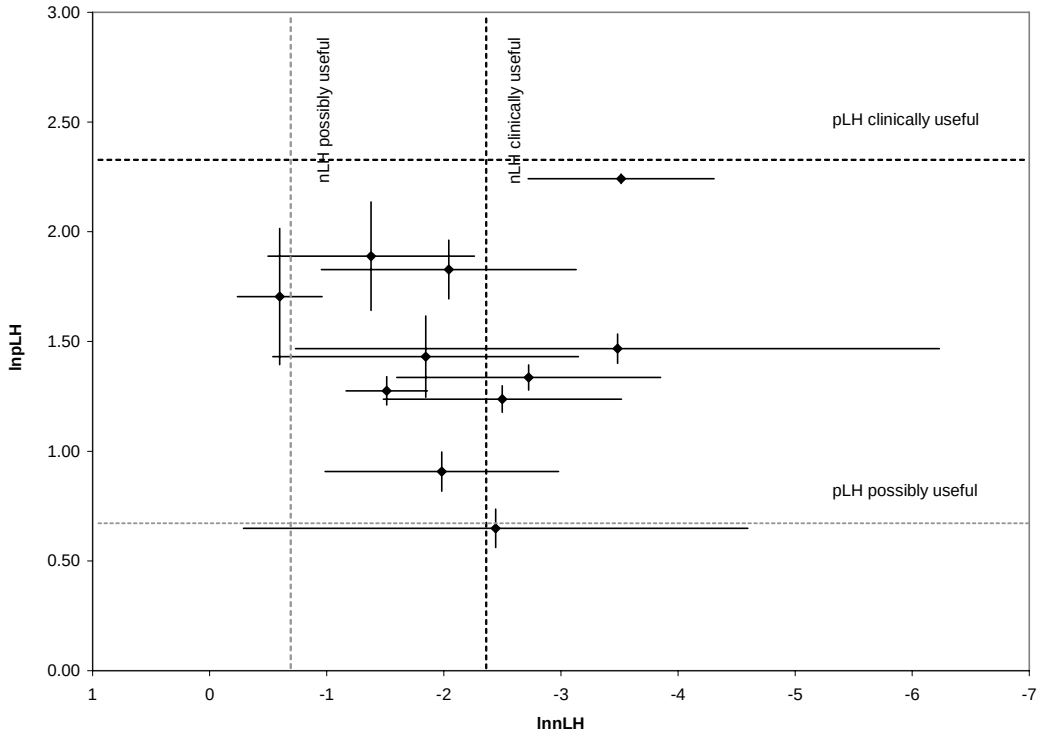
Figure 6. Palpable Lesions Examined by Scintimammography: Meta-Regression Model



InDOR = natural log diagnostic odds ratio

The natural log of the likelihood ratios we calculated from data reported by each study are shown in Figure 7. Only seven out of eleven (63.6%) of the negative likelihood ratios are clearly within the possibly clinically useful or the clinically useful range (less than 0.5, or for natural logs as shown in the graph, less than -0.69). Therefore the negative likelihood ratios are not qualitatively robust, which renders the evidence unacceptably weak for supporting a conclusion about the usefulness of scintimammography in ruling out breast cancer when evaluating palpable lesions.

Figure 7. Likelihood Ratios of Scintimammography for Palpable Lesions



nLh = negative likelihood ratio
pLH = positive likelihood ratio
lnpLH = natural log of positive likelihood ratio
lnnLH = natural log of negative likelihood ratio

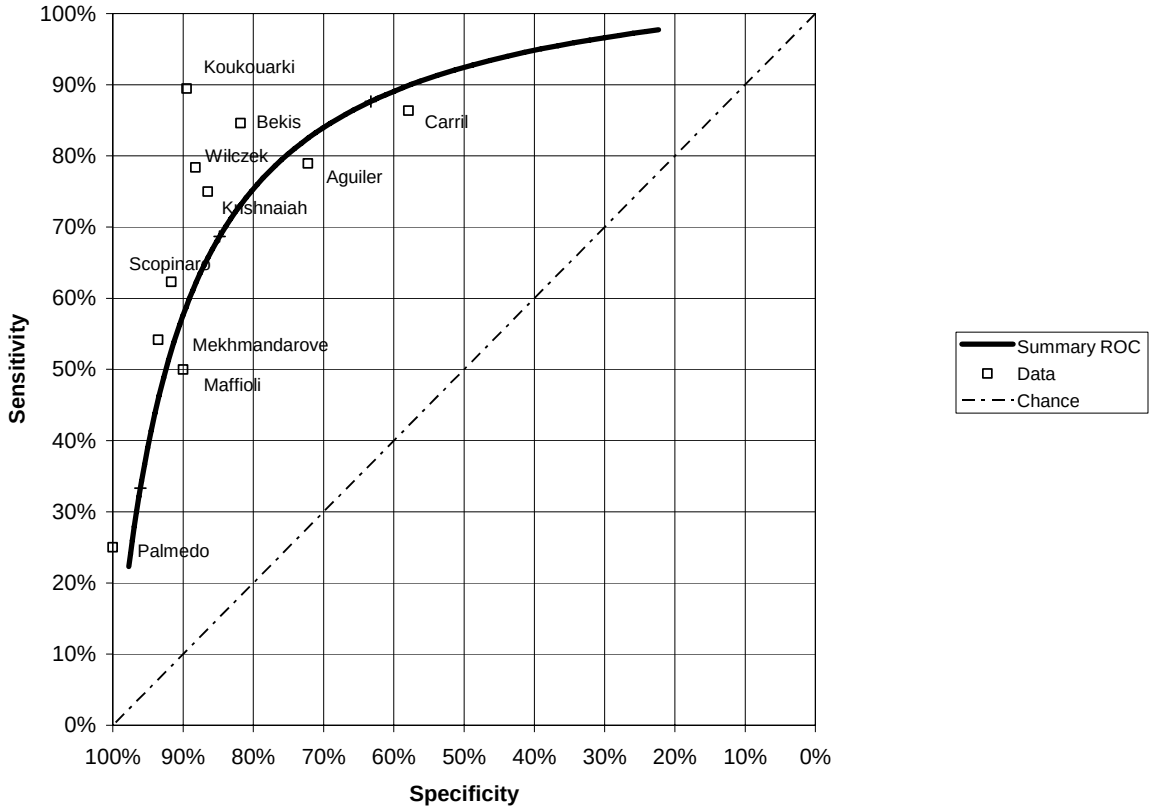
Non-Palpable Lesions

Ten studies reported results for 509 non-palpable lesions (Table 10). The data were not heterogeneous ($I^2 = 0\%$), so we combined them meta-analytically (Table 39, in Appendix E. Evidence Tables). The SROC curve is shown in Figure 8. At the mean of these studies' thresholds, the sensitivity of scintimammography was 68.7% and the specificity was 84.8%. At a fixed sensitivity of 95%, the specificity was 39.2%, the positive predictive value was 91.2%, and the negative predictive value was 54.1%. The prevalence of disease in this population was 57.0%. The evidence base is of moderate quality (median score 8.3), and the summary diagnostic odds ratio was stable as tested by cumulative meta-analysis. Therefore the estimates of diagnostic test performance derived from the SROC curve are of moderate stability.

Table 10. Studies of Scintimammography for Non-palpable Lesions

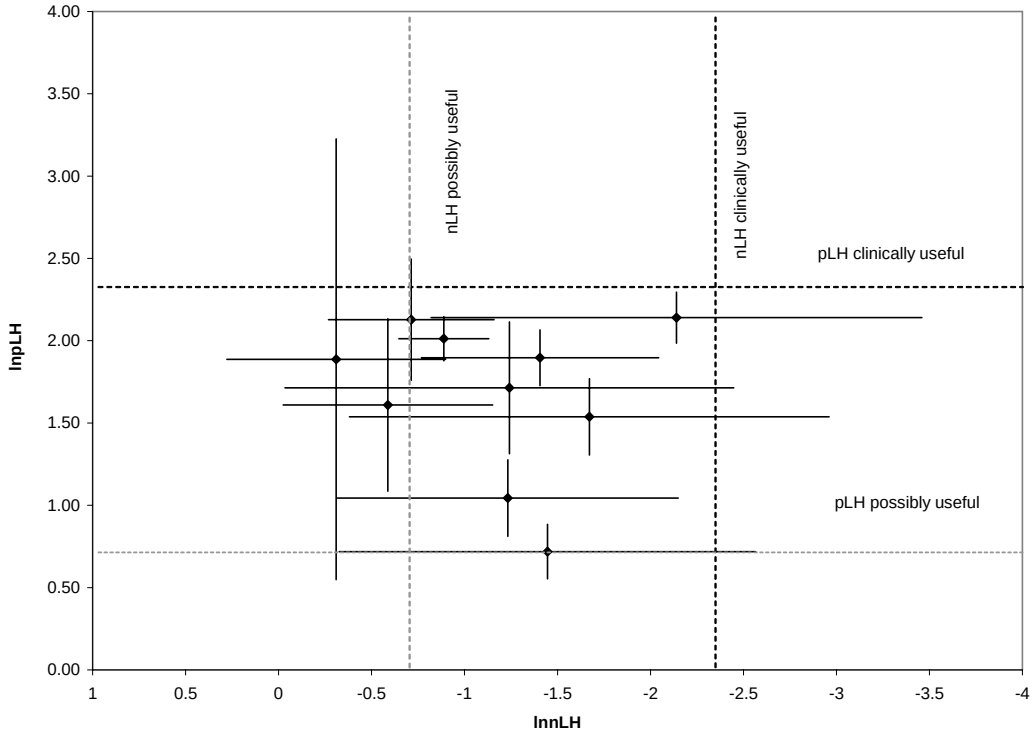
Study	N lesions	Quality score	Sensitivity (95% CI)	Specificity (95% CI)	Positive predictive value (95% CI)	Negative predictive value (95% CI)	Positive likelihood ratio (95% CI)	Negative likelihood ratio (95% CI)
Scopinaro et al. 1997[84]	166	8.3	62.3% (53.7% to 70.2%)	91.7% (78.0% to 97.0%)	96.4% (89.9% to 98.7%)	40.2% (30.3% to 51.1%)	7.48 (6.54 to 8.55)	0.41 (0.32 to 0.52)
Mekhmandarove et al. 1998[86]	55	8.8	54.2% (35.1% to 72.0%)	93.5% (79.1% to 98.1%)	86.7% (61.9% to 96.0%)	72.5% (57.1% to 83.8%)	8.40 (5.81 to 12.13)	0.49 (0.31 to 0.76)
Wilczek et al. 2003[88]	54	8.6	78.4% (62.7% to 88.5%)	88.2% (65.4% to 96.5%)	93.5% (79.1% to 98.1%)	65.2% (44.9% to 81.1%)	6.66 (5.62 to 7.89)	0.25 (0.13 to 0.46)
Krishnaiah et al. 2003[89]	45	7.4	75.0% (40.8% to 92.5%)	86.5% (71.9% to 94.0%)	54.5% (28.1% to 78.6%)	94.1% (80.7% to 98.2%)	5.55 (3.72 to 8.28)	0.29 (0.09 to 0.97)
Carril et al. 1997[109]	41	7.9	86.4% (66.5% to 95.1%)	57.9% (36.3% to 76.7%)	70.4% (51.5% to 84.0%)	78.6% (52.3% to 92.2%)	2.05 (1.74 to 2.42)	0.24 (0.08 to 0.72)
Koukouraki et al. 2001[91]	38	7.5	89.5% (68.4% to 96.8%)	89.5% (68.4% to 96.8%)	89.5% (68.4% to 96.8%)	89.5% (68.4% to 96.8%)	8.50 (7.29 to 9.92)	0.12 (0.03 to 0.44)
Aguilar et al. 2001[110]	37	8.2	78.9% (56.5% to 91.3%)	72.2% (49.1% to 87.3%)	75.0% (53.0% to 88.6%)	76.5% (52.6% to 90.2%)	2.84 (2.25 to 3.58)	0.29 (0.12 to 0.73)
Bekis et al. 2004[111]	35	8.3	84.6% (57.6% to 95.4%)	81.8% (61.3% to 92.5%)	73.3% (48.0% to 88.9%)	90.0% (69.7% to 97.0%)	4.65 (3.69 to 5.87)	0.19 (0.05 to 0.68)
Maffioli et al. 1996[112]	24	7.5	50.0% (26.9% to 73.1%)	90.0% (59.3% to 97.9%)	87.5% (52.6% to 97.4%)	56.3% (33.2% to 76.8%)	5.00 (2.96 to 8.44)	0.56 (0.32 to 0.98)
Palmedo et al. 1996[106,107]	14	8.3	25.0% (5.1% to 69.9%)	100.0% (71.7% to 99.7%)	100.0% (20.8% to 99.2%)	76.9% (49.6% to 91.6%)	6.60 (1.73 to 25.18)	0.73 (0.41 to 1.32)
10 studies	509 lesions	Median 8.3 Moderate	At mean threshold 68.7%	At mean threshold 84.8% At 95% sensitivity 39.2%	At mean threshold 85.7% At 95% sensitivity 67.4%	At mean threshold 67.2% At 95% sensitivity 54.1%	Heterogeneous, no summary estimate calculated	0.41 (0.34 to 0.49)

Figure 8. SROC Curve of Scintimammography to Evaluate Non-Palpable Lesions



We calculated the summary likelihood ratios from the ten studies of patients referred with non-palpable lesions (see Figure 9 and also Table 39 in Appendix E. Evidence Tables for a summary of the likelihood ratio meta-analysis). The data for the negative likelihood ratios were not heterogeneous (I^2 = 38.5%), so we combined them using meta-analysis. The summary negative likelihood ratio was 0.41 (0.34 to 0.49). The evidence base is moderate in quality, but the summary estimate is not stable by cumulative meta-analysis, and therefore is of low stability. The data for the positive likelihood ratios were hetereogeneous (I^2 = 57.2%), and therefore could not be synthesized to produce a single summary estimate. However, 80% (eight out of ten) of the studies reported a positive likelihood ratio within the possible clinically useful range (2 or greater), supported by moderately strong evidence. Because positive likelihood ratios are not directly relevant to the indication under evaluation in this report, we did not attempt to explain the heterogeneity further.

Figure 9. Likelihood Ratios of Scintimammography to Evaluate Non-Palpable Lesions



nLh = negative likelihood ratio
pLH = positive likelihood ratio
lnpLH = natural log of positive likelihood ratio
lnnLH = natural log of negative likelihood ratio

Non-Palpable Lesions with Microcalcifications

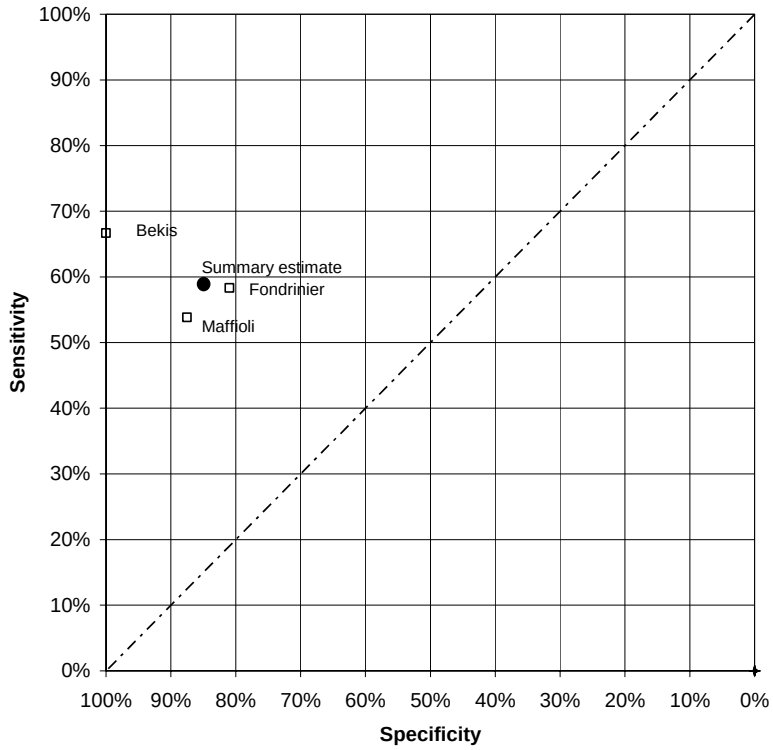
Three studies reported outcomes for 79 non-palpable lesions with microcalcifications evident on mammography (Table 11). The data were not heterogeneous (I^2 = 0%), so we combined them meta-analytically (Table 40, in Appendix E. Evidence Tables). Because there were only three studies, we directly pooled the diagnostic test characteristics instead of deriving a SROC curve. The summary sensitivity is 58.1%, the summary specificity is 86.1%, the summary positive

predictive value is 83.3%, and the summary negative predictive value is 63.3%. The prevalence of disease in this population was 54.4%. The diagnostic test characteristics are shown graphically in Figure 10. Although the studies were of moderate quality and not heterogeneous, all of the summary estimates of test performance were not stable by cumulative meta-analysis, and are, therefore, of low stability.

Table 11. Studies of Scintimammography for Non-palpable Lesions with Microcalcifications

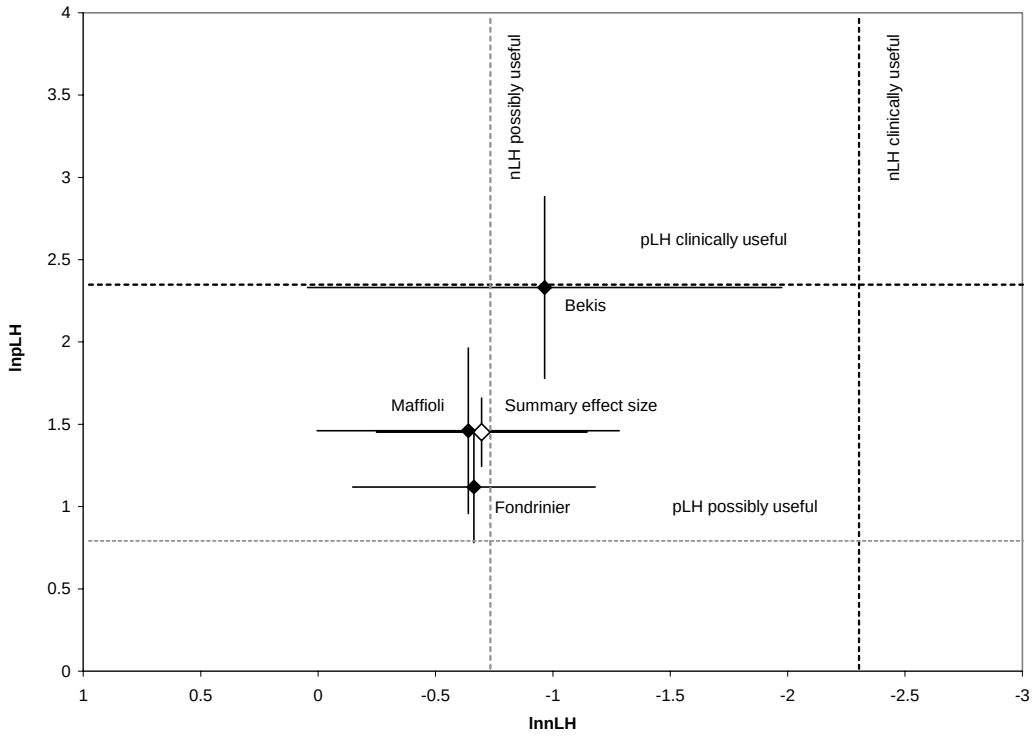
Study	N lesions	Quality score	Sensitivity (95% CI)	Specificity (95% CI)	Positive predictive value (95% CI)	Negative predictive value (95% CI)	Positive likelihood ratio (95% CI)	Negative likelihood ratio (95% CI)
Fondrinier et al. 2004[113]	45	7.2	58.3% (38.8% to 75.4%)	81.0% (59.9% to 92.1%)	77.8% (54.7% to 90.8%)	63.0% (44.2% to 78.4%)	3.06 (2.18 to 4.29)	0.51 (0.31 to 0.86)
Maffioli et al. 1996[112]	21	7.5	53.8% (29.2% to 76.7%)	87.5% (52.6% to 97.4%)	87.5% (52.6% to 97.4%)	53.8% (29.2% to 76.7%)	4.31 (2.60 to 7.13)	0.53 (0.28 to 1.00)
Bekis et al. 2004[111]	13	8.3	66.7% (30.1% to 89.9%)	100.0% (64.0% to 99.6%)	100.0% (50.5% to 99.5%)	77.8% (45.1% to 93.3%)	10.29 (5.92 to 17.87)	0.38 (0.14 to 1.05)
3 studies	79 lesions	Median 7.5 Moderate	58.1% (43.3% to 72.9%)	86.1% (74.8% to 97.4%)	83.3% (72.2% to 94.5%)	63.3% (47.6% to 79.0%)	4.27 (3.47 to 5.26)	0.50 (0.32 to 0.78)

Figure 10. Diagnostic Test Characteristics of Scintimammography to Evaluate Non-palpable Lesions with Microcalcifications



The likelihood ratios from these three studies are shown in Figure 11. Neither the negative or the positive likelihood ratios were heterogeneous (I^2 = % for both), so we calculated summary estimates. The summary negative likelihood ratio is 0.50 (0.32 to 0.78), and the summary positive likelihood ratio is 4.27 (3.47 to 5.26). Neither estimate was stable by cumulative meta-analysis, and therefore the stability of the summary estimates is low.

Figure 11. Likelihood ratios of Scintimammography to Evaluate Non-Palpable Lesions with Microcalcifications



nLh = negative likelihood ratio
pLH = positive likelihood ratio
lnpLH = natural log of positive likelihood ratio
lnnLH = natural log of negative likelihood ratio

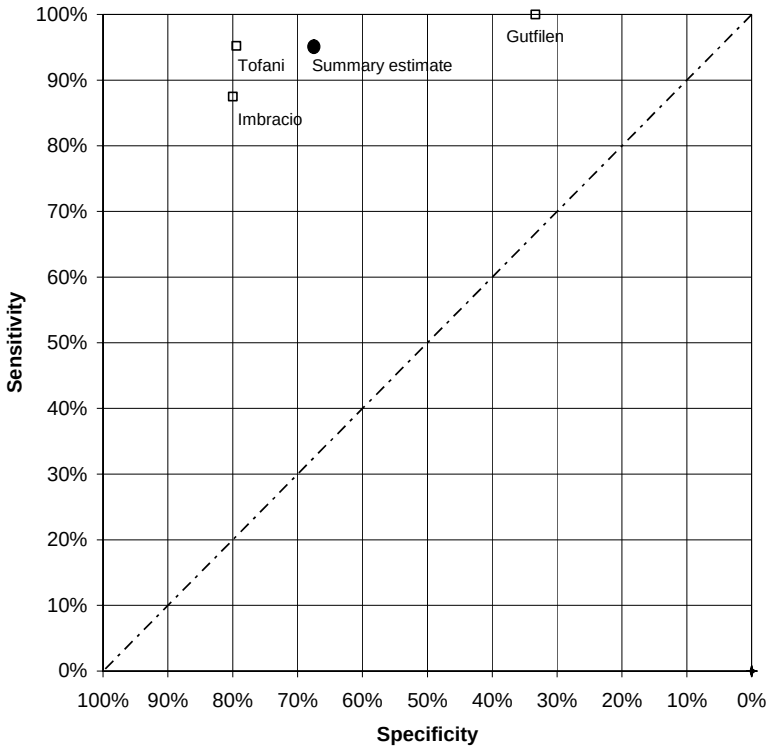
Large (Greater than 10 mm in diameter) Lesions

Three studies reported information about the use of scintimammography to evaluate 306 large lesions (lesions larger than 10 mm in diameter), listed in Table 12. The diagnostic odds ratios were not heterogeneous ($I^2 = 0\%$). Because there were only three studies, we directly pooled the diagnostic test characteristics rather than deriving a SROC curve (Table 41, in Appendix E. Evidence Tables). The summary sensitivity was 95.1%, the summary specificity was 77.8%, the summary positive predictive value was 92.2%, and the summary negative predictive value was 85.1%. The prevalence of disease in the population studied was 73.5%. The summary sensitivity was robust by cumulative meta-analysis, but the summary specificity was not. Therefore the stability of the summary estimates is low. The diagnostic test characteristics are shown graphically in Figure 12.

Table 12. Studies of Scintimammography for Lesions Larger than 10 mm

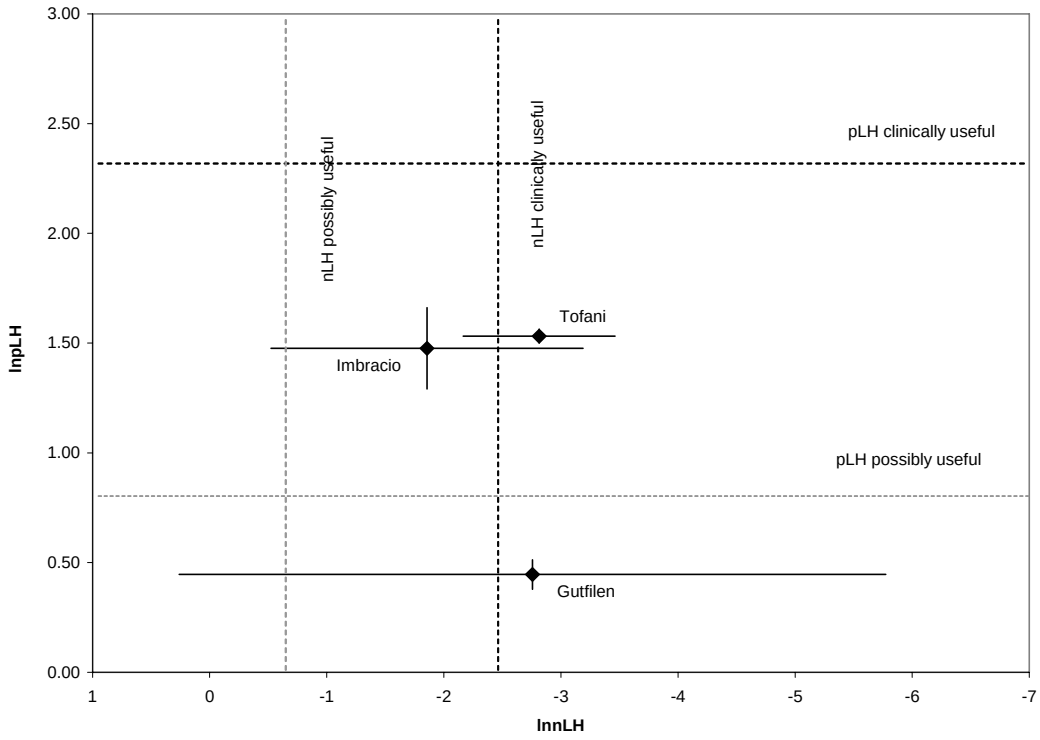
Study	N lesions	Quality score	Sensitivity (95% CI)	Specificity (95% CI)	Positive predictive value (95% CI)	Negative predictive value (95% CI)	Positive likelihood ratio (95% CI)	Negative likelihood ratio (95% CI)
Tofani et al. 1999[85]	257	8.3	95.2% (91.2% to 97.4%)	79.4% (68.3% to 87.3%)	92.8% (88.2% to 95.6%)	85.7% (74.9% to 92.2%)	4.63 (4.48 to 4.78)	0.06 (0.03 to 0.11)
Imbriaco et al. 2001[98]	26	7.6	87.5% (63.7% to 96.3%)	80.0% (48.9% to 94.0%)	87.5% (63.7% to 96.3%)	80.0% (48.9% to 94.0%)	4.38 (3.64 to 5.27)	0.16 (0.04 to 0.59)
Gutfilen et al. 2001[101]	23	6.9	100.0% (83.5% to 99.8%)	33.3% (6.7% to 79.0%)	90.9% (72.0% to 97.3%)	100.0% (20.8% to 99.2%)	1.56 (1.46 to 1.67)	0.06 (0.00 to 1.30)
3 studies	306 lesions	Median 7.6 Moderate	95.1% (92.2% to 97.9%)	77.8% (68.7% to 86.8%)	92.2% (88.7% to 95.7%)	85.1% (77.4% to 92.9%)	Heterogeneous, no summary estimate calculated	0.07 (0.05 to 0.10)

Figure 12. Diagnostic Test Characteristics of the Performance of Scintimammography in Evaluating Large Lesions



We combined the likelihood ratios calculated from the three studies of large lesions (Figure 13.). The negative likelihood ratios were not hetereogeneous (I^2 = 0%), and the summary estimate was 0.07 (0.05 to 0.10). However, the estimate was not stable by cumulative meta-analysis, and therefore its stability is low. The positive likelihood ratios were heterogeneous (I^2 = 88.2%) and could not be directly combined. Only two of the three positive likelihood ratios fell within the possibly clinically useful range, and therefore the evidence is unacceptably weak for supporting a conclusion about the usefulness of scintimammography to rule in breast cancer. Because this indication is not directly relevant to the purposes of this report, we did not attempt to explore the heterogeneity further.

Figure 13. Likelihood Ratios of Scintimammography to Evaluate Large Lesions



nLh = negative likelihood ratio
pLH = positive likelihood ratio
lnpLH = natural log of positive likelihood ratio
lnnLH = natural log of negative likelihood ratio

Other Patient Groups

Two or fewer studies reported outcomes for multiple different patient subgroups (see Table 6 and Table 34 in Appendix E. Evidence Tables). For all of these subgroups, the evidence base was found to be unacceptably weak for evaluation of test performance due to the small size of the evidence bases.

SPECT Imaging

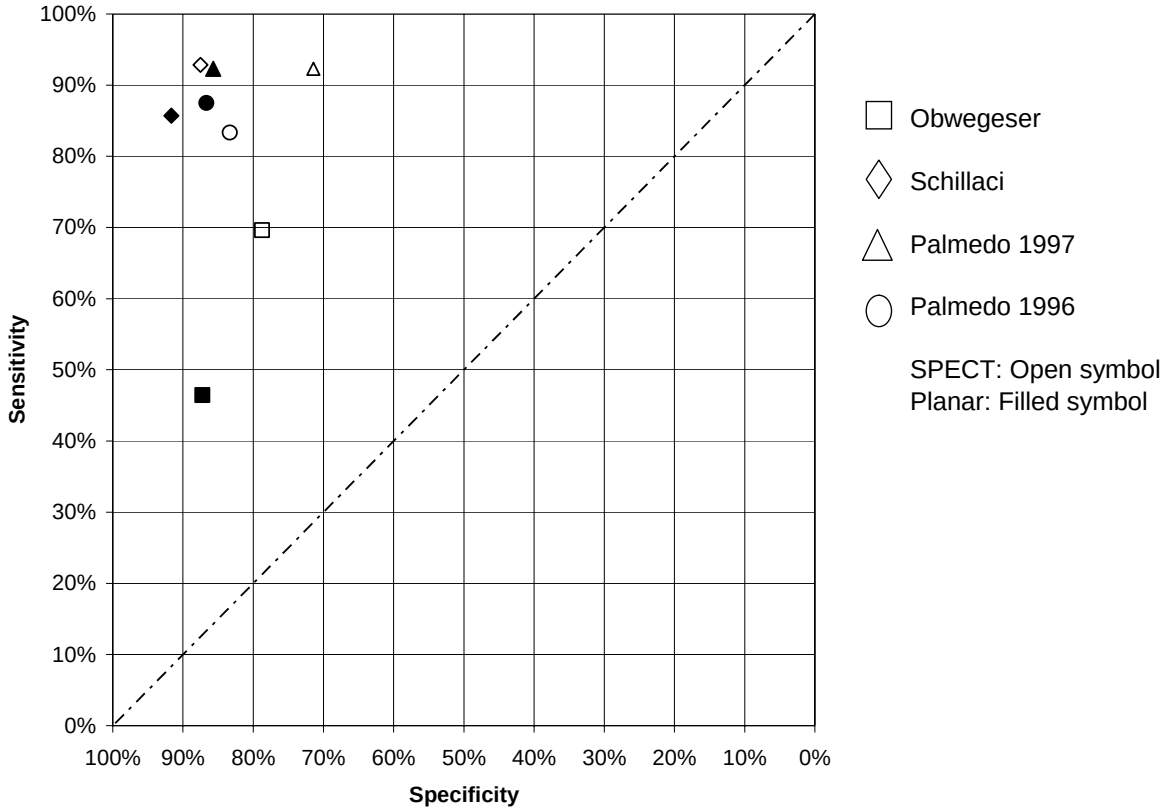
Four studies examined patients using both SPECT and planar imaging methods and reported the data separately for 243 lesions. For each data set, we calculated the ratio of the odds of receiving a false negative result on planar imaging and of the odds of receiving a false negative result on SPECT imaging (Table13; see also Table 42, in Appendix E. Evidence Tables). These odds ratios and their confidence intervals should not be interpreted to be precisely accurate estimates of the performance of these tests. Because the "control" group and the "experimental" group are the same patients, their measurements are not independent. Lack of independence violates many of the statistical assumptions underlying standard analysis methods. However, we can use these odds ratios to roughly estimate the relative performance of each of these tests, and therefore, while we cannot arrive at a quantitative conclusion, we may be able to derive a qualititative conclusion.

Two of the studies found that planar was superior to SPECT, one found that SPECT was superior to planar, and one found no difference. Therefore the data are not qualitatively consistent or robust, and are unacceptably weak for reaching a conclusion about whether SPECT or planar imaging is less likely to miss cases of cancer. The paired diagnostic test characteristics are shown graphically in Figure 14.

Table 13. Analysis of Studies of Scintimammography: SPECT vs. Planar Imaging

Study	N lesions	Quality score	False negatives on SPECT	False negatives on planar	Odds Ratio (95% CI)	p value of difference
Obwegeser et al. 1999[90]	103	7.4	17	30	0.48 (0.25 to 0.94)	0.0327
Schillaci et al. 1997[92]	66	8.8	3	6	0.48 (0.11 to 2.0)	0.309
Palmedo et al. 1996[106,107]	54	7.9	4	3	1.36 (0.29 to 6.36)	0.697
Palmedo et al. 1997[58]	20	8.3	1	1	1.00 (0.058 to 17.12)	1.00
4 studies	243 lesions	Median 8.1 Moderate				

Figure 14. SPECT vs. Planar Diagnostic Test Characteristics



Key Question 3. Are there other factors which affect the accuracy or acceptability of scintimammography?

The following narrative review identifies and discusses key factors that experts in the field believe may affect the performance of scintimammography of the breast.

Equipment Differences

Gamma cameras contain camera heads with collimators that route, to multiple detectors, photons that are emitted from the patient. Gamma cameras are designed to perform either planar (two-dimensional) imaging of the three-dimensional radiopharmaceutical distribution or single photon emission tomography (SPECT). SPECT is a technique that may use one, two, or three heads to create a three-dimensional representation of the administered radiopharmaceutical. Most of today's gamma cameras are capable of both planar and SPECT operation. Some experts believe that SPECT breast imaging has a higher overall accuracy compared to planar breast imaging to differentiate malignant from benign breast lesions smaller than 10 mm.[114] However, the literature has not revealed widespread agreement for the use of SPECT imaging for overall breast cancer detection. Because no consensus has been reached regarding SPECT's utility, specific parameters for SPECT breast imaging have not been recommended.[115,116]

A general purpose gamma camera with the proper collimator may be used for nuclear medicine breast imaging. The collimator should be a low-energy, high-resolution (LEHR) type. LEHR collimators are used for imaging photon energies up to 1 MeV. The photopeak of ^{99m}Tc is 140-keV and a symmetric 10% energy window (±5%) should be centered over the 140-keV photopeak.[116] However, a dedicated gamma camera with a unique collimator-detector arrangement and breast compression capabilities may offer other advantages (e.g., improved signal-to-noise ratio and spatial resolution).[117]

Effect of Operator and Image Interpreter on Scintimammography

Nuclear medicine breast imaging is susceptible to a number of errors that could affect the accuracy of a diagnosis. These errors include infiltration of ^{99m}Tc-sestamibi around the injection site, improper patient positioning or patient motion, and cross talk from ^{99m}Tc-sestamibi uptake of the opposite breast. Nuclear medicine technologists performing breast imaging studies must be aware of the sources of these errors.[116]

Images should be interpreted by the physician from the gamma camera's display. This permits adjustment of image contrast if necessary. Additionally, grayscale rather than color imaging is preferred.[116] Interpreters should consider a focal area of increased uptake positive for cancer regardless of its intensity.[115] Greater accuracy may be obtained when differentiating malignant from benign breast lesions when uptake patterns and sizes of lesions are considered in addition to tracer uptake levels.[118]

Patient Position

Accurate prone positioning of the patient and the breast to be imaged is important. The breast to be imaged should be hanging down and a positioning device should be used to minimize

patient motion.[116,119] Because the prone position can be very uncomfortable to the patient, which may result in movement, a reduction in total exam time should be attempted.[120]

Lesion Size and Patient Characteristics

Scintimammography with ^{99m}Tc-sestamibi is useful for screening women with dense breasts but has limited spatial resolution for demonstrating cancers with diameters smaller than 10 mm.[116,121,122] Spatial resolution is important in breast imaging since the lesions are small. The detectability of lesions that are smaller than the spatial resolution is compromised by the partial volume effect and tissue attenuation.[119]

Sestamibi has a strong affinity for breast tumors, but may also accumulate in areas of inflammation or infection.[123] The sensitivity of scintimammography has been reported to be affected by type of tumor, size of tumor, and the phase of the menstrual cyle.[14] Scintimammography has been reported to be unaffected by the presence of a breast implant or by the density of the breast tissue.[14]

Patient Safety and Comfort

Intravenous injection of ^{99m} Tc-sestamibi has been associated with very few adverse reactions.[80] A case of a patient without a past history of allergies, who developed a rash following administration of ^{99m}Tc-sestamibi, has been reported in the literature.[124] Another study reports, in addition to rash development, patients experiencing a strange taste following injection of ^{99m}Tc-sestamibi.[125] The incidence of adverse reactions to radiopharmaceuticals has been found to be 1,000 times lower than that of x-ray contrast media and other drugs administered in a hospital setting.[126]

Other than removal of all clothing and jewelry above the waist, no special preparation is required of patients undergoing a nuclear medicine imaging study. Compared to other breast imaging procedures, nuclear medicine breast imaging takes longer to perform – forty minutes or more.[127] During a typical study, the patient is placed in a prone position with the breast to be imaged hanging down.[119] Although taut compression of the breast to be imaged is not required, prevention of cross-talk may require compression of the opposite breast.[115,116]

A review of the literature has not revealed risks to the patient. As long as routine practices are followed, nuclear medicine breast imaging can be considered a safe exam for most patients.

Accreditation Factors

The Intersocietal Commission for the Accreditation of Nuclear Medicine Laboratories (ICANL) offers voluntary accreditation to facilities based on a peer review of their staff's qualifications, education, equipment, quality control, and volume of clinical procedures.[76]

All medical and technical staff are required to meet specific minimum experience and education requirements in order for their facility to be accredited by ICANL. Options available to a facility's medical staff range from board certification in nuclear medicine to board certification in a specialty area with a minimum number of years practice and volume of studies interpreted.

The accreditation program requires the technical director and all technologists working in the facility to hold the RT(N) credential from the American Registry of Radiologic Technologists

(ARRT) or the CNMT credential from the Nuclear Medicine Technology Certification Board (NMTCB). In all situations, the physician is ultimately responsible to see that the appropriate images are obtained.

Magnetic Resonance Imaging

Magnetic resonance imaging (MRI) has been used for many medical applications since 1985. MR images are created by recording the signals generated after radio frequency excitation of nuclear particles in tissue exposed to a strong magnetic field. A contrast agent, injected into the bloodstream, accumulates in the vascular system and can aid in locating tumors by highlighting areas containing a dense blood vessel network.

Summary

We found that for suspicious lesions in general, at a fixed 95% sensitivity, the specificity of MRI was 62.8%. At the mean threshold of the studies, the sensitivity was 92.5%, the specificity was 72.4%, and the negative predictive value was 90.5% (for a population with a prevalence of disease of 50.3%). For lesions with microcalcifications, our analysis found that the sensitivity of MRI was 85.9%, the specificity was 75.5%, and the negative predictive value was 84.7% (for a population with a prevalence of disease of 50.3%). The stability of these estimates was moderate for lesions in general and low for lesions with microcalcifications, indicating a small (for lesions in general) or reasonable (for microcalcifications) chance that publication of new evidence could substantially change these summary estimates.

Our analysis found that the negative likelihood ratio of MRI to evaluate women referred for further evaluation of the breast was 0.16. Our findings indicate that if a woman with a suspicious lesion tests negative for breast cancer by MRI, her chance of actually having breast cancer drops from 20%[3] to 3.8%.

The authors of two other technology assessments concluded that the negative predictive value of MRI is too low for routine use to rule out breast cancer after detection of a possible abnormality. Our results agree with this conclusion.

The studies, patients, and analyses used to reach these conclusions are described in the text that follows this section. The performance of MRI in evaluating women with suspicious breast lesions is summarized in Table 14.

[3] A woman in the general screening population who has a positive finding on mammography and/or physical examination has an approximately 20% chance of having breast cancer.

Table 14. Summary Test Performance of MRI

Patient subgroup	N studies	N lesions		Sensitivity	Specificity	Prevalence (Range)	PPV	NPV	+ LHR (95% CI)	- LHR (95% CI)
Suspicious breast lesion	10	1289	At mean threshold At 95% sensitivity STE	92.5% 95.0% Moderate	72.4% 62.8% Moderate	50.3% (25.9% to 73.1%)	77.2% 72.1% Moderate	90.5% 92.5% Moderate	Possibly clinically useful. SOE: Moderate	0.16 (0.13 to 0.19) STE: Low
Micro-calcifications	3	474	Summary estimate STE	85.9% (81.5% to 90.4%) Low	75.5% (70.1% to 80.9%) Low	49.3% (42.3% to 64.4%)	77.3% (71.9% to 82.7%) Low	84.7% (80.1% to 89.2%) Low	Possibly clinically useful. SOE: Moderate	0.20 (0.15 to 0.25) STE: Low
BIRADS 4 or 5	3	201		Too weak		59.7% (36.0% to 78.7%)	Too weak		Possibly clinically useful. SOE: Moderate	Too weak
No micro-calcifications	1	470	Too weak							
Palpable lesion	2	388	Too weak							
Non-palpable lesion	2	487	Too weak							
Indeterminate mammogram or physical exam	1	68	Too weak							
Positive by fine needle aspiration	2	225	Too weak							
Lesions >1.0 cm	2	38	Too weak							
Lesions ≤1.0 cm	1	23	Too weak							
Patients <65 yrs of age	2	77	Too weak							
Premenopausal patients	1	346	Too weak							
Post-menopausal patients	1	474	Too weak							

PPV = positive predictive value NPV = negative predictive value LHR = likelihood ratio
Too weak = The evidence base is unacceptably weak for calculation of summary diagnostic test characteristics or evaluation of test performance
SOE = Strength of evidence
STE = Stability of estimate

Analysis and Results for Key Questions 1. and 2. Diagnostic Test Characteristics, Predictive Values, and Likelihood Ratios of MRI

Other Published Technology Assessments

We identified two systematic reviews of the use of MRI to evaluate women with prior clinical findings that suggest the possibility of breast cancer. The methods and conclusions of these reviews are summarized in Table 15. The authors of both systematic reviews concluded that the negative predictive value of MRI is too low for this indication, and therefore patients did not benefit from being examined by MRI after mammography.

Table 15. Other Published Technology Assessments of MRI

Study	Methods	Conclusions
The Blue Cross/ Blue Shield Technology Evaluation Program, 2004[128]	Systematic review of the literature on the use of MRI to evaluate suspicious breast lesions in order to avoid biopsies. The review included 25 prospective studies and 14 retrospective studies. Reported data were described and a small, informal cost-benefit analysis was performed.	Reported sensitivity for MRI ranged from 91% to 99%; specificity ranged from 31% to 91%; and negative predictive value ranged from 56% to 99%. The authors of the review pointed out that in many of the populations studied, small breast lesions had been specifically excluded, and therefore the diagnostic performance of MRI in the clinic, where smaller lesions are often encountered, may be less accurate than predicted from these studies. The authors of the review performed a small, informal cost-benefit analysis and concluded that the negative predictive value of MRI was too low, even under the best possible conditions, to recommend the use of MRI for this indication. The potential benefit of sparing patients from unnecessary biopsy was not found to outweigh the potential harm of missed or delayed diagnosis of breast cancer.
Hrung et al. 1999[129]	A systematic review focused on women presenting with either a lesion that was palpably abnormal, or a BIRADS category 4 lesion detected by mammography. The review included 16 studies published between 1994 and 1997. Quality of the studies was rated on a 10-point scale (1 = highest quality, 10 = poorest quality). The data from the included studies was combined meta-analytically using the method of Littenburg and Moses.[81] The authors then conducted a cost-effectiveness analysis.	The mean quality score of the included studies was 3.0, indicating low quality. The optimal operating point of MRI, chosen to have a sensitivity of 95%, was found to have a specificity of 67%. Breast MRI is cost-effective relative to needle core biopsy only if MRI performance achieves a sensitivity and specificity of 93%, and needle core biopsy performance is less than the best available estimates. Therefore, the authors concluded that choosing needle core biopsy instead of MRI both increased patients QALYs and lowered the average cost per patient.

Studies

Nineteen diagnostic cohort studies of MRI met the inclusion criteria. Characteristics of the studies and included patients are summarized in Table 43, in Appendix E. Evidence Tables. The overall quality of the studies was moderate (median score of 7.8, range 6.4 to 8.3). A common shortcoming of these studies was not reporting whether readers of tests were blinded to patient information or the results of other tests.

The enrolled patients were incompletely described. Most of the studies enrolled any patients referred for biopsy due to a suspicious lesion discovered on physical exam or mammography. A few studies had more specific enrollment criteria, such as the presence of microcalcifications, specific BIRADS categories after mammography, or specific lesion sizes. The patients ranged in age from 18 to 85. Reported mean ages ranged from 48.4 to 58, suggesting that the patient populations studied are younger than the typical breast cancer population. Only one of the 19 studies reported information on the percentage of patients 65 years of age or older (8% of 49 patients; Imbriaco et al. 2001[98]).

The diagnostic test characteristics we calculated from data reported by each of the studies are presented in Table 45, in Appendix E. Evidence Tables.

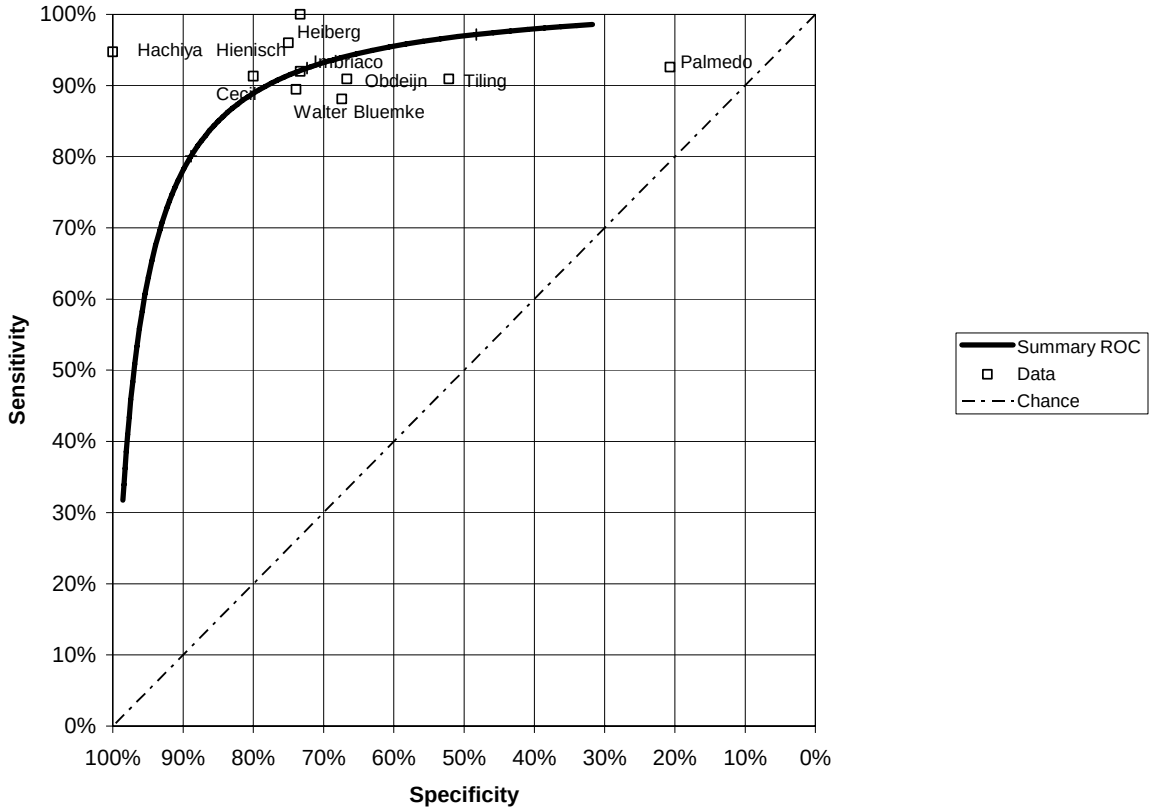
Patients Referred for Evaluation of Suspicious Breast Lesions

Ten studies reported results for 1289 lesions in patients referred for further evaluation of a suspicious breast lesion (abnormal mammogram and/or abnormal physical examination and/or abnormal ultrasound examination), listed in Table 16. The data were not heterogeneous (I^2 = 34%), so we combined them meta-analytically (Table 46 in Appendix E. Evidence Tables). The SROC curve is shown in Figure 15. At the mean threshold used in these studies, the sensitivity of MRI was 92.5% and the specificity was 72.4%. At a fixed sensitivity of 95%, the specificity was 62.8%, the positive predictive value was 72.1%, and the negative predictive value was 92.5%. The prevalence of disease in this population was 50.3%. The evidence base is of moderate quality (median score 7.6), and the summary diagnostic odds ratio was stable as tested by cumulative meta-analysis. Therefore our estimates of diagnostic test performance are of moderate stability.

Table 16. Studies of MRI for Evaluation of Suspicious Breast Lesions

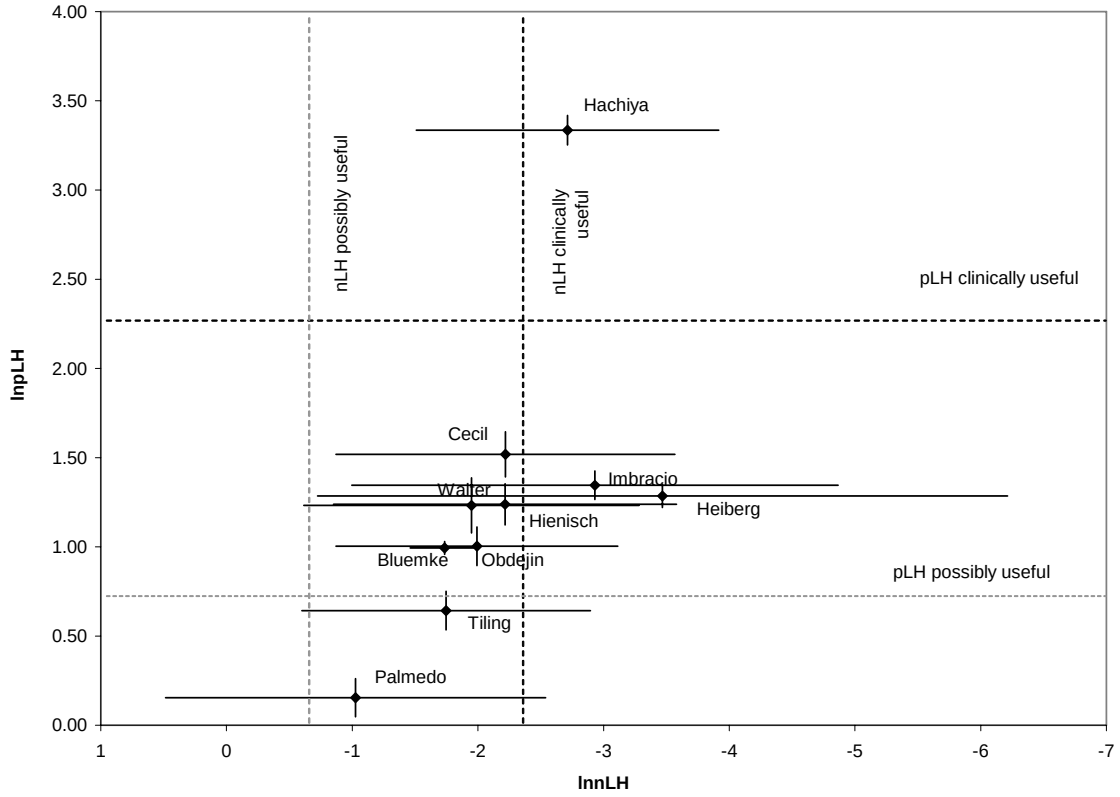
Study	N lesions	Quality score	Sensitivity (95% CI)	Specificity (95% CI)	Positive predictive value (95% CI)	Negative predictive value (95% CI)	Positive likelihood ratio (95% CI)	Negative likelihood ratio (95% CI)
Bluemke et al. 2004[130]	821	7.8	88.1% (84.6 to 91.1)	67.4% (62.7 to 71.9)	72.4% (68.2 to 76.3)	85.4% (81.1 to 89.0)	2.70 (2.61 to 2.80)	0.18 (0.13 to 0.23)
Heiberg et al. 1996[131]	81	6.4	100.0% (84.2% to 99.8%)	73.3% (60.9% to 82.8%)	56.8% (40.9% to 71.3%)	100.0% (91.8% to 99.9%)	3.61 (3.39 to 3.85)	0.03 (0.00 to 0.48)
Tiling et al. 1997[95]	56	7.4	90.9% (76.3% to 96.7%)	52.2% (33.0% to 70.7%)	73.2% (58.0% to 84.2%)	80.0% (54.7% to 92.7%)	1.90 (1.71 to 2.12)	0.17 (0.06 to 0.55)
Palmedo et al. 1996[96]	56	7.9	92.6% (76.4% to 97.8%)	20.7% (10.0% to 38.5%)	52.1% (38.3% to 65.5%)	75.0% (40.8% to 92.5%)	1.17 (1.05 to 1.30)	0.36 (0.08 to 1.62)
Obdejin et al. 1996[132]	54	7.9	90.9% (76.3% to 96.7%)	66.7% (45.3% to 82.7%)	81.1% (65.7% to 90.4%)	82.4% (58.8% to 93.6%)	2.73 (2.45 to 3.04)	0.14 (0.04 to 0.42)
Hachiya et al. 1991[133]	52	7.5	94.7% (82.5% to 98.4%)	100.0% (78.0% to 99.8%)	100.0% (90.1% to 99.9%)	87.5% (63.7% to 96.3%)	28.08 (25.86 to 30.48)	0.07 (0.02 to 0.22)
Imbracio et al. 2001[98]	49	7.6	96.0% (80.2% to 99.1%)	75.0% (55.0% to 87.8%)	80.0% (62.6% to 90.4%)	94.7% (75.1% to 98.8%)	3.84 (3.54 to 4.16)	0.05 (0.01 to 0.37)
Walter et al. 2003[57]	42	7.9	89.5% (68.4% to 96.8%)	73.9% (53.4% to 87.3%)	73.9% (53.4% to 87.3%)	89.5% (68.4% to 96.8%)	3.43 (2.94 to 4.00)	0.14 (0.04 to 0.54)
Hienisch et al. 2003[56]	40	7.2	92.0% (74.8% to 97.6%)	73.3% (48.0% to 88.9%)	85.2% (67.4% to 93.9%)	84.6% (57.6% to 95.4%)	3.45 (3.07 to 3.87)	0.11 (0.03 to 0.43)
Cecil et al. 2001[134]	38	7.2	91.3% (73.0% to 97.4%)	80.0% (54.7% to 92.7%)	87.5% (68.8% to 95.5%)	85.7% (59.8% to 95.7%)	4.57 (4.02 to 5.18)	0.11 (0.03 to 0.42)
10 studies	1289 lesions	median 7.6 Moderate	At mean threshold 92.5%	At mean threshold 72.4% At 95% sensitivity 62.8%	At mean threshold 77.2% At 95% sensitivity 72.1%	At mean threshold 90.5% At 95% sensitivity 92.5%	Heterogeneous, no summary estimate calculated	0.16 (0.13 to 0.19)

Figure 15. SROC of MRI for Evaluation of Suspicious Breast Lesions



We calculated the summary likelihood ratios from the ten studies of patients referred for evaluation of suspicious breast lesions, shown graphically in Figure 16. The data for the negative likelihood ratios were not heterogeneous ($I^2 = 0\%$), so we meta-analytically combined them (Table 46, in Appendix E. Evidence Tables). The summary negative likelihood ratio is 0.16 (95% CI 0.13 to 0.19). The evidence base is of moderate quality, but the summary negative likelihood ratio was not stable as tested by cumulative meta-analysis. Therefore the stability of the summary estimate is low. The data for the positive likelihood ratio were heterogeneous ($I^2 = 97\%$), and therefore we did not attempt to compute a summary estimate. However, eight out of ten of the individual studies positive likelihood ratios are within the possibly/clinically useful range (2 or greater) and, therefore a qualititative conclusion that MRI is possibly clinically useful in detecting breast cancer, supported by a moderate strength of evidence, can be reached. Because positive likelihood ratios are not directly relevant to the purpose of MRI for this report, we did not attempt to explain the heterogeneity further.

Figure 16. Likelihood Ratios of MRI for Evaluation of Suspicious Breast Lesions



nLh = negative likelihood ratio
pLH = positive likelihood ratio
lnpLH = natural log of positive likelihood ratio
lnnLH = natural log of negative likelihood ratio

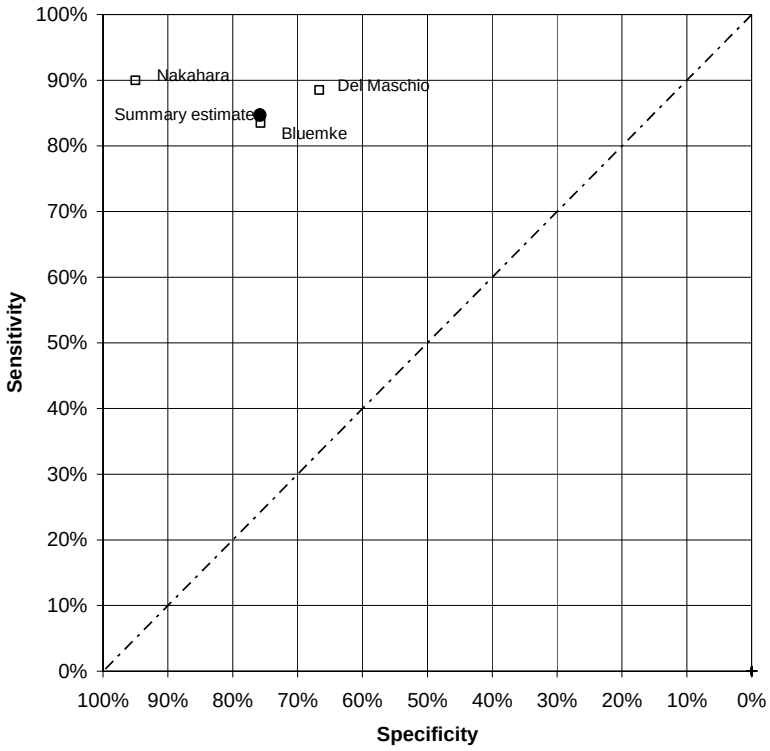
Lesions with Microcalcifications

Three studies reported results of MRI examination of 474 lesions with microcalcifications (Table 17). The data were not hetereogeneous (I^2 = 28%). Because there were only three studies, we did not compute a SROC. Instead, we separately combined the test characteristics from these three studies using meta-analysis (Table 47, in Appendix E. Evidence Tables). The pooled sensitivity was 85.9% (81.5% to 90.4%), and the pooled specificity was 75.5% (70.1% to 80.9%), as is shown in Figure 17. The pooled positive predictive value is 77.3% and the pooled negative predictive value is 84.7%. The prevalence of disease in this population was 49.3%. The evidence base is of moderate quality (median score 7.8), but the summary test characteristics were not stable as tested by cumulative meta-analysis. Therefore the estimates of diagnostic test performance are of low stability.

Table 17. Studies of MRI for Lesions with Microcalcifications

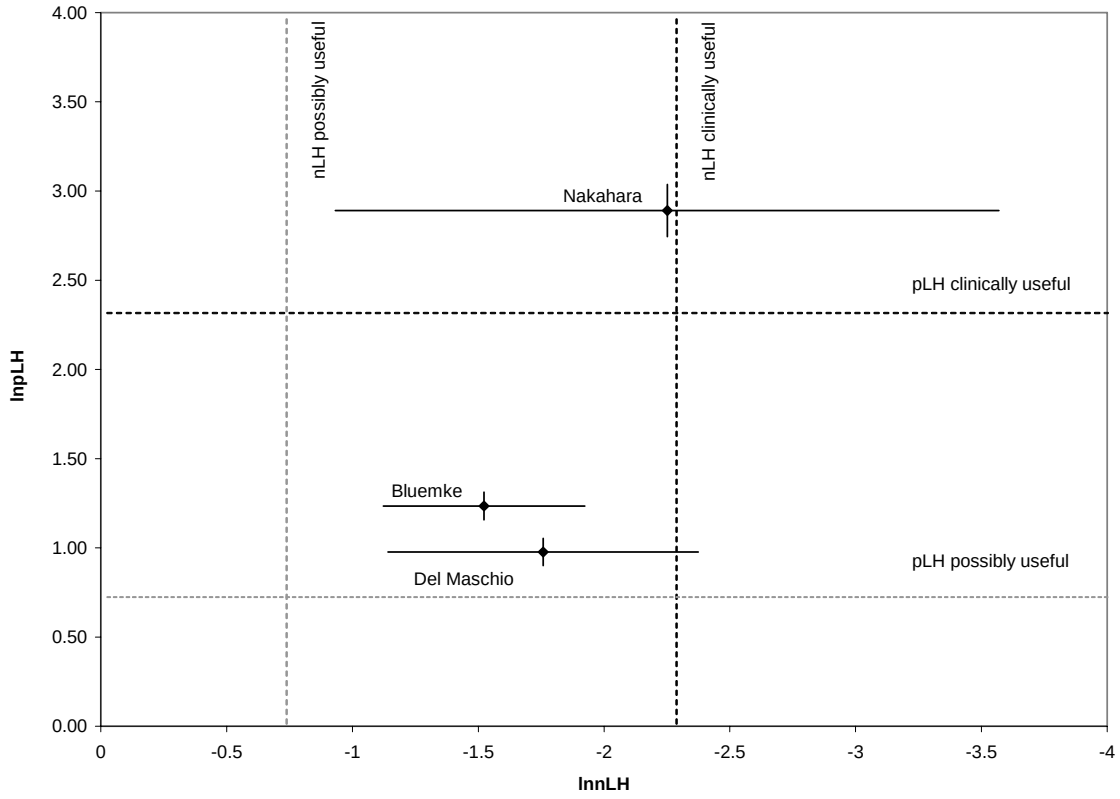
Study	N lesions	Quality score	Sensitivity (95% CI)	Specificity (95% CI)	Positive predictive value (95% CI)	Negative predictive value (95% CI)	Positive likelihood ratio (95% CI)	Negative likelihood ratio (95% CI)
Bluemke et al. 2004[130]	300	7.8	83.5% (75.8 to 89.5)	75.7% (68.6 to 81.9)	71.6% (63.6 to 78.7)	86.2% (79.7 to 91.2)	3.44 (3.18 to 3.71)	0.22 (0.15 to 0.33)
Del Maschio et al. 2002[135]	134	6.5	88.5% (80.1% to 93.6%)	66.7% (52.5% to 78.3%)	82.8% (73.8% to 89.1%)	76.2% (61.4% to 86.4%)	2.66 (2.46 to 2.86)	0.17 (0.09 to 0.32)
Nakahara et al. 2001[136]	40	7.9	90.0% (69.7% to 97.0%)	95.0% (76.1% to 98.9%)	94.7% (75.1% to 98.8%)	90.5% (70.9% to 97.1%)	18.00 (15.55 to 20.83)	0.11 (0.03 to 0.39)
3 studies	474 lesions	median 7.8 Moderate	Summary estimate 85.9% (81.5 to 90.4)	Summary estimate 75.5% (70.1 to 80.9)	Summary estimate 77.3% (71.9 to 82/7)	Summary estimate 84.7% (80.1 to 89.2)	Heterogeneous, no summary estimate calculated	Summary estimate 0.20 (0.15 to 0.25)

Figure 17. Sensitivity and Specificity of MRI for Lesions with Microcalcifications



We calculated the summary likelihood ratios from the three studies of patients with lesions with microcalcifications. The results of the analysis are shown in Figure 18. The data for the positive likelihood ratio were heterogenous (I2 = 81.4%) and therefore were not directly combined. However, all three of the positive likelihood ratios (100%) were within the possibly clinically useful range (2 or larger). The data for the negative likelihood ratio were not heterogeneous (I2 = 0.0%). The summary negative likelihood ratio is 0.20 (0.15 to 0.25). This evidence base is of moderate quality, but the summary negative likelihood estimate was not stable by cumulative meta-analysis, and therefore the estimate of the summary negative likelihood ratio is of low stability.

Figure 18. Likelihood Ratios of MRI for Lesions with Microcalcifications



nLh = negative likelihood ratio
pLH = positive likelihood ratio
lnpLH = natural log of positive likelihood ratio
lnnLH = natural log of negative likelihood ratio

Lesions Scored BIRADS 4 or 5 after Mammography

Three studies reported outcomes for 201 lesions scored as BIRADS 4 or 5 after mammography (Table 18). The data were heterogeneous (I^2 = 70%) and therefore we did not combine them (Table 48, in Appendix E. Evidence Tables). Because there are only three studies, meta-regression cannot be performed to explore the heterogeneity. Therefore the evidence is unacceptably weak for calculating summary diagnostic test characteristics. All three studies reported a positive likelihood ratio within the possibly clinically useful range (2 or larger). However, only one of these studies reported a negative likelihood ratio that was within the possibly clinically useful range (0.5 or less). From this, we can conclude that MRI may be possibly clinically useful for detecting cancer when evaluating lesions scoring a BIRADS 4 or 5 after mammography, but the evidence is unacceptably weak to support a conclusion as to whether MRI may be clinically useful for ruling out cancer when evaluating lesions scoring a BIRADS 4 or 5.

Table 18. Studies of MRI for Lesions of BIRADS 4 or 5

Study	N lesions	Quality score	Sensitivity (95% CI)	Specificity (95% CI)	Positive predictive value (95% CI)	Negative predictive value (95% CI)	Positive likelihood ratio (95% CI)	Negative likelihood ratio (95% CI)
Malich et al. 2001[137]	90	7.9	98.1% (90.1% to 99.6%)	80.6% (64.9% to 90.1%)	88.3% (77.7% to 94.1%)	96.7% (83.1% to 99.3%)	5.05 (4.87 to 5.24)	0.02 (0.00 to 0.16)
Knopp et al. 2003[138]	61	8.3	70.8% (56.8% to 81.7%)	76.9% (49.6% to 91.6%)	91.9% (78.5% to 97.1%)	41.7% (24.6% to 61.2%)	3.07 (2.56 to 3.68)	0.38 (0.22 to 0.65)
Huange et al. 2004[139]	50	7.1	100.0% (82.0% to 99.8%)	62.5% (45.2% to 77.0%)	60.0% (42.3% to 75.3%)	100.0% (83.5% to 99.8%)	2.57 (2.39 to 2.77)	0.04 (0.00 to 0.66)
3 studies	201 lesions	median 7.9 Moderate	Heterogeneous, no summary estimates calculated.					

Other Patient Groups

Two or fewer studies reported results for other patient subgroups (see Table 14 and Table 45 in Appendix E. Evidence Tables). For all of these subgroups, the evidence bases were unacceptably weak for evaluation of test performance due to their small sizes.

Key Question 3. Are there other factors that affect the accuracy or acceptability of the test?

The following narrative review identifies and discusses key factors that experts in the field believe may affect the performance of MRI of the breast.

Equipment Differences

The capability of MRI technology is constantly being improved. Apart from the room and the magnet itself, most other MRI components can be replaced and upgraded without major disruptions to service. However, an upgrade can entail considerable cost, and the long useful lifespan of MRI equipment can discourage frequent upgrading of core equipment. Therefore, age of an MRI installation is not necessarily an indicator of outdated equipment. A wide range of equipment performance exists between MRI installations.

Comparing equipment specification and performance is not straightforward in MRI due to the complex relationship between the factors that control image quality.[140] The following equipment-related factors potentially affect the performance of an MRI system used in breast imaging: magnet field strength, temporal and spatial resolution, surface coils, fat suppression, and contrast media.

Magnet Strength

Magnet strength varies from 0.5T to 3.0T. The primary reason for using higher field strength is to increase the signal available. Increasing the signal allows higher spatial resolution images or faster image acquisition times. However, the magnet strength also modifies the magnetic relaxation properties and therefore the contrast in images. Therefore image acquisition parameters must be significantly adjusted when using different field strengths. Also, these factors must be recognized when comparing the diagnostic performances as a function of field strength. MRI systems that use field strengths below 1 T are usually open gantries and are primarily used for patients who cannot be accommodated inside the bores of higher field strength magnets. Another advantage of the open magnets is the easier access when performing invasive procedures, such as biopsies. So, despite the drawbacks of lower field strength (reduced signal to noise ratio, longer imaging times, and unavailability of fat suppression), there is considerable interest in using low field-strength systems.[141] Overall, the use of 1.5 T and 1.0 T systems for breast MRI is well established, but lower field strength systems are not.

Temporal and Spatial Resolution

Once detected on MRI, two aspects of a breast lesion are of interest to a radiologist: the dynamics of the contrast agent in the lesion and its morphology. Therefore, both temporal and

spatial resolution are important. Since MRI entails a tradeoff between acquisition time and spatial resolution, determining the most effective tradeoff is necessary.[142]

Of interest in the dynamic study is the rate of contrast enhancement and subsequent wash-out. The time scale for this process is about eight minutes, with the first minute containing important information. Specificity is improved if the temporal resolution is increased from two-minutes to one-minute acquisitions.[143] Even more information would be available if the first tens of seconds after contrast enhancement could be captured.[144]

Commercial MRI systems today can achieve a one to two minute image acquisition time (i.e., temporal resolution) in bilateral breast studies.[145-147] To achieve this a 3D image T1 weighted gradient echo pulse sequence with contiguous 2 to 3 mm slices is used. Two dimensional acquisition would be faster, except the slices would not be contiguous.[145] Therefore, the ability to acquire 3D images is required for breast MRI. To obtain this level of temporal resolution requires rapid pulse sequences, which necessitate high specification field gradients (i.e., power and slew rate).[144] The precise specification is not crucial. Instead, it is preferred that a system can acquire full 3D (3mm slice width) images sets of both breasts (i.e., either axial or coronal views) with medium resolution (256 x 256 acquisition matrix) with a temporal resolution of two minutes, or faster.[144]

Also, in addition to the T1 weighted dynamic acquisition, a high spatial resolution T2 weighted image is useful to help distinguish morphology and improve overall interpretation.[146,148-150] Since the acquisition time is not critical for this, most MR equipment is capable of this type of acquisition.

Surface Coils

Surface coils are routinely used in MRI to increase the efficiency of signal detection and, by extension, the image quality. Dedicated breast coils have been available for some time and are considered a prerequisite for breast imaging.[151] The dedicated breast coils allow the patient to lie prone with her breasts in close proximity with the coils. Some coils contain some means to immobilize the breasts with compression. The compression reduces the volume to be imaged (and therefore reduces image acquisition time) and moves the coils closer to the tissue and helps prevent patient movement (so image quality is improved).[145,152]

Most dedicated breast coils enable bilateral imaging without having to reposition the patient and repeat the contrast administration. Bilateral imaging allows radiologists to compare each breast.[146] Therefore, bilateral breast coils are preferred.

The number of individual elements within a coil affects signal to noise ratio. In general, increasing the number of elements improves signal to noise ratio.[145,153,154] However, the possible gain in signal to noise ratio is limited because the coils must get smaller as the number of elements is increased. In breast coils, a two element (per breast) phased array is preferred.[145] Phased array coils are standard for modern dedicated breast coils. Also, multiple elements combined with multiple channel processing enables parallel imaging techniques to be used, which enable faster imaging. However, in breast imaging parallel imaging is not widely used.[155]

Fat Suppression

In contrast-enhanced breast MRI the issue of fat suppression is particularly important due to the fact that the signals from fat and the contrast enhanced lesions can be similar.[145] The fat

signal can be suppressed by simple image subtraction or with a fat-suppressing pulse sequence. The image subtraction technique simply uses a pre-contrast image to subtract the fat signal from the subsequent images. The difficulty with this approach is the susceptibility to artifacts caused by patient motion. Breast compression can be used to ameliorate the problem. Fat-suppression pulse sequences rely on either a saturation pulse, which prolongs the image time, or on a spectrally selective inversion technique.[156] The availability and effectiveness of a fat-suppression methodology will depend on the MRI system. For example, pulse sequence-based fat suppression requires high field homogeneity, which is compromised on open MRI systems.[157] So, pulse sequence derived fat suppression is not always possible, and the subtraction method must be used. The subtraction method is acceptable providing the patient can remain motionless for ten to fifteen minutes.

Contrast Media

A number of non-specific contrast agents have been used widely in MRI since the mid 1980s. Despite molecular differences, the resulting contrast enhancement and safety profile is similar for the commonly used gadolinium chelates.[158] However, a recently approved agent, gadobenate dimeglumine (MultiHance), has also been used. Gadobenate dimeglumine has approximately double the T1 relaxivity of standard gadolinium chelates, which, theoretically, translates to higher contrast enhancement.[138] A multi-center comparison of gadobenate dimeglumine and gadopentetate dimeglumine found that gadobenate dimeglumine may have advantages when assessing the contrast intensity time curves.[138] However, the authors point to the need for further studies due to an imbalance in the number of malignant and nonmalignatant lesions in each patient group. In another comparison of gadopentetate dimeglumine (0.1 mmol/kg) to gadobenate dimeglumine (0.05 to 0.2 mmol/kg), the authors found significantly higher radiologist image quality scores of breast vascularity for the gadobenate dimeglumine.[159] Therefore, whereas traditional contrast agents can be regarded as interchangeable and functionally equivalent, newer agents may be advantageous. Additional clinical trials are required to confirm these findings.

Effect of Operator and Image Interpreter on MRI

MR images are susceptible to a number of artifacts that could cause image distortion and false interpretations. In particular, breast MR images are prone to artifacts caused by sternal wires and prosthesis filling valves.[160] Also, respiratory motion can be a problem, although when the patient is prone the effect is reduced.[160] Technologists conducting MRI breast studies must be aware of the potential problems and be able to compensate accordingly.

Breast imaging is a specialized area for radiologists. Breast MR image interpretation is not simply a matter of characterizing the contrast enhancement curve.[161] Since breast MR images should be interpreted in conjunction with other imaging studies, familiarity with those images is required.[146]

Studies have shown that automating the analysis can complement the interpretation and increase accuracy and efficiency.[162-165] For example, one factor that helps decrease variability is the automatic positioning of the region used to calculate the contrast dynamics.[166] Various automated approaches are commercially available to aid radiologists. These computer-based tools may prove useful for a number of reasons; they could save radiologist time, reduce

subjectivity, and decrease inter- and intra-observer variability. So, the use of automatic image analysis reduces the MRI-specific experience necessary to interpret exams accurately and reproducibly in a shorter time.[163]

Protocol Selection

While all suppliers of MRI equipment provide suggested protocols for different examination types, it is common for users to customize these. The degree of protocol customization largely depends on the clinical users, both radiologists and technologists. Even in tightly controlled studies with a limited number of institutions all using equipment supplied by the same manufacturer, differences in technique were observed.[167] Therefore, standardization of protocols can be difficult. Sometimes protocols should be slightly modified to accommodate differences in patients.

Patient Position

Using dedicated breast coils helps optimize the positioning of the patient. However, it is important that the breast is pulled away from the chest wall and, if compression is used, that contrast enhancement is not impaired by impeded blood flow.[156,168]

Contrast Media Administration

When contrast is taken up by a lesion, one of three characteristic enhancement and wash-out curves are usually observed: continuous enhancement, rapid enhancement followed by a plateau, or rapid enhancement followed by rapid wash-out. Rapid wash-out is indicative of malignancy.[161] However, there is no clear division between the three responses. A prolonged administration or poor synchronization between the injection and image acquisition of contrast media may affect the contrast enhancement versus time curve.[156] So, minimizing the number of variables is preferred. Standardization of contrast injection and synchronization with image acquisition is best achieved with a powered injector.[146] Also, the operator must ensure that all preparation is carried out before the contrast administration. For example, all patient-specific tuning and shimming must be made before administering the bolus of contrast agent to avoid delays.[145]

MRI does not directly visualize the contrast agent; instead the effects of the contrast agent on the magnetic properties of the tissue are visualized. Therefore, the relationship between contrast concentration and image quality is complex.[138,169] When using conventional gadolinium contrast agents, the dose does not appear to be a major factor when used in the normal range (0.1 to 0.2 mmol/kg). However, more data is required for newer contrast agents.

Patient Characteristics

In premenopausal women, the normal parenchyma can demonstrate enhancement that can decrease the specificity of breast MRI studies.[170,171] The amount of enhancement depends on the stage in the menstrual cycle. Therefore, an MRI study should be scheduled during the second week of the menstrual cycle when proliferative changes are at their lowest level and operators should verify this.

Patient Safety and Comfort

The fact that MRI does not use any ionizing radiation is one of the reasons for its widespread use. However, the magnetic field can cause problems and most exams require the patient to lie still for extended periods in a confined and loud environment. Also, all breast MRI exams require the administration of contrast media, but the gadolinium-based contrast agents used in MR imaging have been demonstrated to be safe and well tolerated.[144,172]

A number of hazards exist when a patient is undergoing an MRI exam. Examples include: patient heating, pacemaker malfunction, dislodgment of metallic implants, peripheral nerve stimulation, acoustic noise, and radio frequency induced burns.[173-178] Precautions are already taken at MRI facilities to routinely screen patients for possible contraindications. Patients are routinely asked to wear earplugs and are given an emergency call button. Technologists are trained to avoid situations in which burns may occur. Also, the FDA 510(k) guidelines for MRI equipment require that known hazards (e.g., patient heating) are accounted for in the system's design. However, it is important that users maintain up-to-date information as new devices and implants are used, and as MRI systems develop.[179,180] There is no reason to suspect that MR breast imaging is any different from other MRI exams with respect to safety.

Some aspects of the MRI environment, such as the safety of static magnetic fields and radio frequency radiation, remain issues of some debate. However, at the levels to which patients are exposed during routine MRI scanning, no conclusively determined serious adverse long-term effects have been identified.[181-184] Therefore, so long as routine precautions are followed, breast MRI can be considered a safe exam for most patients.

The requirement for a large-volume homogenous magnetic field for fat suppression means that closed-bore systems are preferred. Therefore, extreme claustrophobia and patient body habitus may prevent a small percentage of patients from undergoing breast MRI studies.[146] Dedicated breast coils allow most patients to be positioned comfortably. A search for reports of patient discomfort did not find any reports of severe discomfort. In fact, in order to decrease patient motion, it is important that the patient be as comfortable as possible.[160] Breast compression does increase the level of discomfort, but the amount is not significant, particularly when compared to the compression that is exerted during x-ray mammography exams.

Accreditation Factors

General-purpose MRI systems are cleared for marketing by United States Food and Drug Administration (FDA) under the 510(k) process. Accessories such as breast coils are cleared separately, also under the 510(k) process. Imaging devices are usually not cleared for specific indications; they are cleared for marketing for all indications in the entire body or in specified parts of the body.

There is no nationwide compulsory accreditation for MRI facilities. The American College of Radiology does administer a voluntary accreditation program.[185] However, the quality, capability, and performance of MRI systems varies widely. For example, when the American College of Radiology first started its voluntary MRI accreditation program, it reported that "59% of facilities failed to qualify (for accreditation) at their first try" due to poor image quality.[185] The failures were attributed to poor homogeneity and calibration. The daily tests used by some manufacturers are not sufficient to ensure acceptable image quality. Therefore, the ACR accreditation can help ensure acceptable image quality.

Ultrasound

Ultrasound waves are high-frequency sound waves that reflect at boundaries between tissues with different acoustic properties. Ultrasound is commonly used to distinguish between solid breast lesions and cysts and to guide biopsy needles;[186] these two uses are outside the scope of this report, and are not addressed here. Here, we address the use of ultrasound to examine solid lesions for signs of malignancy.

Summary

We found that for suspicious lesions in general, the sensitivity of ultrasound examination was 86.1%, the specificity was 66.4%, and the negative predictive value was 93.3% (for a population with a prevalence of disease of 25.7%). The stability of these estimates is moderate, indicating a small chance that publication of new evidence could substantially change these estimates.

Our analysis found that the negative likelihood ratio of ultrasound to evaluate women referred for suspicious breast lesions was 0.21. This result indicates that if a woman with a suspicious lesion is diagnosed as not having cancer by ultrasound, her chance of actually having cancer drops from 20%[4] to 5.0%.

The studies, patients, and analyses used to reach these conclusions are described in the text that follows this section. The performance of ultrasound in evaluating women with suspicious breast lesions is summarized in Table 19. The estimates of summary diagnostic test characteristics are of moderate or low stability, indicating that there is a small (moderate stability) to reasonable (low stability) chance that the magnitude of the estimate could substantially change as a result of the publication of new evidence. Therefore regular monitoring of the literature is recommended.

[4] A woman in thegeneral screening population who has a positive finding on mammography and/or physical examination has an approximately 20% chance of having breast cancer.

Table 19. Summary Test Performance of Ultrasound

Patient subgroup	N studies	N lesions	Sensitivity (95% CI)	Specificity (95% CI)	Prevalence (95% CI)	PPV (95%CI)	NPV (95% CI)	+ LHR (95% CI)	- LHR (95% CI)
Suspicious breast lesion	3	3,258	86.1% (83.8% to 88.5%) Stability of estimate: Moderate	66.4% (64.5% to 68.3%) Stability of estimate: Moderate	25.7% (24.2% to 27.2%)	47.0% (43.5% to 50.4%) Stability of estimate: Moderate	93.3% (92.3% to 94.3%) Stability of estimate: Moderate	Possibly clinically useful Strength of evidence: Moderate	0.21 (0.19 to 0.24) Stability of estimate: Low
Palpable lesion	5	2,090	Not calculated due to heterogeneity		23.7% to 73.1%	Not calculated due to heterogeneity	Not calculated due to heterogeneity	Possibly clinically useful Strength of evidence: Moderate	Possibly clinically useful Strength of evidence: Moderate
Lesions ≤10 mm	1	135	The body of evidence is unacceptably weak for evaluation of test performance.						
Lesions >10 mm	1	1,068	The body of evidence is unacceptably weak for evaluation of test performance.						
Patients >65 yrs of age	0	0	No evidence						
Patients <65 years of age	0	0	No evidence						

PPV = positive predictive value
NPV = negative predictive value
LHR = likelihood ratio

Analysis and Results for Key Questions 1. and 2. Diagnostic Test Characteristics, Predictive Values, and Likelihood Ratios of Ultrasound

Previously Published Technology Assessments

Flobbe et al. published a decision analysis model comparing different strategies for managing patients presenting with palpable breast masses in 2004.[187] Their decision model was based entirely on data from a single clinical study they previously authored (Flobbe et al.[188]). This particular clinical study by Flobbe et al. was excluded from the current report because it was confounded. Findings from the ultrasound exams influenced the way each patient was managed, including whether the patient was evaluated by biopsy. Therefore the data from Flobbe et al. can not be used to accurately estimate the diagnostic characteristics of ultrasound. Because the decision model developed by Flobbe et al. was based entirely upon this confounded study, the results of the decision model are also suspect and will not be discussed here.

Studies

Eight diagnostic cohort studies of ultrasound met the inclusion criteria. Characteristics of the studies and included patients are summarized in Table 49, in Appendix E. Evidence Tables. The quality of the studies was moderate (median score of 7.4, range 6.9 to 7.9). Despite the fact that the accuracy of ultrasound has been reported to be very operator- dependent, only one study (Hachiya et al.[133]) accounted for inter-operator differences in its design. Other common flaws in these studies were not reporting whether readers of tests were blinded to patient information or the results of other tests, not reporting sources of funding, and not reporting details about patient selection (whether the patients were consecutively enrolled or not).

The enrolled patients were incompletely described. Inclusion criteria were often incompletely described, and no demographic information aside from age was reported by any of the studies. Five of the eight studies included only women with palpable lesions. The patients ranged in age from 14 to 98. Reported mean ages range from 38.7 to 54, suggesting that the patient populations studied are younger than the typical breast cancer population. None of the studies reported information about the percentage of patients 65 years of age or older.

The diagnostic test characteristics we computed from data on each of these studies are summarized by study in Table 51, in Appendix E. Evidence Tables.

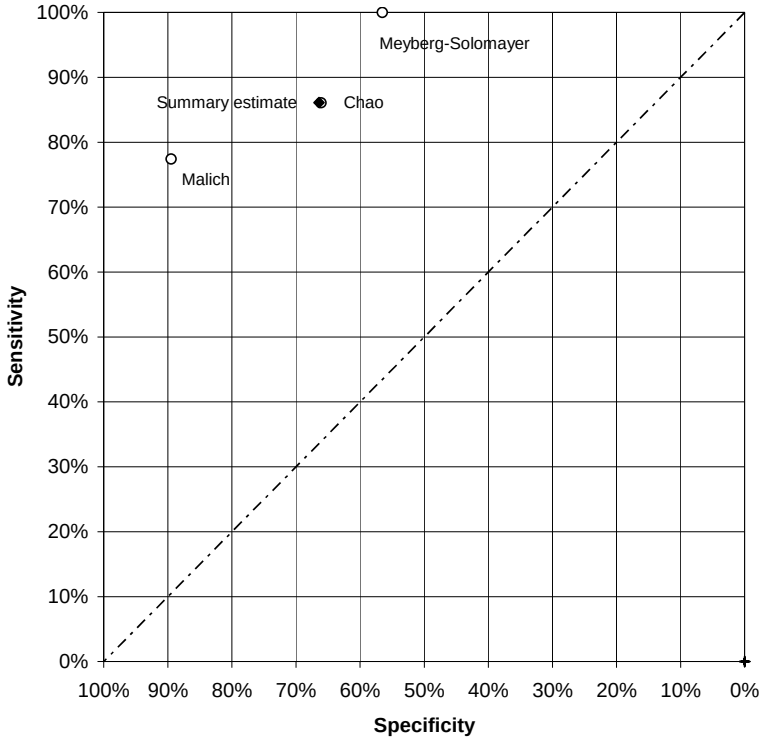
Patients Referred for Evaluation of Suspicious Breast Lesions

Three studies reported results for 3,258 lesions in patients referred for further evaluation of suspicious breast lesions (abnormal mammogram and/or physical examination), listed in Table 20. The data were not heterogeneous ($I^2 = 47\%$) but, because there are only three studies, we did not derive a SROC curve. Instead, we seperately combined the sensitivities and specificities from these studies (see Table 52 in Appendix E. Evidence Tables). The pooled sensitivity was 86.1% (83.8% to 88.5%) and the pooled specificity 66.4% (64.5% to 68.3%), as shown in Figure 19. The pooled positive predictive value is 47% and the negative predictive value is 93.3%. The prevalence in this population was 25.7%. This evidence base is of moderate quality, and the summary test characteristics were stable as tested by cumulative meta-analysis. Therefore the stability of the estimate of diagnostic test performance is moderate.

Table 20. Studies of Ultrasound for Evaluation of Suspicious Breast Lesions

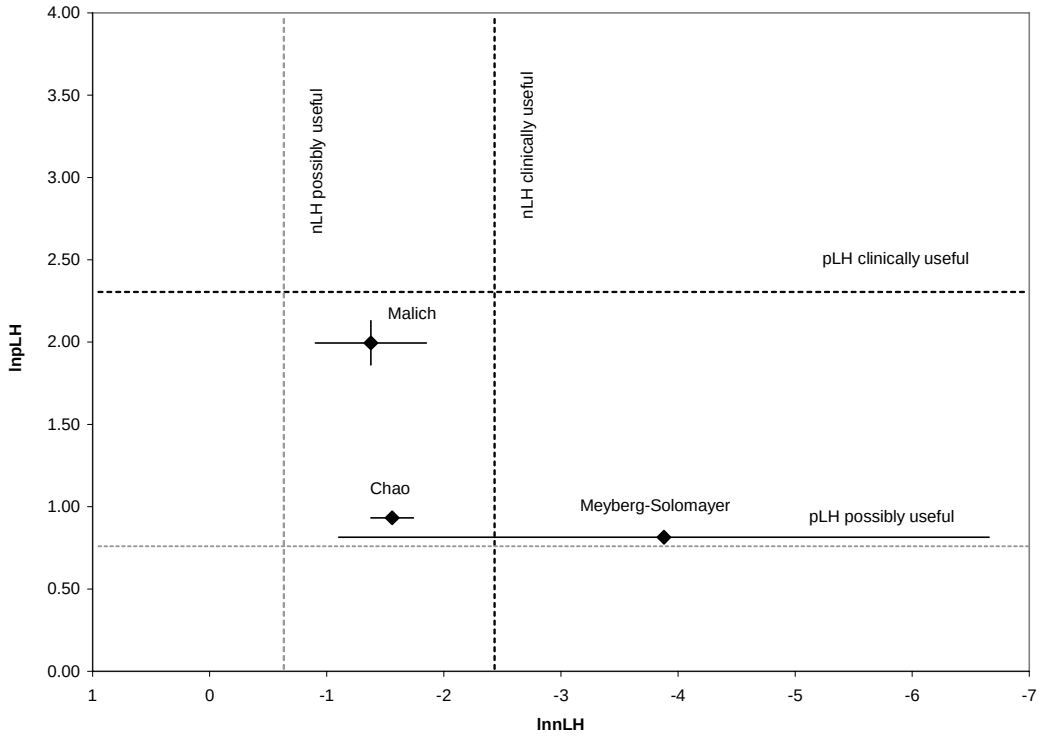
Study	N lesions	Quality score	Sensitivity (95% CI)	Specificity (95% CI)	Positive predictive value (95% CI)	Negative predictive value (95% CI)	Positive likelihood ratio (95% CI)	Negative likelihood ratio (95% CI)
Meyberg-Solomayer et al. 2004[189]	65	7.4	100.0% (91.4 to 99.9)	56.5% (36.8 to 74.3)	80.8% (68.0 to 89.1)	100.0% (76.7 to 99.8)	2.3 (2.2 to 2.3)	0.02 (0.00 to 0.33)
Malich et al. 2001[137]	100	7.9	77.4% (65.5 to 86.0)	89.5% (75.7 to 95.7)	92.3% (81.7 to 96.9)	70.8% (56.8 to 81.7)	7.4 (6.4 to 8.4)	0.25 (0.16 to 0.41)
Chao et al. 1999[190]	3,093	7.5	86.1% (83.4 to 88.4)	66.1% (64.2 to 68.0)	44.1% (41.5 to 46.7)	93.9% (92.6 to 94.9)	2.5 (2.5 to 2.6)	0.21 (0.18 to 0.25)
3 studies	3,258 lesions	median 7.6 Moderate	Summary estimate 86.1% (83.8 to 88.5)	Summary estimate 66.4% (64.5 to 68.2)	Summary estimate 47.0% (43.6 to 50.4)	Summary estimate 93.3% (92.3 to 94.2)	Heterogeneous, no summary estimate calculated	Summary estimate 0.21 (0.24 to 0.19)

Figure 19. Sensitivity and Specificity of Ultrasound for Evaluating Suspicious Breast Lesions



We calculated the summary likelihood ratios from the three studies of patients referred for suspicious breast lesions (see Figure 20). The data for the positive likelihood ratio were heterogenous (I^2 = 67%) and, therefore, we did not perform a meta-analysis. All three of the reported positive likelihood ratios were within the possibly/clinically useful range (2 or greater). The data for the negative likelihood ratio were not heterogenous (I^2 = 38%), and the summary negative likelihood ratio is 0.21 (0.24 to 0.19). However, the summary estimate was not stable by cumulative meta-analysis, so the stability of the estimate is low.

Figure 20. Likelihood Ratios of Ultrasound for Evaluating Suspicious Breast Lesions



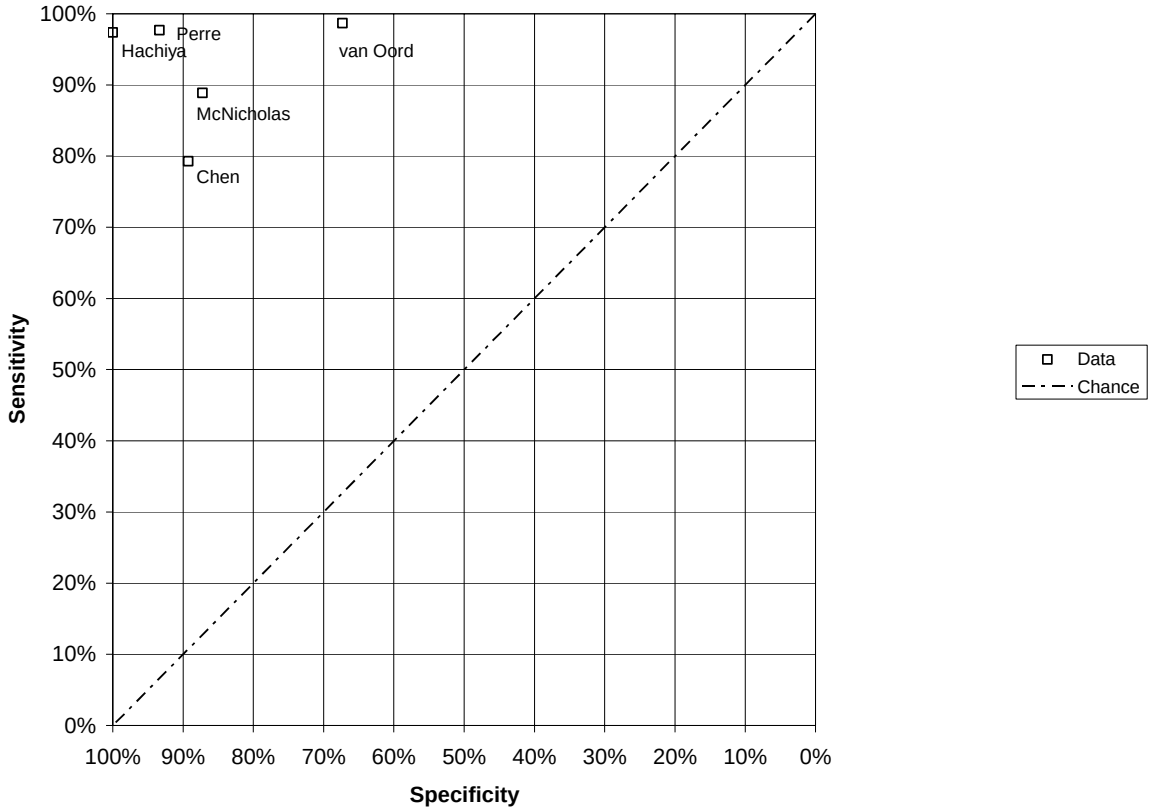
Patients Referred After Detection of Palpable Lesions

Five studies reported results for patients referred for further evaluation after discovery of 2,090 palpable breast lesions (listed in Table 21). The data were heterogeneous (I^2 = 90%) and therefore not directly combinable (Table 53, in Appendix E. Evidence Tables). We performed a meta-regression of the diagnostic odds ratio in order to identify possible causes of the heterogeneity. The results of the meta-regression, including the variables we examined, are summarized in Table 54, in Appendix E. Evidence Tables. None of the variables tested had a statistically significant correlation with the diagnostic odds ratio. Because the heterogeneity could not be explained, we did not compute a summary estimate of diagnostic test performance. The individual study estimates of diagnostic test performance are shown in Figure 21.

Table 21. Studies of Ultrasound for Palpable Lesions

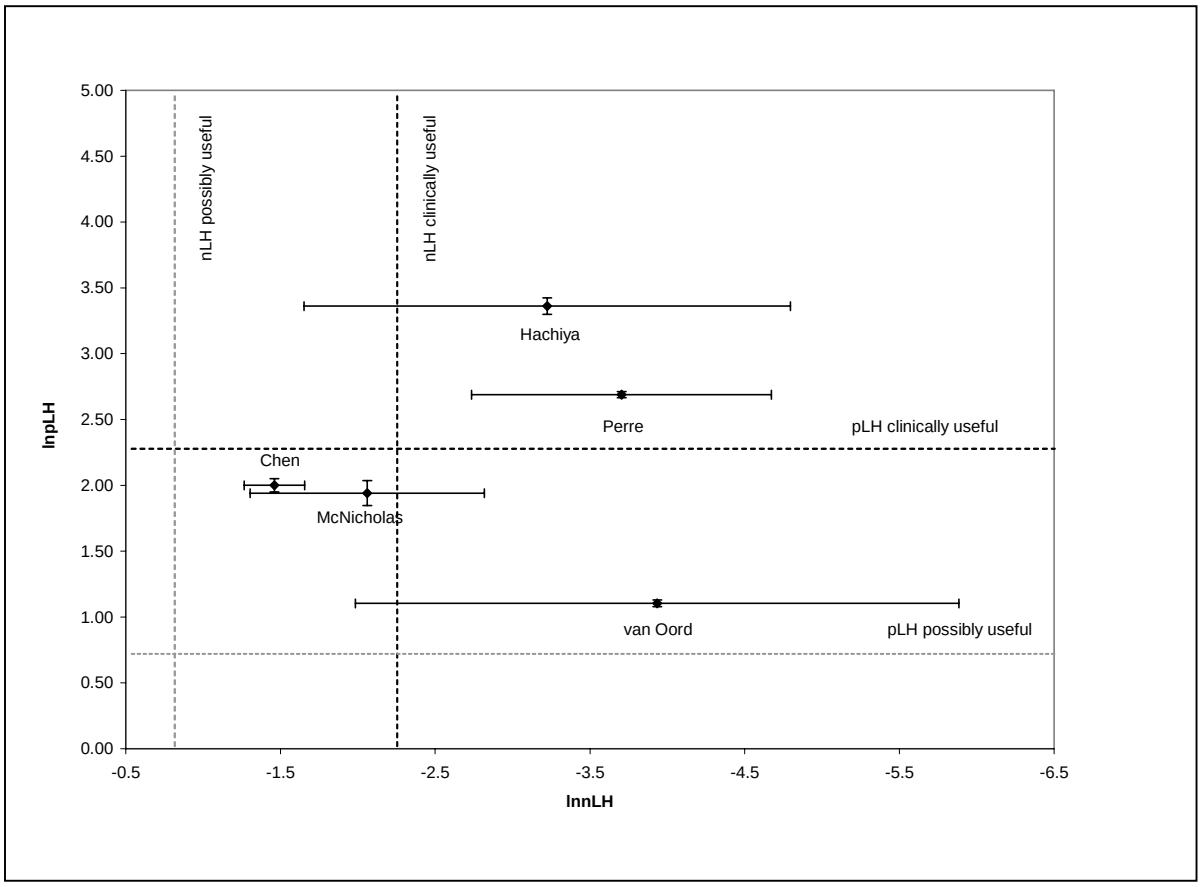
Study	N lesions	Quality score	Sensitivity (95% CI)	Specificity (95% CI)	Positive predictive value (95% CI)	Negative predictive value (95% CI)	Positive likelihood ratio (95% CI)	Negative likelihood ratio (95% CI)
Chen et al. 2004[191]	1,203	6.9	79.3% (75.0 to 83.0)	89.3% (87.0 to 91.2)	78.1% (73.8 to 81.9)	90% (87.7 to 91.8)	7.4 (7.1 to 7.8)	0.23 (0.19 to 0.28)
Perre et al. 1994[192]	400	7.4	97.7% (94.2 to 99.1)	93.4% (89.3 to 95.9)	91.9% (87.0 to 95.0)	98.1% (95.3 to 99.2)	14.7 (14.4 to 15.1)	0.02 (0.01 to 0.06)
van Oord 1991[193]	232	7.2	98.7% (92.8 to 99.7)	67.3% (59.6 to 74.2)	59.5% (50.8 to 67.7)	99.1% (94.8 to 99.8)	3.0 (2.9 to 3.1)	0.02 (0.00 to 0.14)
McNicholas 1993[194]	203	7.8	88.9% (77.7 to 94.7)	87.2% (80.9 to 91.6)	71.6% (59.9 to 81.0)	95.6% (90.7 to 97.9)	7.0 (6.3 to 7.7)	0.13 (0.06 to 0.27)
Hachiya 1991[133]	52	7.1	97.4% (86.3 to 99.4)	100.0% (78.0 to 99.8)	100.0% (90.3 to 99.9)	93.3% (69.9 to 98.6)	28.9 (27.1 to 30.7)	0.04 (0.01 to 0.19)
5 studies	2,090 lesions	median 7.2 Moderate	Heterogeneous, no summary estimates calculated.					

Figure 21. Diagnostic Test Characteristics of Ultrasound Examination of Palpable Lesions



The likelihood ratios from the five studies are shown in Figure 22. Due to the heterogeneity of the data, we did not attempt to pool the likelihood ratios. All of the data points and confidence intervals for both the positive and negative likelihood ratios are within the possibly/clinically useful ranges (2 or greater for the positive likelihood ratio, 0.5 or less for the negative likelihood ratio), suggesting that ultrasound may possibly be clinically useful for evaluating palpable lesions. The median quality of the studies is moderate, and the data are qualitatively robust, therefore the strength of the evidence is moderate.

Figure 22. Likelihood Ratios of Ultrasound for Palpable Lesions



nLh = negative likelihood ratio
pLH = positive likelihood ratio
lnpLH = natural log of positive likelihood ratio
lnnLH = natural log of negative likelihood ratio

Patients with Specific Lesion Sizes

Only one study reported results for lesions of different sizes. Because only one study of moderate quality reported data on these subgroups, the evidence was unacceptably weak for evaluation of test performance.

Patients Older than 65 Years of Age

None of the studies reported information on the percentage of patients older than 65, nor did any of the studies report diagnostic test characteristics for subgroups defined by age. No studies reported outcomes for any other subgroups of patients.

Key Question 3. Are there other factors which affect the accuracy or acceptability of ultrasound?

The following narrative review identifies and discusses key factors that experts in the field believe may affect the performance of ultrasound imaging of the breast.

Equipment Differences

The quality of a breast ultrasound image depends on the specifications of the ultrasound scanner; in particular, the type of transducer and the transducer's frequency.
Breast ultrasound requires a 7 MHz or higher-frequency linear transducer and electronic adjustment of focal zones(s) is recommended.[195,196] Transducers producing higher frequencies improve resolution. The transducer should be manufactured specifically for use in superficial imaging and a frequency of 10 MHz or 12 MHz is preferred.[197] Transducers with frequencies up to 12 MHz are available for breast imaging on most modern ultrasound scanners.

Linear array transducers are used for breast studies because they generate rectangular images. Compared with transducer configurations that produce sector, wedge-shaped images, linear array transducers produce images with a large field-of-view of areas close to the skin surface.

Effect of Operator and Image Interpreter on Ultrasound Imaging

The accuracy of ultrasound strongly depends on the skill of the sonographer/mammographer. Operators must continuously and carefully alter scanning parameters including transducer orientation, pressure, and instrument controls to avoid artifacts in ultrasound images, which can significantly degrade image quality and possibly lead to an incomplete or incorrect diagnosis.[197-199]

Ultrasound images are susceptible to a number of artifacts that could cause image distortion and false interpretations. Sonographers performing ultrasound breast imaging studies must be aware of the potential problems and be able to compensate accordingly. Two common types of artifacts are shadowing and reverberation.

Breast ultrasound images are prone to a form of shadowing caused by the complex combination of absorption and refraction along the border of a mass.[198] This edge shadowing can obscure the mass's lateral margins. Another shadowing form - acoustic shadowing - can completely conceal the posterior margin of a mass, making size measurements problematic.[200]

Although posterior acoustic shadowing is a feature most commonly associated with malignancy, this feature may also be seen with benign breast lesions.[201] Shadowing could also be caused by poor contact between the transducer and the skin[198]

Reverberation, which is the display of parallel echogenic lines caused by reflection of the ultrasound beam back and forth between the transducer and tissue interface, can give the false appearance of solid or complex material along the wall of a cyst.[198]

Aside from artifacts, a number of operator-controlled settings affect the diagnosis from a breast ultrasound study. Dynamic range settings determine the range of echo amplitudes detected and displayed. A low dynamic range setting increases image contrast but may cause the echoes in a solid mass to be absent from the display, thus mimicking a normally anechoic (echo-free) simple cyst. A high dynamic range setting results in an image with little contrast, which hinders differentiation of fat lobules from subtle masses.[198]

The focal zone is the location where the width of the ultrasound beam is narrowest and where the spatial resolution is greatest. Image characteristics outside the focal zone may cause a subtle mass to be less visible, sharp edges to appear ill-defined, and an anechoic simple cyst to have internal echoes.[198]

Settings of gray-scale gain determine the amplification of the returning echo signals. Inappropriately high gain settings could cause spurious echoes to be displayed in a simple cyst resulting in the appearance of a complex cyst or solid mass.[198,202,203]

Studies of inter-observer variability indicate a high degree of agreement when determining the shape of a mass while moderate agreement among observers is found when the degree of posterior acoustic transmission is described.[200] However, when the margins of lesions are poorly defined, or when shadowing is present, measurements of masses are difficult. Interobserver agreement is very high when ultrasound breast imaging is performed as adjunct to physical examination and mammography and with full knowledge of the mammographic findings.[204,205]

Several computer-aided diagnosis (CAD) systems have been developed to provide quantitative assessments and to eliminate variability among observers. The use of computer tools is expected to diminish operator dependency and to aid in diagnosis. Because the use of these systems is not widespread, diagnosis is still based on a subjective evaluation of the findings.[199,206] This further supports the need for well qualified interpreters.

Patient and Lesions Characteristics

Ultrasound is reported to function well in imaging breasts with dense tissue, which are often problematic to evaluate with x-ray mammography.[207,208] However, ultrasound has been reported to have a poor ability to detect microcalcifications in breast tissue.[207] Microcalcifications may be an important early indication of breast cancer.

Patient Safety and Comfort

During a typical ultrasound breast imaging study, the patient is placed in a supine oblique position, with a pillow under the shoulder and the arm extended behind the head.[209] Because taut compression is not required, ultrasound is advantageous, particularly when evaluating a painful breast.

A review of the literature has not revealed risks to the patient. As long as routine practices are followed, ultrasound breast imaging can be considered a safe exam for most patients.

Accreditation Factors

The American College of Radiology (ACR) has instituted a voluntary breast ultrasound accreditation program that offers facilities the opportunity for peer review of their staff's qualifications, equipment, and quality control and quality assurance programs.[196]

A physician supervising and interpreting breast ultrasound examinations is required to meet specific minimum experience and education requirements in order for their facility to be accredited by the ACR.

The accreditation program requires sonographers/mammographers to be certified by the American Registry of Diagnostic Medical Sonography (ARDMS), or post-primary certification ("advanced registry") in breast sonography by the American Registry of Radiologic Technologists (ARRT), or certification by the ARRT or unrestricted state license and qualified to do mammography under Mammography Quality Standards Act (MQSA). The physician is not required to be present during breast ultrasound examinations performed by ARDMS sonographers or ARRT technologists with certification in breast sonography. However, the physician must be in the department during breast ultrasound examinations performed by ARRT technologists without an advanced registry in breast sonography. In all situations, the physician is ultimately responsible to see that the appropriate images are obtained.

Overall Summary

Currently, biopsy is the only accepted method to accurately differentiate benign from malignant breast lesions. Although a biopsy is a safe procedure, patients may experience high anxiety, temporary loss of productivity, and various degrees of surgical trauma and cosmetic alteration; and each procedure places cost and personnel demands on the health care system. Reducing the number of biopsies through non-invasive technologies is desirable. However, how accurate does a test have to be before it becomes an acceptable and appropriate alternative to biopsy? This decision ultimately depends upon what society considers an acceptable rate of missed cases of cancer. An analysis by the Ontario Ministry of Health concluded that a negative predictive value of 98% or greater would be an acceptable level for a diagnostic test to reliably preclude breast biopsy.[80] With a negative predictive value of 98%, 20 cases of breast cancer would be missed in exchange for avoiding 980 unnecessary biopsies.

Our analysis found that in a patient population with a prevalence of breast cancer of 20%, MRI would miss 38 cases of breast cancer in exchange for avoiding 962 unnecessary biopsies; ultrasound would miss 50 cases of breast cancer in exchange for avoiding 950 unnecessary biopsies; PET scanning would miss 76 cases of breast cancer in exchange for avoiding 924 unnecessary biopsies; and scintimammography would miss 93 cases of breast cancer in exchange for avoiding 907 unnecessary biopsies. These numbers are calculated from the summary negative likelihood ratios, and assume 1,000 women with suspicious lesions diagnosed as cancer-free by each particular test will choose to forego biopsy. If a similar analysis is performed using the negative predictive value derived from the SROC curves, similar results are obtained. For example, for scintimammography, at the mean threshold for a population with a prevalence of 20%, the negative predictive value derived from the SROC curve is 91.6%, and the negative predictive value derived from the summary likelihood ratio is 90.7% (95% confidence interval 89.1 to 92.2). The minor difference in results between methods of analysis is within the expected error range.

The numbers above are based upon an average risk of breast cancer for a women referred for breast biopsy. However, an individual woman's risk of breast cancer in the face of a suspicious finding on mammogram or clinical examination may vary widely, depending upon the findings and her own situation; the extent of cancer risk should be discussed by the woman and her health care provider. In general, the higher a woman's risk of cancer before undergoing a non-invasive imaging test, the higher the risk that she has cancer even if the test is negative.

We used a prevalence of breast cancer of 20% in the above calculations because Banks et al. reported a prevalence of breast cancer of 11.6% to 24.4%, depending on age, for more than 100,000 unselected patients with positive mammograms.[46] From the descriptions of the patients enrolled in the majority of the studies, one might expect these patients to be similar to those described by Banks et al. However, a potential limitation of the available studies is the high prevalence of breast cancer in the patients enrolled in them. Prevalences of 50% or higher were reported for most patient groups studied (except for studies of ultrasound, with a more reasonable prevalence of 24%). These high prevalences suggest that the studied women were not representative of the general population of women with suspicious findings after breast cancer screening exams. Therefore the negative predictive values calculated from these studies may not

be directly applicable to the clinic (because negative predictive values are dependent on disease prevalence). Sensitivity, specificity, and negative likelihood ratios are independent of prevalence; however, they are dependent on the spectrum of disease present in the patient population. If the patients enrolled in the included studies had disease that was more advanced than that in the typical clinical population (a possibility suggested by the high prevalence rates), the resulting estimates of test performance may not be directly applicable to the clinical situation. As such, our strength of evidence ratings apply only to the internal validity of these studies, as applied to the populations that were enrolled. Because the evidence bases for PET, scintimammography, and MRI appear to be affected by spectrum bias, care should be taken when attempting to apply these results to any other patient populations.

A clinical use of non-invasive imaging technology that we were unable to address in this report, due to an absence of evidence, is examination of women with probably benign lesions. Women who have probably benign findings on mammography screening, or unusable findings (BIRADS 0 or 3) are commonly referred for more frequent mammography screening tests. Mammography does expose patients to x-rays, and is often reported to be uncomfortable; and of course, women must suffer from considerable mental distress from the months of not knowing for sure if they are healthy. It is possible that non-invasive imaging could be used to accurately detect suspicious cases in this population. These women would be referred for further evaluation, and the other patients could return to routine screening. Because these women have a completely different spectrum of disease than the women enrolled in the published studies, we cannot extrapolate our findings to this population group. Further research is necessary on this possible use of non-invasive imaging technology.

Further research on these diagnostic imaging procedures is desirable. The most informative study design would randomize women with suspicious findings to either be directly followed up by biopsy, or by a non-invasive imaging method first, and only women with positive findings on imaging would be evaluated by biopsy. Ideally, all women in the study would be followed for many years, and patient-oriented outcomes such as mortality and quality of life recorded. Such a study would directly address the question of whether women benefit from being evaluated by non-invasive imaging methods. However, a study of this type would be expensive, time-consuming, and may present ethical issues about delayed cancer diagnoses. A more feasible approach to further research would be to conduct diagnostic cohort studies on patient populations more representative of the general population of women with suspicious findings after breast cancer screening. A cost-benefit analysis of data from such studies would indirectly address the question of whether women benefit from being evaluated by non-invasive imaging methods.

In conclusion, an ideal test to evaluate breast abnormalities found by mammography or breast examination would distinguish women who needed to have a biopsy from those who could safely avoid one. A woman who has a negative test result should be very confident that she does not have breast cancer before deciding to forgo a biopsy. The risk of a test missing a cancer is dependent on a woman's risk of cancer prior to undergoing a test as well as the accuracy of the test. While an "acceptable" level of risk of cancer given a negative test result is dependent on a woman's personal preferences; at least one organization, the Ontario Ministry of Health, has suggested that a 98% negative predictive value threshold would be societally acceptable to reliably preclude breast biopsy. Evidence suggests that for women at average risk of breast cancer receiving a biopsy in the US, all four of the diagnostic tests evaluated in this report fall short of this 98% threshold. While MRI was more sensitive than the other technologies in typical

usage, even this technology would result in a 96% negative predictive value for a woman at average risk; women at higher risk would have an even lower negative predictive value.

References

1. American Cancer Society (ACS). Cancer facts and figures 2003. Atlanta (GA): American Cancer Society (ACS); 2003. 48 p. Also available: http://www.cancer.org/docroot/STT/stt_0.
asp.

2. Jemal A, Murray T, Ward E, et al. Cancer statistics, 2005. CA Cancer J Clin 2005;55(1):10-30.

3. Greenlee RT, Hill-Harmon MB, Murray T, et al. Cancer statistics, 2001. CA Cancer J Clin 2001;51(1):15-36.

4. Guinee VF. Epidemiology of breast cancer. In: Bland KI, Copeland EM 3rd, editors. The breast: comprehensive management of benign and malignant diseases. 2nd ed. Vol. 1. Philadelphia (PA): W.B. Saunders Company; 1998. p. 339-51.

5. Jemal A, Tiwari RC, Murray T, et al. Cancer statistics, 2004. CA Cancer J Clin 2004;54(1):8-29.

6. Yeatman TJ, Bland KI. Assessment and designation of breast cancer stage. In: Bland KI, Copeland EM 3rd, editors. The breast: comprehensive management of benign and malignant diseases. 2nd ed. Vol. 1. Philadelphia (PA): W.B. Saunders Company; 1998. p. 400-17.

7. American Cancer Society (ACS). After the tests: staging. [internet]. Atlanta (GA): American Cancer Society (ACS); 2001 [cited 2002 Jan 2]. [2 p]. Available: http://www.cancer.org.

8. U.S. Preventive Services Task Force (USPSTF). Screening for breast cancer: recommendations and rationale. Washington (DC): Agency for Healthcare Research and Quality (AHRQ); 2002 Feb. 10 p. (AHRQ Pub.; no. 02-507A). Also available: http://www.ahrq.gov/clinic/3rduspstf/
breastcancer.

9. Humphrey LL, Helfand M, Chan BK, et al. Breast cancer screening: a summary of the evidence for the U.S. Preventive Services Task Force. Ann Intern Med 2002 Sep 3;137(5 Part 1):347-67.

10. Rosato FE, Rosato EL. Examination techniques: roles of the physician and patient in evaluating breast diseases. In: Bland KI, Copeland EM 3rd, editors. The breast: comprehensive management of benign and malignant diseases. 2nd ed. Vol. 1. Philadelphia (PA): W.B. Saunders Company; 1998. p. 615-23.

11. Liberman L, Menell JH. Breast imaging reporting and data system (BI-RADS). Radiol Clin North Am 2002 May;40(3):409-30.

12. Diagnostic and Interventional Radiology Department. Mammography. Omaha (NE): Creighton University School of Medicine; 22 p. Also available: http://radiology.creighton.edu/mammo.html.

13. Obenauer S, Hermann KP, Grabbe E . Applications and literature review of the BI-RADS classification. Eur Radiol 2005 May;15(5):1027-36 .

14. Schillaci O, Buscombe JR. Breast scintigraphy today: indications and limitations. Eur J Nucl Med Mol Imaging 2004 Jun;31(Suppl 1):S35-45.

15. Kopans DB. The positive predictive value of mammography. AJR Am J Roentgenol 1992 Mar;158(3):521-6.

16. Khan KS. Systematic reviews of diagnostic tests: a guide to methods and application. Baillieres Best Pract Res Clin Obstet Gynaecol 2005 Feb; 19(1):37-46.

17. Dujardin B, Van den Ende J, Van Gompel A, et al. Likelihood ratios: a real improvement for clinical decision making. Eur J Epidemiol 1994 Feb;10(1):29-36.

18. Gallagher EJ. Clinical utility of likelihood ratios. Ann Emerg Med 1998 Mar;31(3):391-7.

19. Eisenberg JM. Derived thresholds. Determining the diagnostic probabilities at which clinicians initiate testing and treatment. Med Decis Making 1983;3(2):155-68.

20. Barone JE. Evidence-based medicine applied to sentinel lymph node biopsy in patients with breast cancer. Am Surg 2005 Jan;71(1):66-70.

21. Verkooijen HM. Diagnostic accuracy of large-core needle biopsy for nonpalpable breast disease: a meta-analysis. Br J Cancer 2000 Mar;82(5): 1017-21.

22. Kerlikowske K. Evaluation of abnormal mammography results and palpable breast abnormalities. Ann Intern Med 2003 Aug 19;139(4):274-84.

23. Yasmeen S. Frequency and predictive value of a mammographic recommendation for short-interval follow-up. J Natl Cancer Inst 2003 Mar 19;95(6):429-36.

24. Chalmers I, Adams M, Dickersin K, et al. A cohort study of summary reports of controlled trials. JAMA 1990 Mar 9;263(10):1401-5.

25. Neinstein LS. A review of Society for Adolescent Medicine abstracts and Journal of Adolescent Health Care articles. J Adolesc Health Care 1987 Mar; 8(2):198-203.

26. De Bellefeuille C, Morrison CA, Tannock IF. The fate of abstracts submitted to a cancer meeting: factors which influence presentation and subsequent publication. Ann Oncol 1992 Mar;3(3):187-91.

27. Scherer RW, Langenberg P. Full publication of results initially presented in abstracts. In: Cochrane Library [Cochrane methodology review]. Issue 2. Oxford: Update Software; 2001 [cited 2001 Apr 23]. [35 p]. Available: http://www.cochranelibrary.com.

28. Marx WF, Cloft HJ, Do HM, et al. The fate of neuroradiologic abstracts presented at national meetings in 1993: rate of subsequent publication in peer-reviewed, indexed journals. AJNR Am J Neuroradiol 1999 ;20(6):1173-7.

29. Yentis SM, Campbell FA, Lerman J . Publication of abstracts presented at anaesthesia meetings. Can J Anaesth 1993 Jul;40(7):632-4.

30. Carneiro AV. Applicability of clinical trial results to the individual patient: practice guidelines. Rev Port Cardiol 2003 Feb;22(2):259-68.

31. Armitage P, Berry G. Statistical methods in medical research. 3rd ed. Oxford, England: Blackwell Scientific; 1994. 620 p.

32. Lopis R. A personal encounter with a mystery illness. Aust Fam Physician 1991 Mar;20(3):316-7.

33. Bojia F. Comparison of fine-needle aspiration cytology and excisional biopsy of breast lesions. East Afr Med J 2001 May;78(5):226-8.

34. Vetrani A. Fine-needle aspiration biopsies of breast masses. An additional experience with 1153 cases (1985 to 1988) and a meta-analysis. Cancer 1992 Feb 1;69(3):736-40.

35. Abu-Salem OT. Fine needle aspiration biopsy (FNAB) of breast lumps: comparison study between pre- and post-operative histological diagnosis. Arch Inst Pasteur Tunis 2002;79(1-4):59-63.

36. Ljung BM. Diagnostic accuracy of fine-needle aspiration biopsy is determined by physician training in sampling technique. Cancer 2001 Aug 25;93(4):263-8.

37. West S, King V, Carey TS, et al. Systems to rate the strength of scientific evidence. Contract no. 290-97-0011. Rockville (MD): Agency for Healthcare Research and Quality (AHRQ); 2002 Apr. 199 p. (Evidence report/technology assessment; no. 47).

38. Atkins D, Best D, Briss PA, et al. Grading quality of evidence and strength of recommendations. BMJ 2004 Jun 19;328(7454):1490.

39. Higgins JP, Thompson SG. Quantifying heterogeneity in a meta-analysis. Stat Med 2002 Jun 15;21(11):1539-58.

40. Higgins JP, Thompson SG, Deeks JJ, et al. Measuring inconsistency in meta-analysis. BMJ 2003 Sep 6;327:557-60.

41. Ioannidis JP, Contopoulos-Ioannidis DG, Lau J. Recursive cumulative meta-analysis: a diagnostic for the evolution of total randomized evidence from group and individual patient data. J Clin Epidemiol 1999 Apr;52(4):281-91.

42. Lau J, Schmid CH, Chalmers TC. Cumulative meta-analysis of clinical trials builds evidence for exemplary medical care. J Clin Epidemiol 1995 Jan;48(1):45-57; 59-60.

43. Moses LE, Shapiro D, Littenberg B. Combining independent studies of a diagnostic test into a summary ROC curve: data-analytic approaches and some additional considerations. Stat Med 1993 Jul 30;12(14):1293-316.

44. Deeks JJ. Systematic reviews of evaluations of diagnostic and screening tests. In: Egger M, Smith GD, Altman DG, editors. Systematic reviews in health care: meta-analysis in context. 2nd ed. London, England: BMJ Books; 2001. p. 248-82.

45. Stengel D, Bauwens K, Sehouli J, et al. A likelihood ratio approach to meta-analysis of diagnostic studies. J Med Screen 2003;10(1):47-51.

46. Banks E. Influence of personal characteristics of individual women on sensitivity and specificity of mammography in the Million Women Study: cohort study. BMJ 2004 Aug 28;329(7464):477.

47. Higgins JP, Thompson SG. Controlling the risk of spurious findings from meta-regression. Stat Med 2004 Jun 15;23(11):1663-82.

48. Harbord R, Statistical Software Components, Boston College Department of Economics. METAREG: Stata module to perform meta-analysis regression. [software download]. Stockholm, Sweden: EconPapers; 2004 Sep 22 [cited 2004 Nov 3]. Available: http://econpapers.hhs.se/software/bocbocode/s446201.htm.

49. Elan Avinza approved. Pharm Approval Monthly 2002 Apr 1;7(4):56.

50. Agency for Healthcare Research and Quality (AHRQ). FDG positron emission tomography for evaluating breast cancer - systematic review. Rockville (MD): Agency for Healthcare Research and Quality (AHRQ); 2001. Also available: http://www.mrw.interscience.wiley.com/cochrane/clhta/articles/HTA-20020419/frame.html.

51. Holle LH, Trampert L, Lung-Kurt S, et al. Investigations of breast tumors with fluorine-18-fluorodeoxyglucose and SPECT. J Nucl Med 1996 Apr;37(4):615-22.

52. Brix G, Henze M, Knopp MV, et al . Comparison of pharmacokinetic MRI and [18F] fluorodeoxyglucose PET in the diagnosis of breast cancer: initial experience. Eur Radiol 2001;11(10):2058-70.

53. Crowe JP Jr, Adler LP, Shenk RR, et al. Positron emission tomography and breast masses: comparison with clinical, mammographic, and pathological findings. Ann Surg Oncol 1994 Mar;1(2):132-40.

54. Yutani K, Tatsumi M, Shiba E, et al. Comparison of dual-head coincidence gamma camera FDG imaging with FDG PET in detection of breast cancer and axillary lymph node metastasis. J Nucl Med 1999 Jun;40(6):1003-8.

55. Yutani K, Shiba E, Kusuoka H, et al. Comparison of FDG-PET with MIBI-SPECT in the detection of breast cancer and axillary lymph node metastasis. J Comput Assist Tomogr 2000;24(2):274-80.

56. Heinisch M, Gallowitsch HJ, Mikosch P, et al. Comparison of FDG-PET and dynamic contrast-enhanced MRI in the evaluation of suggestive breast lesions. Breast 2003 Feb;12(1):17-22.

57. Walter C, Scheidhauer K, Scharl A, et al. Clinical and diagnostic value of preoperative MR mammography and FDG-PET in suspicious breast lesions. Eur Radiol 2003 Jul;13(7):1651-6.

58. Palmedo H, Bender H, Grunwald F, et al. Comparison of fluorine-18 fluorodeoxyglucose positron emission tomography and technetium-99m methoxyisobutylisonitrile scintimammography in the detection of breast tumours. Eur J Nucl Med 1997 Sep;24(9):1138-45.

59. Tse NY, Hoh CK, Hawkins RA, et al. The application of positron emission tomographic imaging with fluorodeoxyglucose to the evaluation of breast disease. Ann Surg 1992 Jul;216(1):27-34.

60. Bergmann H. An inter-laboratory comparison study of image quality of PET scanners using the NEMA NU 2-2001 procedure for assessment of image quality. Phys Med Biol 2005 May 21;50(10):2193-207.

61. Levine EA, Freimanis RI, Perrier ND, et al. Positron emission mammography: initial clinical results. Ann Surg Oncol 2003;10(1):86-91.

62. McDonough MD, DePeri ER, Mincey BA. The role of positron emission tomographic imaging in breast cancer. Curr Oncol Rep 2004 Jan;6(1):62-8.

63. Avril N, Schelling M, Dose J, et al. Utility of PET in Breast Cancer. Clin Positron Imaging 1999 Oct;2(5):261-71.

64. Hoh CK, Schiepers C. 18-FDG imaging in breast cancer. Semin Nucl Med 1999 Jan;29(1):49-56.

65. Bleckmann C, Dose J, Bohuslavizki KH, et al. Effect of attenuation correction on lesion detectability in FDG PET of breast cancer. J Nucl Med 1999 Dec;40(12):2021-4.

66. Doshi NK, Shao Y, Silverman RW, et al. Design and evaluation of an LSO PET detector for breast cancer imaging. Med Phys 2000 Jul;27(7):1535-43.

67. Yutani K, Tatsumi M, Uehara T, et al. Effect of patients' being prone during FDG PET for the diagnosis of breast cancer. AJR Am J Roentgenol 1999 Nov;173(5):1337-9.

68. Rosen EL, Turkington TG, Soo MS, et al. Detection of primary breast carcinoma with a dedicated, large-field-of-view FDG PET mammography device: initial experience. Radiology 2005 Feb;234(2):527-34.

69. Bhargava P, Zhuang H, Kumar R, et al. Iatrogenic artifacts on whole-body F-18 FDG PET imaging. Clin Nucl Med 2004 Jul;29(7):429-39.

70. Murthy K, Aznar M, Bergman AM, et al. Positron emission mammographic instrument: initial results. Radiology 2000 Apr;215(1):280-5.

71. Boerner AR, Weckesser M, Herzog H, et al. Optimal scan time for fluorine-18 fluorodeoxyglucose positron emission tomography in breast cancer. Eur J Nucl Med 1999 Mar;26(3):226-30.

72. Wahl RL. Current status of PET in breast cancer imaging, staging, and therapy. Semin Roentgenol 2001 Jul;36(3):250-60.

73. Marshall C, Mustafa S, Wheatley DC, et al. A comparison of 18F-FDG gamma camera PET, mammography and ultrasonography in demonstrating primary disease in locally advanced breast cancer. Nucl Med Commun 2004 Jul;25(7):721-5.

74. Cronin B, Marsden PK, O'Doherty MJ. Are restrictions to behaviour of patients required following fluorine-18 fluorodeoxyglucose positron emission tomographic studies. Eur J Nucl Med 1999 Feb;26(2):121-8.

75. Silberstein EB. Prevalence of adverse reactions to positron emitting radiopharmaceuticals in nuclear medicine. Pharmacopeia Committee of the Society of Nuclear Medicine. J Nucl Med 1998 Dec;39(12):2190-2.

76. Katanick SL. Fundamentals of ICANL accreditation. J Nucl Med Technol 2005 Mar;33(1):19-23.

77. Hillner BE. Decision analysis: MIBI imaging of nonpalpable breast abnormalities. J Nucl Med 1997 Nov;38(11):1772-8.

78. Allen MW, Hendi P, Schwimmer J, et al. Decision analysis for the cost effectiveness of sestamibi scintimammography in minimizing unnecessary biopsies. Q J Nucl Med 2000 Jun;44(2):168-85.

79. Liberman M, Sampalis F, Mulder DS, et al. Breast cancer diagnosis by scintimammography: a meta-analysis and review of the literature (Provisional record). Breast Cancer Res Treat 2003;80(1):115-26.

80. Medical Advisory Secretariat, Ontario Ministry of Health and Long-Term Care. Scintimammography. Toronto: Ontario Ministry of Health and Long-Term Care; 2003. Also available: http://www.mrw.interscience.wiley.com/cochrane/clhta/articles/HTA-20040735/frame.html.

81. Littenberg B, Moses LE. Estimating diagnostic accuracy from multiple conflicting reports: a new meta-analytic method. Med Decis Making 1993;13(4):313-21.

82. Sampalis FS, Denis R, Picard D, et al. International prospective evaluation of scintimammography with (99m)technetium sestamibi. Am J Surg 2003 Jun;185(6):544-9.

83. Khalkhali I, Baum JK, Villanueva-Meyer J, et al. (99m)Tc sestamibi breast imaging for the examination of patients with dense and fatty breasts: multicenter study. Radiology 2002 Jan;222(1):149-55.

84. Scopinaro F, Schillaci O, Ussof W, et al. A three center study on the diagnostic accuracy of 99mTc-MIBI scintimammography. Anticancer Res 1997;17(3B):1631-4.

85. Tofani A, Sciuto R, Semprebene A, et al. 99Tcm-MIBI scintimammography in 300 consecutive patients: factors that may affect accuracy. Nucl Med Commun 1999 Dec;20(12):1113-21.

86. Mekhmandarov S, Sandbank J, Cohen M, et al. Technetium-99m-MIBI scintimammography in palpable and nonpalpable breast lesions. J Nucl Med 1998 Jan;39(1):86-91.

87. Danielsson R, Bone B, Gad A, et al. Sensitivity and specificity of planar scintimammography with 99mTc-sestamibi. Acta Radiol 1999 Jul;40(4):394-9.

88. Wilczek B, Aspelin P, Bone B, et al. Complementary use of scintimammography with 99m-Tc-MIBI to triple diagnostic procedure in palpable and non-palpable breast lesions. Acta Radiol 2003 May;44(3):288-93.

89. Krishnaiah G, Sher-Ahmed A, Ugwu-Dike M, et al. Technetium-99m sestamibi scintimammography complements mammography in the detection of breast cancer. Breast J 2003;9(4):288-94.

90. Obwegeser R, Berghammer P, Rodrigues M, et al. A head-to-head comparison between technetium-99m-tetrofosmin and technetium-99m-MIBI scintigraphy to evaluate suspicious breast lesions. Eur J Nucl Med 1999 Dec;26(12):1553-9.

91. Koukouraki S, Koukourakis MI, Vagios E, et al. The role of 99m Tc-sestamibi scintimammography and colour Doppler ultrasonography in the evaluation of breast lesions. Nucl Med Commun 2001 Nov;22(11):1243-8.

92. Schillaci O, Scopinaro F, Danieli R, et al. 99Tcm-sestamibi scintimammography in patients with suspicious breast lesions: comparison of SPET and planar images in the detection of primary tumours and axillary lymph node involvement. Nucl Med Commun 1997 Sep;18(9):839-45.

93. Villanueva-Meyer J, Leonard MH Jr, Briscoe E, et al. Mammoscintigraphy with technetium-99m-sestamibi in suspected breast cancer. J Nucl Med 1996 Jun;37(6):926-30.

94. Chen SL, Yin YQ, Chen JX, et al. The usefulness of technetium-99m-MIBI scintimammography in diagnosis of breast cancer: using surgical histopathologic diagnosis as the gold standard. Anticancer Res 1997;17(3B):1695-8.

95. Tiling R, Sommer H, Pechmann M, et al. Comparison of technetium-99m-sestamibi scintimammography with contrast- enhanced MRI for diagnosis of breast lesions. J Nucl Med 1997 Jan;38(1):58-62.

96. Palmedo H, Grunwald F, Bender H, et al. Scintimammography with technetium-99m methoxyisobutylisonitrile: comparison with mammography and magnetic resonance imaging. Eur J Nucl Med 1996 Aug;23(8):940-6.

97. Leidenius MH, Leppanen EA, Tykka HT, et al. The role of Tc99m-sestamibi scintimammography in combination with the triple assessment of primary breast cancer. Eur J Surg Oncol 2002 Mar;28(2):108-12.

98. Imbriaco M, Del Vecchio S, Riccardi A, et al. Scintimammography with 99mTc-MIBI versus dynamic MRI for non-invasive characterization of breast masses. Eur J Nucl Med 2001 Jan;28(1):56-63.

99. Papantoniou V, Christodoulidou J, Papadaki E, et al. 99mTc-(V)DMSA scintimammography in the assessment of

breast lesions: comparative study with 99mTc-MIBI. Eur J Nucl Med 2001 Jul;28(7):923-8.

100. Sanidas EE, Koukouraki S, Velidaki A, et al. Contribution of 99mTc-anti-carcinoembryonic antigen antibody and 99mTc-sestamibi scintimammography in the evaluation of high risk palpable breast lesions. Nucl Med Commun 2003 Mar; 24(3):291-6.

101. Gutfilen B, Fonseca LM. Comparison of Tc-99m THY and Tc-99m MIBI scans for diagnosis of breast lesions. J Exp Clin Cancer Res 2001 Sep;20(3):385-91.

102. Yuen-Green M, Wasnich R, Caindec-Ranchez S, et al. New method for breast cancer detection using TC-99m sestamibi scintimammography. Hawaii Med J 1996 Feb;55(2):26-8.

103. Alonso O, Massardo T, Delgado LB, et al. Is (99m)Tc-sestamibi scintimammography complementary to conventional mammography for detecting breast cancer in patients with palpable masses. J Nucl Med 2001 Nov;42(11):1614-21.

104. Ambrus E, Rajtar M, Ormandi K, et al. Value of 99m-Tc MIBI and 99m-Tc(V) DMSA scintigraphy in evaluation of breast mass lesions. Anticancer Res 1997;17(3B):1599-605 .

105. Burak Z, Argon M, Memis A, et al. Evaluation of palpable breast masses with 99Tcm-MIBI: a comparative study with mammography and ultrasonography. Nucl Med Commun 1994 Aug;15(8):604-12.

106. Palmedo H, Schomburg A, Grunwald F, et al. Scintimammography with Tc-99m MIBI in patients with suspicion of primary breast cancer. Nucl Med Biol 1996 Aug;23(6):681-4.

107. Palmedo H, Schomburg A, Grunwald F, et al. Technetium-99m-MIBI scintimammography for suspicious breast lesions. J Nucl Med 1996 Apr;37(4):626-30.

108. Chen S, Liu W, Mao Y, et al. 99 mTc-MIBI and 99 mTc-MDP scintimammography for detecting breast carcinoma. Chin Med J 2000;113(5):400-3.

109. Carril JM, Gomez-Barquin R, Quirce R, et al. Contribution of 99mTc-MIBI scintimammography to the diagnosis of non- palpable breast lesions in relation to mammographic probability of malignancy. Anticancer Res 1997;17(3B):1677-81.

110. Aguilar J, Andres B, Nicolas F, et al. Value of 99mTc-MIBI scintimammography in women with impalpable breast lesions seen on mammography. Eur J Surg 2001 May;167(5):344-6.

111. Bekis R. 99mTc sestamibi scintimammography. Screening mammographic non-palpable suspicious breast lesions: preliminary results. Nucl Med (Stuttg) 2004 Feb;43(1):16-20.

112. Maffioli L, Agresti R, Chiti A, et al. Prone scintimammography in patients with non-palpable breast lesions. Anticancer Res 1996;16(3A):1269-73.

113. Fondrinier E, Muratet JP, Anglade E, et al. Clinical experience with 99mTc-MIBI scintimammography in patients with breast microcalcifications. Breast 2004 Aug;13(4):316-20.

114. Spanu A, Schillaci O, Meloni GB, et al. The usefulness of 99mTc-tetrofosmin SPECT scintimammography in the detection of small size primary breast carcinomas. Int J Oncol 2002 Oct;21(4):831-40.

115. Khalkhali I, Vargas HI. The role of nuclear medicine in breast cancer detection: functional breast imaging. Radiol Clin North Am 2001 Sep;39(5):1053-68.

116. Khalkhali I, Caravaglia G, Abdel-Nabi HH, et al. Society of Nuclear Medicine procedure guideline for breast scintigraphy. version 2.0. Reston (VA): Society of Nuclear Medicine; 2004 Jun 2. 4 p.
Also available: http://interactive.snm.org/
docs/Breast_v2.0.pdf.

117. Coover LR, Caravaglia G, Kuhn P . Scintimammography with dedicated breast camera detects and localizes occult carcinoma. J Nucl Med 2004 Apr;45(4):553-8.

118. Yildiz A, Colak T, Gungor F, et al. Diagnostic value of 99mTc MIBI scintimammography in patients with breast lesions. Rev Esp Med Nucl 2001 Jun; 20(4):276-81.

119. Bombardieri E, Crippa F, Baio SM, et al. Nuclear medicine advances in breast cancer imaging. Tumori 2001;87(5):277-87.

120. Clarke EA, Notghi A, Harding LK . Can reduced imaging times be used for scintimammography. Nucl Med Commun 1999 Oct;20(10):883-6.

121. Edell SL, Eisen MD. Current imaging modalities for the diagnosis of breast cancer. Del Med J 1999 Sep;71(9):377-82.

122. Zegel H, Heller L, Edell S, et al. Tc-99m sestamibi scintimammography in the mammographically dense breast. Clin Nucl Med 1999 Dec;24(12):968-74.

123. Monticciolo DL. How useful are new radiographic tools for detecting breast cancer. Postgrad Med 2004 Mar;115(3):76-7.

124. Thomson LE, Allman KC. Erythema multiforme reaction to sestamibi. J Nucl Med 2001 Mar;42(3):534.

125. Hesslewood SR, Keeling DH. Frequency of adverse reactions to radiopharmaceuticals in Europe. Eur J Nucl Med 1997 Sep;24(9):1179-82 .

126. Silberstein EB, Ryan J. Prevalence of adverse reactions in nuclear medicine. Pharmacopeia Committee of the Society of Nuclear Medicine. J Nucl Med 1996 Jan;37(1):185-92.

127. Klaus AJ, Klingensmith WC 3rd, Parker SH, et al. Comparative value of 99mTc-sestamibi scintimammography and sonography in the diagnostic workup of breast masses. AJR Am J Roentgenol 2000 Jun;174(6):1779-83.

128. Blue Cross Blue Shield Association (BCBS). Breast MRI for detection or diagnosis of primary or recurrent breast cancer. Chicago (IL): Blue Cross Blue Shield Association; 2004 Apr. 88 p. (TEC Assessment Program; vol. 19, no. 1). Also available: http://www.mrw.interscience.wiley.com/cochrane/clhta/articles/HTA-20040234/frame.html.

129. Hrung JM, Sonnad SS, Schwartz J S, et al. Accuracy of MR imaging in the work-up of suspicious breast lesions: a diagnostic meta-analysis (Structured abstract). Acad Radiol 1999;6(7):387-97.

130. Bluemke DA, Gatsonis CA, Chen MH, et al. Magnetic resonance imaging of the breast prior to biopsy. JAMA 2004 Dec 8;292 (22):2735-42.

131. Heiberg EV, Perman WH, Herrmann VM, et al. Dynamic sequential 3D gadolinium-enhanced MRI of the whole breast. Magn Reson Imaging 1996;14(4):337-48.

132. Obdeijn IM, Kuijpers TJ, van Dijk P, et al. MR lesion detection in a breast cancer population. J Magn Reson Imaging 1996;6(6):849-54.

133. Hachiya J, Seki T, Okada M, et al. MR imaging of the breast with Gd-DTPA enhancement: comparison with mammography and ultrasonography. Radiat Med 1991;9(6):232-40.

134. Cecil KM, Schnall MD, Siegelman ES, et al. The evaluation of human breast lesions with magnetic resonance imaging and proton magnetic resonance spectroscopy. Breast Cancer Res Treat 2001 Jul;68(1):45-54.

135. Del Maschio A, Bazzocchi M, Giuseppetti GM, et al. Breast MRI: report on a multicentric national trial by the Study Section of Magnetic Resonance and Breast Imaging. Radiol Med (Torino) 2002 Oct;104(4):262-72 .

136. Nakahara H, Namba K, Fukami A, et al. Three-dimensional MR imaging of mammographically detected suspicious microcalcifications. Breast Cancer 2001;8(2):116-24.

137. Malich A, Boehm T, Facius M, et al. Differentiation of mammographically suspicious lesions: evaluation of breast ultrasound, MRI mammography and electrical impedance scanning as adjunctive technologies in breast cancer detection. Clin Radiol 2001 Apr;56(4):278-83.

138. Knopp MV, Bourne MW, Sardanelli F, et al. Gadobenate dimeglumine-enhanced MRI of the breast: analysis of dose response and comparison with gadopentetate dimeglumine. AJR Am J Roentgenol 2003 Sep;181(3):663-76.

139. Huang W, Fisher PR, Dulaimy K, et al. Detection of breast malignancy: diagnostic MR protocol for improved specificity. Radiology 2004 Aug;232(2):585-91.

140. Magnetic resonance imaging systems. Health Devices 2005 Apr 1;34(4):117-38.

141. Paakko E, Reinikainen H, Lindholm EL, et al. Low-field versus high-field MRI in diagnosing breast disorders. Eur Radiol 2005 Jul;15(7):1361-8.

142. Yen YF, Han KF, Daniel BL, et al. Dynamic breast MRI with spiral trajectories: 3D versus 2D. J Magn Reson Imaging 2000;11(4):351-9.

143. Sardanelli F, Rescinito G, Giordano GD, et al. MR dynamic enhancement of breast lesions: high temporal resolution during the first-minute versus eight-minute study. J Comput Assist Tomogr 2000;24(5):724-31.

144. Sardanelli F, Lozzelli A, Fausto A. MR imaging of the breast: indications, established technique, and new directions. Eur Radiol 2003 Nov;13 Suppl 3(N28-36.

145. Orel SG, Schnall MD. MR imaging of the breast for the detection, diagnosis, and staging of breast cancer. Radiology 2001 Jul;220(1):13-30.

146. Lee CH. Problem solving MR imaging of the breast. Radiol Clin North Am 2004 Sep;42(5):919-34.

147. Szabo BK, Aspelin P, Wiberg MK, et al. Dynamic MR imaging of the breast. Analysis of kinetic and morphologic diagnostic criteria. Acta Radiol 2003 Jul;44(4):379-86.

148. Kuhl CK, Klaschik S, Mielcarek P, et al. Do T2-weighted pulse sequences help with the differential diagnosis of enhancing lesions in dynamic breast MRI? J Magn Reson Imaging 1999 Feb;9(2):187-96.

149. Schnall MD, Rosten S, Englander S, et al. A combined architectural and kinetic interpretation model for breast MR images. Acad Radiol 2001 Jul;8(7):591-7.

150. Agoston AT, Daniel BL, Herfkens RJ, et al. Intensity-modulated parametric mapping for simultaneous display of rapid dynamic and high-spatial-resolution breast MR imaging data. Radiographics 2001;21(1):217-26.

151. American College of Radiology (ACR). ACR Breast Imaging and Reporting Data System, Breast Imaging Atlas. Reston (VA): American College of Radiology (ACR); 2003. ACR BI-RADS - Magnetic Resonance Imaging.

152. Helbich TH. Contrast-enhanced magnetic resonance imaging of the breast. Eur J Radiol 2000 Jun;34(3):208-19.

153. Artemov D, Revelon G, Atalar E, et al. Switchable multicoil array for MR micro-imaging of breast lesions. Magn Reson Med 1999 Mar;41(3):569-74.

154. Shen GX, Wu J, Boada FE. Multiple channel phased arrays for echo planar imaging. MAGMA 2000 Dec;11(3):138-43.

155. Friedman PD, Swaminathan SV, Smith R. SENSE imaging of the breast. AJR Am J Roentgenol 2005 Feb;184(2):448-51.

156. Morris EA. Breast cancer imaging with MRI. Radiol Clin North Am 2002 May;40(3):443-66.

157. Orel SG. MR imaging of the breast. Radiol Clin North Am 2000 Jul;38(4):899-913.

158. Sardanelli F, Iozzelli A, Fausto A. Contrast agents and temporal resolution in breast MR imaging. J Exp Clin Cancer Res 2002 Sep;21(3 Suppl):69-75.

159. National Cancer Data Base (NCDB) [database online]. Chicago (IL): American College of Surgeons (ACS); 1996-1998 [cited 1999 Mar 5]. Relative survival of breast cancer patients by AJCC stage of disease at diagnosis,

1985-1990 cases [table]. [3 p].
Available: http://www.facs.org/about_college/acsdept/cancer_dept/programs/ncdb/breastcancer4.html.

160. Coulthard A, Potterton AJ. Pitfalls of breast MRI. Br J Radiol 2000 Jun;73(870):665-71.

161. Kuhl CK, Schild HH. Dynamic image interpretation of MRI of the breast. J Magn Reson Imaging 2000 Dec;12(6):965-74.

162. Chen W, Giger ML, Lan L, et al. Computerized interpretation of breast MRI: investigation of enhancement-variance dynamics. Med Phys 2004 May;31(5):1076-82.

163. Partridge SC, Heumann EJ, Hylton NM. Semi-automated analysis for MRI of breast tumors. Stud Health Technol Inform 1999;62:259-60.

164. Szabo BK, Wiberg MK, Bone B, et al. Application of artificial neural networks to the analysis of dynamic MR imaging features of the breast. Eur Radiol 2004 Jul;14(7):1217-25.

165. Kelcz F, Furman-Haran E, Grobgeld D, et al. Clinical testing of high-spatial-resolution parametric contrast-enhanced MR imaging of the breast. AJR Am J Roentgenol 2002 Dec;179(6):1485-92.

166. Liney GP, Gibbs P, Hayes C, et al. Dynamic contrast-enhanced MRI in the differentiation of breast tumors: user-defined versus semi-automated region-of-interest analysis. J Magn Reson Imaging 1999 Dec;10 (6):945-9.

167. Heywang-Kobrunner SH, Bick U, Bradley WG Jr, et al. International investigation of breast MRI: results of a multicentre study (11 sites) concerning diagnostic parameters for contrast-enhanced MRI based on 519 histopathologically correlated lesions. Eur Radiol 2001;11(4):531-46.

168. Heywang-Kobrunner SH, Heinig A, Pickuth D, et al. Interventional MRI of the breast: lesion localisation and biopsy. Eur Radiol 2000;10(1):36-45.

169. Schedel H. Magnetic Resonance Female Breast Imaging (MRFBI) - Evaluation of the Changes in Signal Intensity over Time Pre- and Post-administration of 0.2 mmol/kg Gd-DTPA. Zentralbl Gynakol 2002 Feb;124(2):104-10 .

170. Morris EA. Review of breast MRI: indications and limitations. Semin Roentgenol 2001 Jul;36(3):226-37.

171. Rieber A, Nussle K, Merkle E, et al. MR mammography: influence of menstrual cycle on the dynamic contrast enhancement of fibrocystic disease. Eur Radiol 1999;9(6):1107-12.

172. Runge VM. Safety of approved MR contrast media for intravenous injection. J Magn Reson Imaging 2000 Aug 1;12(2):205-13.

173. Medical magnetic resonance (MR) procedures: protection of patients. Health Phys 2004 Aug;87(2):197-216.

174. Shellock FG. Radiofrequency energy-induced heating during MR procedures: a review. J Magn Reson Imaging 2000 Jul;12(1):30-6.

175. Fiek M, Remp T, Reithmann C, et al. Complete loss of ICD programmability after magnetic resonance imaging. Pacing Clin Electrophysiol 2004 Jul;27(7):1002-4.

176. Shellock FG. Magnetic resonance safety update 2002: implants and devices. J Magn Reson Imaging 2002 Nov;16(5):485-96.

177. Foster JR, Hall DA, Summerfield AQ, et al. Sound-level measurements and calculations of safe noise dosage during EPI at 3 T. J Magn Reson Imaging 2000 Jul;12(1):157-63.

178. Dempsey MF, Condon B, Hadley DM . Investigation of the factors responsible for burns during MRI. J Magn Reson Imaging 2001 Apr;13(4):627-31.

179. Shellock FG, Crues JV. MR procedures: biologic effects, safety, and patient care. Radiology 2004 Sep;232(3):635-52.

180. Shellock FG. Biomedical implants and devices: Assessment of magnetic field interactions with a 3.0-Tesla MR

system. J Magn Reson Imaging 2002 Dec; 16(6):721-32.

181. Feychting M. Health effects of static magnetic fields--a review of the epidemiological evidence. Prog Biophys Mol Biol 2005;87(2-3):241-6.

182. Hong CZ, Shellock FG. Short-term exposure to a 1.5 tesla static magnetic field does not affect somato-sensory-evoked potentials in man. Magn Reson Imaging 1990;8(1):65-9.

183. Shellock FG, Schaefer DJ, Gordon CJ. Effect of a 1.5 T static magnetic field on body temperature of man. Magn Reson Med 1986 Aug;3(4):644-7.

184. Budinger TF. Nuclear magnetic resonance (NMR) in vivo studies: known thresholds for health effects. J Comput Assist Tomogr 1981 Dec;5(6):800-11.

185. Brice J. Experts put MRI accreditation program to the test. Diagn Imaging 2001 Jul;23(7):44-7, 49.

186. Skaane P. Ultrasonography as adjunct to mammography in the evaluation of breast tumors. Acta Radiol Suppl 1999;420:7-47.

187. Flobbe K. Costs and effects of ultrasonography in the evaluation of palpable breast masses. Int J Technol Assess Health Care 2004 Fall;20(4):440-8.

188. Flobbe K, Bosch AM, Kessels AG, et al. The additional diagnostic value of ultrasonography in the diagnosis of breast cancer. Arch Intern Med 2003 May 26;163(10):1194-9.

189. Meyberg-Solomayer GC, Kraemer B, Bergmann A, et al. Does 3-D sonography bring any advantage to noninvasive breast diagnostics. Ultrasound Med Biol 2004 May;30(5):583-9.

190. Chao TC, Lo YF, Chen SC, et al. Prospective sonographic study of 3093 breast tumors. J Ultrasound Med 1999 May;18(5):363-70.

191. Chen SC, Cheung YC, Su CH, et al. Analysis of sonographic features for the differentiation of benign and malignant breast tumors of different sizes. Ultrasound Obstet Gynecol 2004 Feb;23(2):188-93.

192. Perre CI, Koot VC, de Hooge P, et al. The value of ultrasound in the evaluation of palpable breast tumours: a prospective study of 400 cases. Eur J Surg Oncol 1994 Dec;20(6):637-40.

193. van Oord JC. The value of ultrasound mammography in palpable breast masses. Rofo 1991 Jul;155(1):63-6.

194. McNicholas MM, Mercer PM, Miller JC, et al. Color Doppler sonography in the evaluation of palpable breast masses. AJR Am J Roentgenol 1993 Oct;161(4):765-71.

195. American Institute of Ultrasound in Medicine. AIUM standard for the performance of breast ultrasound examination. Laurel (MD): American Institute of Ultrasound in Medicine; 2002. 6 p.
Also available: http://www.aium.org/
publications/clinical/breast.pdf.

196. American College of Radiology. Breast ultrasound accreditation program requirements.
Reston (VA): American College of Radiology; 2005. 9 p.
Also available: http://www.acr.org/s_acr/bin.asp?CID=591&DID=12137&DOC=FILE
.PDF.

197. Watson L. The role of ultrasound in breast imaging. Radiol Technol 2000;71(5):441-59; quiz 460-2.

198. Baker JA, Soo MS, Rosen EL. Artifacts and pitfalls in sonographic imaging of the breast. AJR Am J Roentgenol 2001 May;176(5):1261-6 .

199. Drukker K, Giger ML, Horsch K, et al. Computerized lesion detection on breast ultrasound. Med Phys 2002 Jul;29(7):1438-46.

200. Baker JA, Kornguth PJ, Soo MS, et al. Sonography of solid breast lesions: observer variability of lesion

description and assessment. AJR Am J Roentgenol 1999 Jun;172(6):1621-5.

201. Weinstein SP, Conant EF, Mies C, et al. Posterior acoustic shadowing in benign breast lesions: sonographic-pathologic correlation. J Ultrasound Med 2004 Jan;23(1):73-83.

202. Rahbar G, Sie AC, Hansen GC, et al. Benign versus malignant solid breast masses: US differentiation. Radiology 1999 Dec;213(3):889-94 .

203. Murad M, Bari V. Ultrasound differentiation of benign versus malignant solid breast masses. J Coll Physicians Surg Pak 2004 Mar;14(3):166-9.

204. Skaane P, Engedal K, Skjennald A. Interobserver variation in the interpretation of breast imaging. Comparison of mammography, ultrasonography, and both combined in the interpretation of palpable noncalcified breast masses. Acta Radiol 1997 Jul;38(4 Pt 1):497-502.

205. Bosch AM, Kessels AG, Beets GL, et al. Interexamination variation of whole breast ultrasound. Br J Radiol 2003 May;76(905):328-31.

206. Shimamoto K, Sawaki A, Ikede M, et al. Interobserver agreement in sonographic diagnosis of breast tumors. Eur J Ultrasound 1998 Sep;8(1):25-31.

207. Smith AP, Hall PA, Marcello DM. Emerging technologies in breast cancer detection. Radiol Manage 2004;26(4):16-24; quiz 25-7.

208. Siegmann KC, Muller-Schimpfle M, Schick F, et al. MR imaging-detected breast lesions: histopathologic correlation of lesion characteristics and signal intensity data. AJR Am J Roentgenol 2002 Jun;178(6):1403-9 .

209. Boonjunwetwat D, Prathombutr A. Imaging of benign papillary neoplasm of the breast: mammographic, galactographic and sonographic findings. J Med Assoc Thai 2000 Aug;83(8):832-8.

210. Adler LP, Crowe JP, al-Kaisi NK, et al. Evaluation of breast masses and axillary lymph nodes with [F-18] 2-deoxy-2-fluoro-D-glucose PET. Radiology 1993 Jun;187(3):743-50.

211. Avril N, Bense S, Ziegler SI, et al. Breast imaging with fluorine-18-FDG PET: quantitative image analysis. J Nucl Med 1997 Aug;38(8):1186-91.

212. Avril N, Dose J, Janicke F, et al. Metabolic characterization of breast tumors with positron emission tomography using F-18 fluorodeoxyglucose. J Clin Oncol 1996 Jun;14(6):1848-57.

213. Avril N, Rose CA, Schelling M, et al. Breast imaging with positron emission tomography and fluorine-18 fluorodeoxyglucose: use and limitations. J Clin Oncol 2000 Oct 15;18(20):3495-502.

214. Buck A, Schirrmeister H, Kuhn T, et al. FDG uptake in breast cancer: correlation with biological and clinical prognostic parameters. Eur J Nucl Med Mol Imaging 2002 Oct;29(10):1317-23.

215. Danforth DN Jr, Aloj L, Carrasquillo JA, et al. The role of 18F-FDG-PET in the local/regional evaluation of women with breast cancer. Breast Cancer Res Treat 2002 Sep;75(2):135-46.

216. Fujiwara T, Miyake M, Watanuki S, et al. Easy detection of tumor in oncologic whole-body PET by projection reconstruction images with maximum intensity projection algorithm. Ann Nucl Med 1999 Jun;13(3):199-203 .

217. Inoue T, Yutani K, Taguchi T, et al. Preoperative evaluation of prognosis in breast cancer patients by [(18)F]2-Deoxy-2-fluoro-D-glucose-positron emission tomography. J Cancer Res Clin Oncol 2004 May;130(5):273-8.

218. Murthy K, Aznar M, Thompson CJ, et al. Results of preliminary clinical trials of the positron emission mammography system PEM-I: a dedicated breast imaging system producing glucose metabolic images using FDG. J Nucl Med 2000 Nov;41(11):1851-8.

219. Nieweg OE, Wong WH, Singletary SE, et al. Positron emission tomography of glucose metabolism in breast cancer. Potential for tumor detection, staging, and evaluation of chemotherapy. Ann N Y Acad Sci 1993 Nov 30;698:423-8.

220. Noh DY, Yun IJ, Kim JS, et al. Diagnostic value of positron emission tomography for detecting breast cancer. World J Surg 1998 Mar;22(3):223-7; discussion 227-8.

221. Pietrzyk U, Scheidhauer K, Scharl A, et al. Presurgical visualization of primary breast carcinoma with PET emission and transmission imaging. J Nucl Med 1995 Oct;36(10):1882-4.

222. Scheidhauer K, Scharl A, Pietrzyk U, et al. Qualitative [18F]FDG positron emission tomography in primary breast cancer: clinical relevance and practicability. Eur J Nucl Med 1996 Jun;23(6):618-23.

223. Smyczek-Gargya B. PET with [18F]fluorothymidine for imaging of primary breast cancer: a pilot study. Eur J Nucl Med Mol Imaging 2004 May;31(5):720-4.

224. Arslan N, Ozturk E, Ilgan S, et al. 99Tcm-MIBI scintimammography in the evaluation of breast lesions and axillary involvement: a comparison with mammography and histopathological diagnosis. Nucl Med Commun 1999 Apr;20(4):317-25.

225. Aziz A, Hashmi R, Ogawa Y, et al. Tc-99m-MIBI scintimammography; SPECT versus planar imaging. Cancer Biother Radiopharm 1999 Dec;14(6):495-500.

226. Becherer A, Helbich T, Staudenherz A, et al. The diagnostic value of planar and SPET scintimammography in different age groups. Nucl Med Commun 1997 Aug;18(8):710-8.

227. Buscombe JR, Cwikla JB, Holloway B, et al. Prediction of the usefulness of combined mammography and scintimammography in suspected primary breast cancer using ROC curves. J Nucl Med 2001 Jan;42(1):3-8.

228. Buscombe JR, Cwikla JB, Thakrar DS, et al. Uptake of Tc-99m MIBI related to tumour size and type. Anticancer Res 1997;17(3B):1693-4.

229. Clifford EJ, Lugo-Zamudio C. Scintimammography in the diagnosis of breast cancer. Am J Surg 1996 Nov;172(5):483-6.

230. Cwikla JB, Buscombe JR, Kelleher SM, et al. Comparison of accuracy of scintimammography and X-ray mammography in the diagnosis of primary breast cancer in patients selected for surgical biopsy. Clin Radiol 1998 Apr;53(4):274-80.

231. Danielsson R, Sanchez-Crespo A, Pegerfalk A, et al. 99mTc-sestamibi uptake and histological malignancy grade in invasive breast carcinoma. Eur J Nucl Med Mol Imaging 2003 May;30(5):662-6.

232. Dunnwald LK, Hartnett SD, Mankoff DA. Utility and reproducibility of semiquantitative analysis of sestamibi breast images. J Nucl Med Technol 1997 Jun;25(2):106-9.

233. Farias-Jimenez J, Rivera A, Carlo VM. Technetium-99m sestamibi in the diagnosis of breast cancer. The Mayaguez Medical Center experience. Bol Asoc Med P R 2001;93(1-12):9-11.

234. Fleming RM. Mitochondrial uptake of sestamibi distinguishes between normal, inflammatory breast changes, pre-cancers, and infiltrating breast cancer. Integr Cancer Ther 2002 Sep;1(3):229-37.

235. Fleming RM, Dooley WC. Breast enhanced scintigraphy testing distinguishes between normal, inflammatory breast changes, and breast cancer: a prospective analysis and comparison with mammography. Integr Cancer Ther 2002 Sep;1(3):238-45.

236. Horne T, Pappo I, Cohenpour M, et al. 99Tcm-MIBI scintimammography for the detection of breast malignancies: the contribution of the count ratio to specificity. Nucl Med Commun 1999 Jun;20(6):511-6.

237. Horne T, Pappo I, Cohen-Pour M, et al. 99Tc(m)-tetrofosmin scintimammography for detecting breast cancer: a comparative study with 99Tc(m)-MIBI. Nucl Med Commun 2001 Jul;22(7):807-11.

238. Howarth D, Sillar R, Lan L, et al. Scintimammography: an adjunctive test for the detection of breast cancer. Med J Aust 1999 Jun 21;170(12):588-91.

239. Howarth D, Sillar R, Clark D, et al. Technetium-99m sestamibi scintimammography: the influence of

histopathological characteristics, lesion size and the presence of carcinoma in situ in the detection of breast carcinoma. Eur J Nucl Med 1999 Nov;26(11):1475-81.

240. Iraniha S, Khalkhali I I, Cutrone JA, et al. Breast Cancer Imaging: Can Tc-99m Sestamibi Scintimammography Fit In? Medscape Womens Health 1997 Apr;2(4):2.

241. Khalkhali I, Cutrone JA, Mena IG, et al. Scintimammography: the complementary role of Tc-99m sestamibi prone breast imaging for the diagnosis of breast carcinoma. Radiology 1995 Aug;196(2):421-6.

242. Khalkhali I, Cutrone J, Mena I, et al. Technetium-99m-sestamibi scintimammography of breast lesions: clinical and pathological follow-up. J Nucl Med 1995 Oct;36(10):1784-9.

243. Khalkhali I, Mena I, Jouanne E, et al. Prone scintimammography in patients with suspicion of carcinoma of the breast. J Am Coll Surg 1994 May;178(5):491-7.

244. Khalkhali I, Iraniha S, Cutrone JA, et al. Scintimammography with Tc-99m sestamibi. Acta Med Austriaca 1997;24(2):46-9.

245. Kim SJ. Incremental diagnostic value of quantitative analysis of double phase Tc-99m MIBI scintimammography for the detection of primary breast cancer additive to visual analysis. Breast Cancer Res Treat 2004 Jan;83(2):129-38.

246. Lumachi F, Marzola MC, Zucchetta P, et al. Breast cancer detection with 99m-Tc-sestamibi scintigraphy, mammography, and fine-needle aspiration cytology: comparative study in 64 surgically treated patients. Ann Surg Oncol 1999 Sep;6(6):568-71.

247. Lumachi F, Ferretti G, Povolato M, et al. Accuracy of technetium-99m sestamibi scintimammography and X-ray mammography in premenopausal women with suspected breast cancer. Eur J Nucl Med 2001 Dec;28(12):1776-80.

248. Lumachi F, Ferretti G, Povolato M, et al. Sestamibi scintimammography in pT1 breast cancer: alternative or complementary to X-ray mammography. Anticancer Res 2001;21(3C):2201-5.

249. Lumachi F, Zucchetta P, Marzola MC, et al. Positive predictive value of 99mTc sestamibi scintimammography in patients with non-palpable, mammographically detected, suspicious, breast lesions. Nucl Med Commun 2002 Nov 1;23(11):1073-8.

250. Lumachi F, Ferretti G, Povolato M, et al. Usefulness of 99m-Tc-sestamibi scintimammography in suspected breast cancer and in axillary lymph node metastases detection. Eur J Surg Oncol 2001 Apr;27(3):256-9.

251. Massardo T, Alonso O, Kabasakal L, et al. Diagnostic value of 99mTc-methylene diphosphonate and 99mTc-pentavalent DMSA compared with 99mTc-sestamibi for palpable breast lesions. J Nucl Med 2002 Jul;43(7):882-8.

252. Maurer AH, Caroline DF, Jadali FJ, et al. Limitations of craniocaudal thallium-201 and technetium-99m-sestamibi mammoscintigraphy. J Nucl Med 1995 Sep;36(9):1696-700.

253. Myslivecek M, Koranda P, Kaminek M, et al. Technetium-99m-MIBI scintimammography by planar and SPECT imaging in the diagnosis of breast carcinoma and axillary lymph node involvement. Nucl Med Rev 2004;7(2):151-5.

254. Nishiyama Y, Yamamoto Y, Ono Y, et al. Comparative evaluation of 99mTc-MIBI and 99mTc-HMDP scintimammography for the diagnosis of breast cancer and its axillary metastases. Eur J Nucl Med 2001 Apr;28 (4):522-8.

255. Palmedo H, Biersack HJ, Lastoria S, et al. Scintimammography with technetium-99m methoxyisobutylisonitrile: results of a prospective European multicentre trial. Eur J Nucl Med 1998 Apr;25(4):375-85.

256. Papantoniou V, Christodoulidou J, Papadaki E, et al. Uptake and washout of 99mTcV-dimercaptosuccinic acid and 99mTc-sestamibi in the assessment of histological type and grade in breast cancer. Nucl Med Commun 2002 May;23(5):461-7.

257. Polan RL, Klein BD, Richman RH. Scintimammography in patients with minimal mammographic or clinical findings. Radiographics 2001;21(3):641-53; discussion 653-5.

258. Prats E, Aisa F, Abos MD, et al . Mammography and 99mTc-MIBI scintimammography in suspected breast cancer. J Nucl Med 1999 Feb;40(2):296-301.

259. Ren C, Jin S, Zhou Q, et al. Clinical significance of 99mTc-MIBI breast imaging in the diagnosis of early breast cancer. Asian J Surg 2002 Apr;25(2):126-9.

260. Sampalis FS, Denis R, Picard D, et al. International prospective evaluation of scintimammography with technetium-99m sestamibi: interim results. Am J Surg 2001 Oct;182(4):399-403.

261. Scopinaro F, Pani R, De Vincentis G, et al. ,High-resolution scintimammography improves the accuracy of technetium-99m methoxyisobutylisonitrile scintimammography: use of a new dedicated gamma camera. Eur J Nucl Med 1999 Oct;26(10):1279-88.

262. Sillar R, Howarth D, Clark D. The initial Australian experience of technetium-99M sestamibi scintimammography: a complementary test in the management of breast cancer. Aust N Z J Surg 1997 Jul;67(7):433-7.

263. Taillefer R. Technetium-99m-sestamibi prone scintimammography to detect primary breast cancer and axillary lymph node involvement. J Nucl Med 1995 Oct;36(10):1758-65.

264. Tiling R, Tatsch K, Sommer H, et al. Technetium-99m-sestamibi scintimammography for the detection of breast carcinoma: comparison between planar and SPECT imaging. J Nucl Med 1998 May;39(5):849-56.

265. Tiling R, Kessler M, Untch M, et al. Initial evaluation of breast cancer using Tc-99m sestamibi scintimammography. Eur J Radiol 2005;53(2):206-12.

266. Tolmos J, Cutrone JA, Wang B, et al. Scintimammographic analysis of nonpalpable breast lesions previously identified by conventional mammography. J Natl Cancer Inst 1998 Jun 3;90(11):846-9.

267. Vargas HI, Agbunag RV, Kalinowski A, et al. The clinical utility of Tc-99m sestamibi scintimammography in detecting multicentric breast cancer. Am Surg 2001 Dec;67(12):1204-8.

268. Barbacioru C, Arunachalam A, Cowden D, et al. Pharmacokinetic mapping of breast tumors: a new statistical analysis technique for dynamic magnetic resonance imaging. AMIA Symposium Proceedings 2003;783.

269. Baum E, Fischer U, Vosshenrich R, et al. Classification of hypervascularized lesions in CE MR imaging of the breast. Eur Radiol 2002 May 1;12(5):1087-92.

270. Baum F, Fischer U, Vosshenrich R, et al. Classification of hypervascularized lesions in CE MR imaging of the breast. Eur Radiol 2002 May;12(5):1087-92.

271. Meriggi F. Surgical therapy of colorectal metastases - state of the art. Rom J Gastroenterol 2004;13(3):233-5.

272. Bone B, Szabo BK, Perbeck LG, et al. Can contrast-enhanced MR imaging predict survival in breast cancer. Acta Radiol 2003 Jul;44(4):373-8.

273. Carriero A, Di Credico A, Mansour M, et al. Maximum intensity projection analysis in magnetic resonance of the breast. J Exp Clin Cancer Res 2002 Sep;21(3 Suppl):77-81.

274. Choi BG, Kim HH, Kim EN, et al. New subtraction algorithms for evaluation of lesions on dynamic contrast-enhanced MR mammography. Eur Radiol 2002 Dec;12(12):3018-22.

275. Fenlon HM, Phelan NC, O'Sullivan P, et al. Benign versus malignant breast disease: comparison of contrast-enhanced MR imaging and Tc-99m tetrofosmin scintimammography. Radiology 1997 Oct;205(1):214-20.

276. Furman B, Gardner MS, Romilly P, et al. Effect of 0.5 Tesla magnetic resonance imaging on the surgical management of breast cancer patients. Am J Surg 2003 Oct;186(4):344-7.

277. Gibbs P. Differentiation of benign and malignant sub-1 cm breast lesions using dynamic contrast enhanced MRI. Breast 2004 Apr;13(2):115-21.

278. German Agency for Health Technology Assessment. Assessment of magnetic resonance imaging in diagnostics of breast carcinoma - systematic review. Koln: German Agency for Health Technology Assessment at the German Institute for Medical Documentation and Information (DAHTA) (DIMDI); 1998. Also available: http://www.mrw.interscience.wiley.com/cochrane/clhta/articles/HTA-999821/frame.html.

279. Guo Y, Cai YQ, Cai ZL, et al. Differentiation of clinically benign and malignant breast lesions using diffusion-weighted imaging. J Magn Reson Imaging 2002 Aug;16(2):172-8.

280. Jacobs MA, Barker PB, Bottomley PA, et al. Proton magnetic resonance spectroscopic imaging of human breast cancer: a preliminary study. J Magn Reson Imaging 2004 Jan;19(1):68-75.

281. Jacobs MA. Benign and malignant breast lesions: diagnosis with multiparametric MR imaging. Radiology 2003 Oct;229(1):225-32.

282. Kaiser W. MRI of the female breast. First clinical results. Arch Int Physiol Biochim 1985 Dec;93(5):67-76.

283. Khatri VP, Stuppino JJ, Espinosa MH, et al. Improved accuracy in differentiating malignant from benign mammographic abnormalities: a simple, improved magnetic resonance imaging method. Cancer 2001 Aug 1;92(3):471-8.

284. Kim SJ, Morris EA, Liberman L, et al. Observer variability and applicability of BI-RADS terminology for breast MR imaging: invasive carcinomas as focal masses. AJR Am J Roentgenol 2001 Sep;177(3):551-7.

285. Kramer S, Schulz-Wendtland R, Hagedorn K, et al. Magnetic resonance imaging and its role in the diagnosis of multicentric breast cancer. Anticancer Res 1998;18(3C):2163-4.

286. Lee SG, Orel SG, Woo IJ, et al. MR imaging screening of the contralateral breast in patients with newly diagnosed breast cancer: preliminary results. Radiology 2003 Mar;226(3):773-8.

287. Lucht RE, Knopp MV, Brix G. Classification of signal-time curves from dynamic MR mammography by neural networks. Magn Reson Imaging 2001 Jan;19(1):51-7.

288. Nagashima T, Suzuki M, Yagata H, et al. Dynamic-enhanced MRI predicts metastatic potential of invasive ductal breast cancer. Breast Cancer 2002;9(3):226-30.

289. Nakahara H, Namba K, Wakamatsu H, et al. Extension of breast cancer: comparison of CT and MRI. Radiat Med 2002;20(1):17-23.

290. Obenauer S, Schorn C, Stelter B, et al. Contrast-enhanced high in-plane resolution dynamic MRI of the breast. Are there advantages in comparison to standard dynamic MRI? Clin Imaging 2002;26(3):161-5.

291. Reinikainen H, Paakko E, Suramo I, et al. Dynamics of contrast enhancement in MR imaging and power Doppler ultrasonography of solid breast lesions. Acta Radiol 2002 Sep;43(5):492-500.

292. Rigauts H, Casselman J, Steyaert L, et al. Contribution of MRI and color Doppler sonography in breast cancer diagnosis. J Belge Radiol 1993 Aug;76(4):226-31.

293. Schelfout K, Van Goethem M, Kersschot E, et al. Contrast-enhanced MR imaging of breast lesions and effect on treatment. Eur J Surg Oncol 2004 Jun;30(5):501-7.

294. Schelfout K, Van Goethem M, Kersschot E, et al. Preoperative breast MRI in patients with invasive lobular breast cancer. Eur Radiol 2004 Jul;14(7):1209-16.

295. Shahar KH, Solaiyappan M, Bluemke DA. Quantitative differentiation of breast lesions based on three-dimensional morphology from magnetic resonance imaging. J Comput Assist Tomogr 2002;26(6):1047-53.

296. Slanetz PJ, Edmister WB, Yeh ED, et al. Occult contralateral breast carcinoma incidentally detected by breast magnetic resonance imaging. Breast J 2002;8(3):145-8.

297. Szabo BK, Aspelin P, Kristoffersen Wiberg M, et al. Invasive breast cancer: correlation of dynamic MR features with prognostic factors. Eur Radiol 2003 Nov;13(11):2425-35.

298. Teifke A, Hlawatsch A, Beier T, et al. Undetected malignancies of the breast: dynamic contrast-enhanced MR imaging at 1.0 T. Radiology 2002 Sep;224(3):881-8.

299. Trecate G, Tess JD, Vergnaghi D, et al. Breast microcalcifications studied with 3D contrast-enhanced high-field magnetic resonance imaging: more accuracy in the diagnosis of breast cancer. Tumori 2002;88(3):224-33.

300. Tuncbilek N, Unlu E, Karakas HM, et al. Evaluation of tumor angiogenesis with contrast-enhanced dynamic magnetic resonance mammography. Breast J 2003;9(5):403-8.

301. Van Goethem M, Schelfout K, Dijckmans L, et al. MR mammography in the pre-operative staging of breast cancer in patients with dense breast tissue: comparison with mammography and ultrasound. Eur Radiol 2004 May;14(5):809-16.

302. Wedegartner U, Bick U, Wortler K, et al. Differentiation between benign and malignant findings on MR-mammography: usefulness of morphological criteria . Eur Radiol 2001;11(9):1645-50.

303. Louie L, Velez N, Earnest D, et al. Management of nonpalpable ultrasound-indeterminate breast lesions. Surgery 2003 Oct;134(4):667-73; discussion 673-4.

304. Kristoffersen Wiberg M, Aspelin P, Sylvan M, et al. Comparison of lesion size estimated by dynamic MR imaging, mammography and histopathology in breast neoplasms. Eur Radiol 2003 Jun;13(6):1207-12.

305. Yeh ED, Slanetz PJ, Edmister WB, et al. Invasive lobular carcinoma: spectrum of enhancement and morphology on magnetic resonance imaging. Breast J 2003;9(1):13-8.

306. Zuiani C, Francescutti GE, Londero V, et al. Ductal carcinoma in situ: is there a role for MRI. J Exp Clin Cancer Res 2002 Sep;21(3 Suppl):89-95.

307. Allen SA, Cunliffe WJ, Gray J, et al. Pre-operative estimation of primary breast cancer size: a comparison of clinical assessment, mammography and ultrasound. Breast 2001;10(4):299-305.

308. Arger PH, Sehgal CM, Conant EF, et al. Interreader variability and predictive value of US descriptions of solid breast masses: pilot study. Acad Radiol 2001 Apr;8(4):335-42.

309. Baker JA, Soo MS. Breast US: assessment of technical quality and image interpretation. Radiology 2002 Apr;223(1):229-38.

310. Beeckman P, De Clerck S, Jong B, et al. The ultrasound aspect of the skin and subcutaneous fat layer in various benign and malignant breast conditions. J Belge Radiol 1991;74(4):283-8.

311. Bhatti PT, LeCarpentier GL, Roubidoux MA, et al. Discrimination of sonographically detected breast masses using frequency shift color Doppler imaging in combination with age and gray scale criteria. J Ultrasound Med 2001 Apr;20(4):343-50.

312. Bosch AM, Kessels AG, Beets GL, et al. Preoperative estimation of the pathological breast tumour size by physical examination, mammography and ultrasound: a prospective study on 105 invasive tumours. Eur J Radiol 2003 Dec;48(3):285-92.

313. Butler RS, Venta LA, Wiley EL, et al. Sonographic evaluation of infiltrating lobular carcinoma. AJR Am J Roentgenol 1999 Feb;172(2):325-30.

314. Britton PD, Coulden RA. The use of duplex Doppler ultrasound in the diagnosis of breast cancer. Clin Radiol 1990 Dec;42(6):399-401.

315. Buadu LD, Murakami J, Murayama S, et al. Colour Doppler sonography of breast masses: a multiparameter analysis. Clin Radiol 1997;52:917-23.

316. Carson PL, Moskalik AP, Govil A, et al. The 3D and 2D color flow display of breast masses. Ultrasound Med

Biol 1997;23(6):837-49.

317. Kawashima M, Tamaki Y, Nonaka T, et al. MR imaging of mucinous carcinoma of the breast. AJR Am J Roentgenol 2002 Jul;179(1):179-83.

318. Chandawarkar RY, Shinde SR. Preoperative diagnosis of carcinoma of the breast: is a "cost-cutter" algorithm tenable? J Surg Oncol 1997 Feb;64(2):153-8.

319. Chao TC, Lo YF, Chen SC, et al. Color Doppler ultrasound in benign and malignant breast tumors. Breast Cancer Res Treat 1999 Sep;57(2):193-9.

320. Chen DR, Jeng LB, Kao A, et al. Comparing thallium-201 spect mammoscintigraphy and ultrasonography to detect breast cancer in mammographical dense breasts. Neoplasma 2003;50(3):222-6.

321. Chen DR, Chang RF, Kuo WJ, et al. Diagnosis of breast tumors with sonographic texture analysis using wavelet transform and neural networks. Ultrasound Med Biol 2002 Oct;28(10):1301-10.

322. Chen WM, Chang RF, Moon WK, et al. Breast cancer diagnosis using three-dimensional ultrasound and pixel relation analysis. Ultrasound Med Biol 2003 Jul;29(7):1027-35.

323. Choi HY, Kim HY, Baek SY, et al. Significance of resistive index in color Doppler ultrasonogram: differentiation between benign and malignant breast masses. Clin Imaging 1999;23(5):284-8.

324. Ciatto S, Rosselli del Turco M, Catarzi S, et al. The contribution of ultrasonography to the differential diagnosis of breast cancer. Neoplasma 1994;41(6):341-5.

325. Cilotti A, Bagnolesi P, Moretti M, et al. Comparison of the diagnostic performance of high-frequency ultrasound as a first- or second-line diagnostic tool in non-palpable lesions of the breast. Eur Radiol 1997;7(8):1240-4.

326. Cimitan M, Volpe R, Candiani E, et al. The use of thallium-201 in the preoperative detection of breast cancer: an adjunct to mammography and ultrasonography. Eur J Nucl Med 1995 Oct;22(10):1110-7.

327. Cosgrove DO, Kedar RP, Bamber JC, et al. Breast diseases: color Doppler US in differential diagnosis. Radiology 1993 Oct;189(1):99-104.

328. Cosmacini P, Veronesi P, Galimberti V, et al. Ultrasonographic evaluation of palpable breast masses: analysis of 134 cases. Tumori 1990 Oct 31;76(5):495-8.

329. Debniak B, Debniak T, Breborowicz GH, et al. The usefulness of Doppler angiography in differential diagnosis of benign and malignant breast lesions. Int J Fertil Womens Med 2004;49(3):138-43.

330. Delorme S, Zuna I, Huber S, et al. Colour Doppler sonography in breast tumours: an update. Eur Radiol 1998;8(2):189-93.

331. Dixon JM, Walsh J, Paterson D, et al. Colour Doppler ultrasonography studies of benign and malignant breast lesions. Br J Surg 1992 Mar;79(3):259-60.

332. Dock W. Duplex sonography of mammary tumors: a prospective study of 75 patients. J Ultrasound Med 1993 Feb;12(2):79-82.

333. Drukker K, Giger ML, Mendelson EB. Computerized analysis of shadowing on breast ultrasound for improved lesion detection. Med Phys 2003 Jul;30(7):1833-42.

334. Edde DJ. Whole-breast compression ultrasonography with the patient in the sitting position. Can Assoc Radiol J 1994 Aug;45(4):324-6.

335. Staren ED. Surgical office-based ultrasound of the breast. Am Surg 1995 Jul;61(7):619-26; discussion 627.

336. Eltahir A, Jibril JA, Squair J, et al. The accuracy of 'one-stop' diagnosis for 1,110 patients presenting to a symptomatic breast clinic. J R Coll Surg Edinb 1999 Aug;44(4):226-30.

337. Finlay ME, Liston JE, Lunt LG, et al. Assessment of the role of ultrasound in the differentiation of radial scars and stellate carcinomas of the breast. Clin Radiol 1994 Jan;49(1):52-5.

338. Forsberg F, Goldberg BB, Merritt CR, et al. Diagnosing breast lesions with contrast-enhanced 3-dimensional power Doppler imaging. J Ultrasound Med 2004 Feb;23(2):173-82.

339. Foxcroft LM, Evans EB, Porter AJ. The diagnosis of breast cancer in women younger than 40. Breast 2004 Aug;13(4):297-306.

340. Fung HM, Jackson FI. Clinically and mammographically occult breast lesions demonstrated by ultrasound. J R Soc Med 1990 Nov;83(11):696-8.

341. Garra BS, Krasner BH, Horii SC, et al. Improving the distinction between benign and malignant breast lesions: the value of sonographic texture analysis. Ultrason Imaging 1993 Oct;15(4):267-85.

342. Gefen S, Tretiak OJ, Piccoli CW, et al. ROC analysis of ultrasound tissue characterization classifiers for breast cancer diagnosis. IEEE Trans Med Imaging 2003 Feb;22(2):170-7.

343. Germer U, Tetzlaff A, Geipel A, et al. Strong impact of estrogen environment on Doppler variables used for differentiation between benign and malignant breast lesions. Ultrasound Obstet Gynecol 2002 Apr;19(4):380-5.

344. Giuseppetti GM , Baldassarre S, Marconi E. Color Doppler sonography. Eur J Radiol 1998 May;27(Suppl 2):S254-8.

345. Golshan M, Fung BB, Wolfman J, et al. The effect of ipsilateral whole breast ultrasonography on the surgical management of breast carcinoma. Am J Surg 2003 Oct;186(4):391-6.

346. Golub RM. Differentiation of breast tumors by ultrasonic tissue characterization. J Ultrasound Med 1993 Oct;12(10):601-8.

347. Graf O, Helbich TH, Fuchsjaeger MH, et al. Follow-up of palpable circumscribed noncalcified solid breast masses at mammography and US: can biopsy be averted? Radiology 2004 Dec;233(3):850-6.

348. Hieken TJ, Harrison J, Herreros J, et al. Correlating sonography, mammography, and pathology in the assessment of breast cancer size. Am J Surg 2001 Oct;182(4):351-4.

349. Hieken TJ, Velasco JM. A prospective analysis of office-based breast ultrasound. Arch Surg 1998 May;133(5):504-7; discussion 507-8.

350. Hollerweger A, Rettenbacher T, Macheiner P, et al. New signs of breast cancer: high resistance flow and variations in resistive indices evaluation by color Doppler sonography. Ultrasound Med Biol 1997;23(6):851-6.

351. Houssami N, Irwig L, Simpson JM, et al. Sydney Breast Imaging Accuracy Study: Comparative sensitivity and specificity of mammography and sonography in young women with symptoms. AJR Am J Roentgenol 2003 Apr;180(4):935-40.

352. Huang CS, Wu CY, Chu JS, et al. Microcalcifications of non-palpable breast lesions detected by ultrasonography: correlation with mammography and histopathology. Ultrasound Obstet Gynecol 1999 Jun;13(6):431-6.

353. Huber S, Helbich T, Kettenbach J, et al. Effects of a microbubble contrast agent on breast tumors: computer-assisted quantitative assessment with color Doppler US--early experience. Radiology 1998 Aug;208(2):485-9.

354. Huber S, Delorme S, Knopp MV, et al. Breast tumors: computer-assisted quantitative assessment with color Doppler US. Radiology 1994 Sep;192(3):797-801.

355. Jackson VP, Kelly-Fry E, Rothschild PA, et al. Automated breast sonography using a 7.5-MHz PVDF transducer: preliminary clinical evaluation. Work in progress. Radiology 1986;159(3):679-84.

356. Joo S, Yang YS, Moon WK, et al. Computer-aided diagnosis of solid breast nodules: use of an artificial neural network based on multiple sonographic features. IEEE Trans Med Imaging 2004 Oct;23(10):1292-300.

357. Kaiser JS, Helvie MA, Blacklaw RL, et al. Palpable breast thickening: role of mammography and US in cancer detection. Radiology 2002 Jun;223(3):839-44.

358. Kaplan SS. Clinical utility of bilateral whole-breast US in the evaluation of women with dense breast tissue. Radiology 2001 Dec;221(3):641-9.

359. Kedar RP, Cosgrove DO, Bamber JC, et al. Automated quantification of color Doppler signals: a preliminary study in breast tumors. Radiology 1995 Oct;197(1):39-43.

360. Kimme-Smith C, Bassett LW, Gold RH. High frequency breast ultrasound. Hand-held versus automated units; examination for palpable mass versus screening. J Ultrasound Med 1988;7(2):77-81.

361. Kolb TM, Lichy J, Newhouse JH. Comparison of the performance of screening mammography, physical examination, and breast US and evaluation of factors that influence them: an analysis of 27,825 patient evaluations. Radiology 2002 Oct;225(1):165-75.

362. Kolb TM, Lichy J, Newhouse JH. Occult cancer in women with dense breasts: detection with screening US--diagnostic yield and tumor characteristics. Radiology 1998 Apr;207(1):191-9.

363. Ultrasound detects missed cancers. Biomed Instrum Technol 1997;31(3):217-8.

364. Kook SH, Kwag HJ. Value of contrast-enhanced power Doppler sonography using a microbubble echo-enhancing agent in evaluation of small breast lesions. J Clin Ultrasound 2003 Jun;31(5):227-38.

365. Kook SH, Park HW, Lee YR, et al . Evaluation of solid breast lesions with power Doppler sonography. J Clin Ultrasound 1999 Jun;27(5):231-7.

366. Krestan CR, Riedl C, Memarsadeghi M, et al. 3D-power Doppler ultrasound of breast lesions with a microbubble contrast agent. Acad Radiol 2002 Aug;9(Suppl 2):S384-5.

367. Kuijpers TJ, Obdeijn AI, Kruyt RH, et al. Solid breast neoplasms: differential diagnosis with pulsed Doppler ultrasound. Ultrasound Med Biol 1994;20(6):517-20.

368. Lam WW, Chu WC, Tse GM, et al. Sonographic appearance of mucinous carcinoma of the breast. AJR Am J Roentgenol 2004 Apr;182(4):1069-74.

369. Leconte I, Feger C, Galant C, et al. Mammography and subsequent whole-breast sonography of nonpalpable breast cancers: the importance of radiologic breast density. AJR Am J Roentgenol 2003 Jun;180(6):1675-9 .

370. Lee SK, Lee T, Lee KR, et al. Evaluation of breast tumors with color Doppler imaging: a comparison with image-directed Doppler ultrasound. J Clin Ultrasound 1995;23(6):367-73.

371. Lee WJ, Chu JS, Huang CS, et al . Breast cancer vascularity: color Doppler sonography and histopathology study. Breast Cancer Res Treat 1996;37(3):291-8.

372. Leung JW, Kornguth PJ, Gotway MB. Utility of targeted sonography in the evaluation of focal breast pain. J Ultrasound Med 2002 May;21(5):521-6.

373. Lister D, Evans AJ, Burrell HC, et al. The accuracy of breast ultrasound in the evaluation of clinically benign discrete, symptomatic breast lumps. Clin Radiol 1998 Jul;53(7):490-2.

374. Nunes LW, Englander SA, Charafeddine R, et al. Optimal post-contrast timing of breast MR image acquisition for architectural feature analysis. J Magn Reson Imaging 2002 Jul;16(1):42-50.

375. Madjar H, Prompeler HJ, Sauerbrei W, et al. Differential diagnosis of breast lesions by color Doppler. Ultrasound Obstet Gynecol 1995 Sep;6(3):199-204.

376. Madjar H, Sauerbrei W, Prompeler HJ, et al. Color Doppler and duplex flow analysis for classification of breast lesions. Gynecol Oncol 1997 Mar; 64(3):392-403.

377. Madjar H, Prompeler HJ, Sauerbrei W, et al. Color Doppler flow criteria of breast lesions. Ultrasound Med Biol

1994;20(9):849-58.

378. Madjar H, Sauerbrei W, Munch S, et al. Continuous-wave and pulsed Doppler studies of the breast: clinical results and effect of transducer frequency . Ultrasound Med Biol 1991;17(1):31-9.

379. Madjar H, Prompeler HJ, Del Favero C, et al. A new Doppler signal enhancing agent for flow assessment in breast lesions. Eur J Ultrasound 2000 Dec;12(2):123-30.

380. Marini C, Traino C, Cilotti A, et al. Differentiation of benign and malignant breast microcalcifications: mammography versus mammography-sonography combination. Radiol Med (Torino) 2003;105(1-2):17-26.

381. Martinez AM, Medina CJ, Bustos C, et al. Assessment of breast lesions using Doppler with contrast agents. Eur J Gynaecol Oncol 2003;24(6):527-30.

382. Medl M, Peters-Engl C, Leodolter S. The use of color-coded Doppler sonography in the diagnosis of breast cancer. Anticancer Res 1994;14(5B):2249-51.

383. Mesaki K, Hisa N, Kubota K, et al. Differentiation of benign and malignant breast tumors using Doppler spectral parameters including acceleration time index. Oncol Rep (Athens) 2003;10(4):945-50.

384. Milz P, Lienemann A, Kessler M, et al. Evaluation of breast lesions by power Doppler sonography. Eur Radiol 2001;11(4):547-54.

385. Moss HA, Britton PD, Flower CDR, et al. How reliable is modern breast imaging in differentiating benign from malignant breast lesions in the symptomatic population? Clin Radiol 1999;54(10):676-82.

386. Obwegeser R, Berghammer P, Kubista E. Ultrasound-imaging of lateral thoracic arteries to detect breast cancer. Lancet 1999 Jun 5;353(9168):1938.

387. Ohlinger R, Klein GM, Kohler G. Ultrasound of the breast - value of sonographic criteria for the differential diagnosis of solid lesions. Ultraschall Med 2004 Feb;25(1):48-53.

388. Ozdemir A, Kilic K, Ozdemir H, et al. Contrast-enhanced power doppler sonography in breast lesions: effect on differential diagnosis after mammography and gray scale sonography. J Ultrasound Med 2004;23(2):183-96.

389. Zdemir A, Kilic K, Ozdemir H, et al. Contrast-enhanced power Doppler sonography in breast lesions: effect on differential diagnosis after mammography and gray scale sonography. J Ultrasound Med 2004 Feb;23 (2):183-95; quiz 196-7.

390. Ozdemir A, Ozdemir H, Maral I, et al. Differential diagnosis of solid breast lesions: contribution of Doppler studies to mammography and gray scale imaging . J Ultrasound Med 2001 Oct;20(10):1091-101; quiz 1102.

391. Ozdemir A, Oznur II, Vural G, et al. Tl-201 scintigraphy, mammography and ultrasonography in the evaluation of palpable and nonpalpable breast lesions: a correlative study. Eur J Radiol 1997 Feb;24(2):145-54.

392. Pamilo M, Soiva M, Anttinen I, et al. Ultrasonography of breast lesions detected in mammography screening. Acta Radiol 1991 May;32(3):220-5.

393. Park JM, Yoon GS, Kim SM, et al. Sonographic detection of multifocality in breast carcinoma. J Clin Ultrasound 2003;31(6):293-8.

394. Peters-Engl C, Medl M, Mirau M, et al. Color-coded and spectral Doppler flow in breast carcinomas--relationship with the tumor microvasculature. Breast Cancer Res Treat 1998 Jan;47(1):83-9.

395. Peters-Engl C, Medl M, Leodolter S. The use of colour-coded and spectral Doppler ultrasound in the differentiation of benign and malignant breast lesions . Br J Cancer 1995 Jan;71(1):137-9.

396. Pillsbury SG Jr, Haugen JA, Roux S. Reliability of multimodal evaluation of abnormal screening mammogram results. Am J Obstet Gynecol 1996 Jun;174(6):1683-6; discussion 1686-7.

397. Pinero A, Reus M, Illana J, et al. Palpable breast lesions: utility of Doppler sonography for diagnosis of

malignancy. Breast 2003 Aug;12(4):258-63.

398. Pritt B, Ashikaga T, Oppenheimer RG, et al. Influence of breast cancer histology on the relationship between ultrasound and pathology tumor size measurements. Mod Pathol 2004 Aug;17(8):905-10.

399. Puglisi F, Zuiani C, Bazzocchi M, et al. Role of mammography, ultrasound and large core biopsy in the diagnostic evaluation of papillary breast lesions. Oncology 2003;65(4):311-5.

400. Ramlau C, Sledzikowski J. Combination of thermographic and ultrasound methods for the diagnosis of female breast cancer. Eur J Gynaecol Oncol 1993;14 Suppl:152-4.

401. Ranieri E, D'Andrea MR, D'Alessio A, et al. Ultrasound in the detection of breast cancer associated with isolated clustered microcalcifications, mammographically identified. Anticancer Res 1997;17(4A):2831-5.

402. Raza S, Baum JK. Solid breast lesions: evaluation with power Doppler US. Radiology 1997 Apr;203(1):164-8.

403. Reinikainen H, Rissanen T, Paivansalo M, et al. B-mode, power Doppler and contrast-enhanced power Doppler ultrasonography in the diagnosis of breast tumors. Acta Radiol 2001 Jan;42(1):106-13.

404. Richter K. Clinical amplitude/velocity reconstructive imaging (CARI)--a new sonographic method for detecting breast lesions. Br J Radiol 1995 Apr; 68(808): 375-84.

405. Richter K. Detection of diffuse breast cancers with a new sonographic method. J Clin Ultrasound 1996 May;24(4):157-68.

406. Richter K, Heywang-Kobrunner SH . Sonographic differentiation of benign from malignant breast lesions: value of indirect measurement of ultrasound velocity. AJR Am J Roentgenol 1995 Oct;165(4):825-31.

407. Richter K, Heywang-Kobrunner SH . Quantitative parameters measured by a new sonographic method for differentiation of benign and malignant breast disease. Invest Radiol 1995 Jul;30(7):401-11.

408. Richter K, Heywang-Kobrunner SH, Winzer KJ, et al. Detection of malignant and benign breast lesions with an automated US system: results in 120 cases. Radiology 1997 Dec;205(3):823-30.

409. Rosen EL, Soo MS. Tissue harmonic imaging sonography of breast lesions: improved margin analysis, conspicuity, and image quality compared to conventional ultrasound. Clin Imaging 2001;25(6):379-84.

410. Rubin E, Mennemeyer ST, Desmond RA, et al. Reducing the cost of diagnosis of breast carcinoma: impact of ultrasound and imaging-guided biopsies on a clinical breast practice (Structured abstract). Cancer 2001;91(2):324-32.

411. Saarenmaa I, Salminen T, Geiger U, et al. The effect of age and density of the breast on the sensitivity of breast cancer diagnostic by mammography and ultasonography. Breast Cancer Res Treat 2001 May;67(2):117-23.

412. Sahin-Akyar G, Sumer H. Color Doppler ultrasound and spectral analysis of tumor vessels in the differential diagnosis of solid breast masses. Invest Radiol 1996 Feb;31(2):72-9.

413. Saitoh R, Kojima R, Ito K, et al. Real-time high resolution ultrasonography of solid breast masses: use of a 10-MHz mechanical sector transducer with a water bag. Radiat Med 1994;12(5):201-8.

414. Schelling M, Gnirs J, Braun M, et al. Optimized differential diagnosis of breast lesions by combined B-mode and color Doppler sonography. Ultrasound Obstet Gynecol 1997 Jul;10(1):48-53.

415. Schroeder RJ, Maeurer J, Vogl TJ, et al. D-galactose-based signal-enhanced color Doppler sonography of breast tumors and tumorlike lesions. Invest Radiol 1999 Feb;34(2):109-15.

416. Schutze B, Marx C, Fleck M, et al. Diagnostic evaluation of sonographically visualized breast lesions by using a new clinical amplitude/velocity reference imaging technique (CARI sonography). Invest Radiol 1998 Jun;33(6):341-7.

417. Sehgal CM, Cary TW, Kangas SA, et al. Computer-based margin analysis of breast sonography for differentiating

malignant and benign masses. J Ultrasound Med 2004 Sep;23(9):1201-9.

418. Selinko VL, Middleton LP, Dempsey PJ. Role of sonography in diagnosing and staging invasive lobular carcinoma. J Clin Ultrasound 2004 Sep;32(7):323-32.

419. Seo BK, Oh YW, Kim HR, et al. Sonographic evaluation of breast nodules: comparison of conventional, real-time compound, and pulse-inversion harmonic images. Korean J Radiol 2002;3(1):38-44.

420. Shetty MK, Shah YP, Sharman RS. Prospective evaluation of the value of combined mammographic and sonographic assessment in patients with palpable abnormalities of the breast. J Ultrasound Med 2003 Mar;22(3):263-8.

421. Skaane P. The additional value of US to mammography in the diagnosis of breast cancer. A prospective study. Acta Radiol 1999 Sep;40(5):486-90.

422. Skaane P, Sauer T. Ultrasonography of malignant breast neoplasms. Analysis of carcinomas missed as tumor. Acta Radiol 1999 Jul;40(4):376-82.

423. Skaane P, Skjorten F. Ultrasonographic evaluation of invasive lobular carcinoma. Acta Radiol 1999 Jul;40(4):369-75.

424. Skaane P, Olsen JB, Sager EM, et al. Variability in the interpretation of ultrasonography in patients with palpable noncalcified breast tumors. Acta Radiol 1999 Mar;40(2):169-75.

425. Snelling JD, Abdullah N, Brown G, et al. Measurement of tumour size in case selection for breast cancer therapy by clinical assessment and ultrasound. Eur J Surg Oncol 2004 Feb;30(1):5-9.

426. Spreafico C, Lanocita R, Frigerio LF, et al. The Italian experience with SH U 508 A (Levovist) in breast disease. Radiol Med (Torino) 1994 May;87(5 Suppl 1):59-64.

427. Stavros AT, Thickman D, Rapp CL, et al. Solid breast nodules: use of sonography to distinguish between benign and malignant lesions. Radiology 1995 Jul;196(1):123-34.

428. Steinberg BD, Carlson DL, Birnbaum JA. Sonographic discrimination between benign and malignant breast lesions with use of disparity processing. Acad Radiol 2001 Aug;8(8):705-12.

429. Szopinski KT, Pajk AM, Wysocki M, et al. Tissue harmonic imaging: utility in breast sonography. J Ultrasound Med 2003 May;22(5):479-87.

430. Tavassoli K, Cavalla P, Porcelli A, et al. Ultrasound diagnostic criteria in breast disease. Panminerva Med 1997 Sep;39(3):178-82.

431. Taylor KJ, Merritt C, Piccoli C, et al. Ultrasound as a complement to mammography and breast examination to characterize breast masses. Ultrasound Med Biol 2002 Jan;28(1):19-26.

432. Tohnosu N, Okuyama K, Koide Y, et al. A comparison between ultrasonography and mammography, computed tomography and digital subtraction angiography for the detection of breast cancers. Surg Today 1993;23(8):704-10.

433. Vetto JT, Pommier RF, Schmidt WA, et al. Diagnosis of palpable breast lesions in younger women by the modified triple test is accurate and cost-effective (Structured abstract). Arch Surg 1996;131(9):967-72.

434. Wang HC, Chen DR, Kao CH, et al . Detecting breast cancer in mammographically dense breasts: comparing technetium-99m tetrofosmin mammoscintigraphy and ultrasonography. Cancer Invest 2002;20(7-8):932-8.

435. Whitehouse PA, Baber Y, Brown G, et al. The use of ultrasound by breast surgeons in outpatients: an accurate extension of clinical diagnosis. Eur J Surg Oncol 2001 Nov;27(7):611-6.

436. Whitehouse PA, Barber Y, Brown G, et al. Office ultrasound by breast surgeons: a rapid, accurate adjuct to clinical diagnosis. Breast Cancer Res Treat 2001;69(3):237.

437. Wilkens TH, Burke BJ, Cancelada DA, et al. Evaluation of palpable breast masses with color Doppler sonography and gray scale imaging. J Ultrasound Med 1998 Feb;17(2):109-15.

438. Wright IA, Pugh ND, Lyons K, et al. Power Doppler in breast tumours: a comparison with conventional colour Doppler imaging. Eur J Ultrasound 1998 Aug;7(3):175-81.

439. Yang WT, Metreweli C. Assessment of factors affecting palpability of breast masses using ultrasonography in Chinese women. J Ultrasound Med 1996 Dec;15(12):807-12.

440. Yang WT, Tse GM. Sonographic, mammographic, and histopathologic correlation of symptomatic ductal carcinoma in situ. AJR Am J Roentgenol 2004 Jan;182(1):101-10.

441. Yang WT, Suen M, Ahuja A, et al . In vivo demonstration of microcalcification in breast cancer using high resolution ultrasound. Br J Radiol 1997 Jul;70(835):685-90.

442. Yang WT, Metreweli C, Lam PK, et al. Benign and malignant breast masses and axillary nodes: evaluation with echo-enhanced color power Doppler US. Radiology 2001 Sep;220(3):795-802.

443. Yang WT, Mok CO, King W, et al. Role of high frequency ultrasonography in the evaluation of palpable breast masses in Chinese women: alternative to mammography. J Ultrasound Med 1996 Sep;15(9):637-44.

444. Youssefzadeh S, Eibenberger K, Helbich T, et al. Use of resistance index for the diagnosis of breast tumours. Clin Radiol 1996 Jun;51(6):418-20.

445. Zonderland HM, Coerkamp EG, Hermans J, et al. Diagnosis of breast cancer: contribution of US as an adjunct to mammography. Radiology 1999 Nov;213(2):413-22.

446. Bagni B, Franceschetto A, Casolo A, et al. Scintimammography with 99mTc-MIBI and magnetic resonance imaging in the evaluation of breast cancer. Eur J Nucl Med Mol Imaging 2003 Oct;30(10):1383-8.

447. Berg WA, Gutierrez L, NessAiver MS, et al. Diagnostic accuracy of mammography, clinical examination, US, and MR imaging in preoperative assessment of breast cancer. Radiology 2004 Dec;233(3):830-49.

448. Boetes C, Mus RD, Holland R, et al. Breast tumors: comparative accuracy of MR imaging relative to mammography and US for demonstrating extent. Radiology 1995 Dec;197(3):743-7.

449. Hardy JR, Powles TJ, Judson I, et al. How many tests are required in the diagnosis of palpable breast abnormalities. Clin Oncol (R Coll Radiol) 1990 May;2(3):148-52.

450. Hata T, Takahashi H, Watanabe K, et al. Magnetic resonance imaging for preoperative evaluation of breast cancer: a comparative study with mammography and ultrasonography. J Am Coll Surg 2004 Feb;198(2):190-7.

451. Hlawatsch A, Teifke A, Schmidt M, et al. Preoperative assessment of breast cancer: sonography versus MR imaging. AJR Am J Roentgenol 2002 Dec;179(6):1493-501.

452. Leinsinger GL, Friedl L, Tiling R, et al. Comparison of dynamic MR imaging of the breast and sestamibi scintimammography for evaluation of indeterminate mammographic lesions. Eur Radiol 2001;11(10):2050-7.

453. Liang W, Lawrence WF, Burnett CB, et al. Acceptability of diagnostic tests for breast cancer. Breast Cancer Res Treat 2003 May;79(2):199-206.

454. Malur S, Wurdinger S, Moritz A, et al. Comparison of written reports of mammography, sonography and magnetic resonance mammography for preoperative evaluation of breast lesions, with special emphasis on magnetic resonance mammography. Breast Cancer Res 2001;3(1):55-60.

455. Muller-Schimpfle M, Stoll P, Stern W, et al. Do mammography, sonography, and MR mammography have a diagnostic benefit compared with mammography and sonography? AJR Am J Roentgenol 1997 May;168(5):1323-9.

456. Rieber A, Schirrmeister H, Gabelmann A, et al. Pre-operative staging of invasive breast cancer with MR mammography and/or PET: boon or bunk. Br J Radiol 2002 Oct ;75(898):789-98.

457. Sommer H, Tiling R, Pechmann M, et al. Evaluation of mammographic breast lesions with Tc-99m sestamibi scintimammography and contrast enhanced MRI. Zentralbl Gynakol 1997;119(1):6-11.

458. Tiling R, Khalkhali I, Sommer H, et al. Limited value of scintimammography and contrast-enhanced MRI in the evaluation of microcalcification detected by mammography. Nucl Med Commun 1998 Jan;19(1):55-62.

459. Wang HC, Sun SS, Kao A, et al. Comparison of technetium-99m methoxyisobutylisonitrile scintimammography and ultrasonography in the diagnosis of breast cancer in patients with mammographically dense breast. Cancer Invest 2002;20(3):318-23.

460. Yang WT, Lam WW, Cheung H, et al. Sonographic, magnetic resonance imaging, and mammographic assessments of preoperative size of breast cancer. J Ultrasound Med 1997 Dec;16(12):791-7.

461. Bone B, Wiberg MK, Szabo BK, et al. Comparison of 99mTc-sestamibi scintimammography and dynamic MR imaging as adjuncts to mammography in the diagnosis of breast cancer. Acta Radiol 2003 Jan;44(1):28-34.

462. Maunda KY, Chande H, Mselle TF, et al. 99mTc sestamibi scintimammography in the diagnosis of palpable breast masses. Nucl Med Commun 2003 Feb;24(2):141-4.

463. Brem RF, Schoonjans JM, Kieper DA, et al. High-resolution scintimammography: a pilot study. J Nucl Med 2002 Jul;43(7):909-15.

464. Danielsson R, Bone B, Agren B, et al. Comparison of planar and SPECT scintimammography with 99mTc-sestamibi in the diagnosis of breast carcinoma. Acta Radiol 1999 Mar;40(2):176-80.

465. De Vincentis G, Gianni W, Pani R, et al. Role of 99 mTc-Sestamibi scintimammography by SPEM camera in the management of breast cancer in the elderly. Breast Cancer Res Treat 1998;48(2):159-63.

466. Uriarte I, Carril JM, Quirce R, et al. Optimization of X-ray mammography and technetium-99m methoxyisobutylisonitrile scintimammography in the diagnosis of non-palpable breast lesions. Eur J Nucl Med 1998 May;25(5):491-6.

467. Alonso JC, Soriano A, Zarca MA, et al. Breast cancer detection with sestamibi-Tc-99m and Tl-201 radionuclides in patients with non conclusive mammography. Anticancer Res 1997;17(3B):1661-5.

468. Scopinaro F, Ierardi M, Porfiri LM, et al. 99mTc-MIBI prone scintimammography in patients with high and intermediate risk mammography. Anticancer Res 1997;17(3B):1635-8.

469. Tiling R, Khalkhali I, Sommer H, et al. Role of technetium-99m sestamibi scintimammography and contrast-enhanced magnetic resonance imaging for the evaluation of indeterminate mammograms. Eur J Nucl Med 1997 Oct;24(10):1221-9.

470. Kristoffersen Wiberg M, Aspelin P, Perbeck L, et al. Value of MR imaging in clinical evaluation of breast lesions. Acta Radiol 2002 May;43(3):275-81.

U.S. Department of Health and Human Services
Mike Leavitt, *Secretary*

Office of Public Health and Science
Richard H. Carmona, M.D., M.P.H., F.A.C.S., *Surgeon General of the United States*

Agency for Healthcare Research and Quality
Carolyn Clancy, M.D., *Director*

U.S. Department of
Health and Human Services

Public Health Service
Agency for Healthcare Research and Quality

AHRQ Publication No. 06-EHC005-EF
February 2006