Effectiveness of Non-invasive Diagnostic Tests for Breast Abnormalities – An Update to the 2006 Report

This report is based on research conducted by the XXX (EPC) under contract to the Agency for Healthcare Research and Quality (AHRQ), Rockville, MD (Contract No. XXX). The findings and conclusions in this document are those of the author(s), who are responsible for its content, and do not necessarily represent the views of AHRQ. No statement in this report should be construed as an official position of AHRQ or of the U.S. Department of Health and Human Services.

The information in this report is intended to help clinicians, employers, policymakers, and others make informed decisions about the provision of health care services. This report is intended as a reference and not as a substitute for clinical judgment.

This report may be used, in whole or in part, as the basis for the development of clinical practice guidelines and other quality enhancement tools, or as a basis for reimbursement and coverage policies. AHRQ or U.S. Department of Health and Human Services endorsement of such derivative products may not be stated or implied.

Draft Comparative Effectiveness Review
Number XX

Effectiveness of Non-invasive Diagnostic Tests for Breast Abnormalities – An Update to the 2006 Report

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.

Prepared by:

Investigators

AHRQ Publication No.

This document is in the public domain and may be used and reprinted without permission except those copyrighted materials noted, for which further reproduction is prohibited without the specific permission of copyright holders.

Suggested Citation:

No investigators have any affiliations or financial involvement (e.g., employment, consultancies, honoraria, stock options, expert testimony, grants or patents received or pending, or royalties) that conflict with material presented in this report.

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 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 (CERs) 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 http://effectivehealthcare.ahrq.gov/reference/purpose.cfm

AHRQ expects that CERs will be helpful to health plans, providers, purchasers, government programs, and the health care system as a whole. In addition, AHRQ is committed to presenting information in different formats so that consumers who make decisions about their own and their family's health can benefit from the evidence.

Transparency and stakeholder input are essential to the Effective Health Care Program. 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. Comparative Effectiveness Reviews will be updated regularly.

Carolyn M. Clancy, M.D.
Director
Agency for Healthcare Research and Quality

Jean Slutsky, P.A., M.S.P.H.
Director, Center for Outcomes and Evidence
Agency for Healthcare Research and Quality

Stephanie Chang, M.D., M.P.H.
Director, Evidence-based Practice Center Program
Agency for Healthcare Research and Quality

William Lawrence, M.D., M.S.
Task Order Officer, Evidence-based Practice Center Program
Agency for Healthcare Research and Quality

Acknowledgments

Technical Expert Panel

AHRQ Contacts

Structured Abstract

Objectives: To systematically review the literature on the diagnostic accuracy of non-invasive imaging technologies that have been proposed to be useful as part of the work-up after recall of women with suspicious breast abnormalities identified on routine screening. This report is an update of a Comparative Effectiveness Review (CER) of the same title originally published in 2006.

Data Sources: The medical literature, including PubMed and EMBASE, was searched from December 1994 through September 2010. Diagnostic cohort studies that enrolled the patient population of interest, and used current generation scanners and protocols of the non-invasive imaging technologies of interest, were selected. Case-control studies were not included, nor were meeting presentations or very small (<10 patients) studies.

Review Methods: Data was abstracted from the included studies and meta-analyzed using a bivariate mixed-effects binomial regression model. The summary likelihood ratios and Bayes' theorem were used to calculate the post-test probability of having a benign or malignant lesion. Heterogeneity in the data was explored with meta-regressions using standard methodology. The strength of evidence supporting each major conclusion was graded as High, Moderate, Low, or Insufficient. The grade was developed by considering four important domains: the risk of bias in the evidence base (internal validity, or quality of the studies), the consistency of the findings, the precision of the results, and the directness of the evidence.

Results: We identified 41 studies of MRI. The summary sensitivity of MRI was 91.7% (95% CI: 88.5 to 94.1%) and the summary specificity was 77.5% (95% CI: 71.0 to 82.9%). The estimate of accuracy was judged to be supported by a Moderate to Low strength of evidence (Low for the estimate of specificity due to the wide confidence interval).

We identified seven studies of PET. The summary sensitivity of PET was 83.0% (95% CI: 73.0 to 89.0%) and the summary specificity was 74.0% (95% CI: 58.0 to 86.0%). The estimate of accuracy was judged to be supported by a Low strength of evidence.

We identified 10 studies of scintimammography. The summary sensitivity of scintimammography was 84.7% (95% CI: 78.0 to 89.7%) and the summary specificity was 77.0% (95% CI: 64.7 to 85.9%). The estimate of accuracy was judged to be supported by a Low strength of evidence.

We identified 21 studies of B-mode grayscale ultrasound, six studies of color Doppler ultrasound, and seven studies of power Doppler ultrasound. For B-mode grayscale, summary sensitivity was 92.4% (95% CI: 84.6 to 96.4%) and the summary specificity was 75.8% (95% CI: 60.8 to 86.3%); for color Doppler, summary sensitivity was 88.5% (95% CI: 74.4 to 95.4%) and summary specificity was 76.4% (95% CI: 61.7 to 86.7%); for power Doppler, summary sensitivity was 70.8% (95% CI: 47 to 86.6%) and summary specificity was 72.6% (95% CI: 59.9 to 82.5%). These estimates of accuracy were all judged to be supported by a Low strength of evidence.

Conclusions: The use of non-invasive imaging in addition to standard workup may be clinically useful for women with a low (2 to 20%) suspicion of malignancy. When choosing which non-invasive imaging technology to use for this purpose, diagnostic B-mode grayscale ultrasound and MRI appear to more accurate than PET, scintimammography, or Doppler ultrasound.

Contents

Executive Summary 1
Background 1
Methods 2
Results 3
Discussion 5
Remaining Issues 6
Conclusions 7

Evidence Report 9

Chapter 1. Introduction 11

Background 11
Breast Cancer 11
Breast Cancer Diagnosis 11
Non-invasive Imaging 13
Conceptual Framework 13
Diagnostic Test Characteristics 16

Scope and Key Questions 18

Chapter 2. Methods 19

Topic Development 19
Search Strategy 20
Study Selection 20
Data Abstraction 24
Study Quality Evaluation 24
Strength of Evidence Assessment 24
Applicability 26
Data Analysis and Synthesis 26
Peer Review and Public Commentary 27

Chapter 3. Results 29

Magnetic Resonance Imaging (MRI) 30
Background 30
Findings from 2006 Review 32

Evidence Base 32
Question 1. What is the accuracy of MRI for diagnosis of breast cancer in women referred for further evaluation after identification of a possible breast abnormality on routine screening (mammography and/or clinical or self-detection of a palpable lesion)? 32
Question 2. Are there demographic (e.g., age) and clinical risk factors (e.g., morphologic characteristics of the lesion) that affect the accuracy of the tests considered in Question 1? ... 33
Question 3. Are there other factors and considerations that may affect the accuracy or acceptability of MRI? 33
Previously Published Systematic Reviews 34
Conclusion 36

Positron Emission Tomography (PET) 41
Background 41
Findings from 2006 Review 42
Evidence Base 43
Question 1. What is the accuracy of PET for diagnosis of breast cancer in women referred for further evaluation after identification of a possible breast abnormality on routine screening (mammography and/or clinical or self-detection of a palpable lesion)? 43
Question 2. Are there demographic (e.g., age) and clinical risk factors (e.g., morphologic characteristics of the lesion) that affect the accuracy of the tests considered in Question 1? ... 44
Question 3. Are there other factors and considerations that may affect the accuracy or acceptability of the tests considered in Questions 1 and 2? 44
Previously Published Systematic Reviews 44
Conclusion 45

Scintimammography 48
Background 48
Findings from 2006 Review 49
Evidence Base 49
Question 1. What is the accuracy of scintimammography for diagnosis of breast cancer in women referred for further evaluation after identification of a possible breast abnormality on routine screening (mammography and/or clinical or self-detection of a palpable lesion)? .. 49
Question 2. Are there demographic (e.g., age) and clinical risk factors (e.g., morphologic characteristics of the lesion) that affect the accuracy of the tests considered in Question 1? ... 50
Question 3. Are there other factors and considerations that may affect the accuracy or acceptability of the tests considered in Questions 1 and 2? 50
Previously Published Systematic Reviews 50
Conclusion 52

Ultrasound....56
Background....56
Findings from 2006 Review....57
Evidence Base....57
Question 1. What is the accuracy of ultrasound for diagnosis of breast cancer in women referred for further evaluation after identification of a possible breast abnormality on routine screening (mammography and/or clinical or self-detection of a palpable lesion)?....58
Question 2. Are there demographic (e.g., age) and clinical risk factors (e.g., morphologic characteristics of the lesion) that affect the accuracy of the tests considered in Question 1?....59
Question 3. Are there other factors and considerations that may affect the accuracy or acceptability of the tests considered in Questions 1 and 2?....59
Previously Published Systematic Reviews....60
Conclusion....60
Comparative Accuracy....65
Chapter 4. Discussion....67
Changes since 2006....68
Limitations of the Evidence Base....68
Applicability....68
Possible Impact of Key Assumptions on the Conclusions....69
Future Research Needed....70
References and Included Studies....71
List of Acronyms/Abbreviations....85
Glossary of Selected Terms....87
Definition of Terms....87

Tables

Table 1. Summary of key findings 7
Table 2. Example of a 2x2 table 16
Table 3. Non-invasive current technologies to be evaluated 22
Table 4. Other published technology assessments of MRI 34
Table 5. Included studies: magnetic resonance imaging (MRI) 36
Table 6. Magnetic resonance imaging (MRI) accuracy 39
Table 7. Clinical interpretations of magnetic resonance accuracy: benign finding on MRI 40
Table 8. Clinical interpretations of MRI accuracy: malignant finding on MRI 40
Table 9. Included studies: PET and PET/CT 46
Table 10. PET accuracy 46
Table 11. Clinical interpretations of PET accuracy: benign finding on PET 47
Table 12. Clinical interpretations of PET accuracy: malignant finding on PET 47
Table 13. Other published technology assessments of scintimammography 50
Table 14. Included studies: scintimammography 53
Table 15. Scintimammography accuracy 54
Table 16. Clinical interpretations of scintimammography accuracy: benign finding on scintimammography 55
Table 17. Clinical interpretations of scintimammography accuracy: malignant finding on scintimammography 55
Table 18. Included studies: ultrasound 61
Table 19. Ultrasound accuracy: accuracy of different types of ultrasound 63
Table 20. Clinical interpretations of ultrasound accuracy: benign finding on ultrasound 64
Table 21. Clinical interpretations of ultrasound accuracy: malignant finding on ultrasound 64
Table 22. Summary accuracy results 66

Figures

Figure 1. Analytical framework 15
Figure 2. Study selection process 29

Appendices

Appendix A: Search strategy and exact search strings
Appendix B. Sample Data Abstraction Forms
Appendix C: Evidence tables
Appendix D: List of Excluded Studies
Appendix E: Peer Reviewers

Appendices and Evidence Tables for this report are provided electronically at http://www.ahrq.gov/downloads/pub/evidence/pdf/XXX/XXX.pdf**.**

Effectiveness of Noninvasive Diagnostic Tests for Breast Abnormalities – An Update to the 2006 Report

Executive Summary

The Effective Health Care Program was initiated in 2005 to provide valid evidence about the comparative effectiveness of different medical interventions. The object is to help consumers, health care providers, and others in making informed choices among treatment alternatives. Through its Comparative Effectiveness Reviews, the program supports systematic appraisals of existing scientific evidence regarding treatments for high-priority health conditions. It also promotes and generates new scientific evidence by identifying gaps in existing scientific evidence and supporting new research. The program puts special emphasis on translating findings into a variety of useful formats for different stakeholders, including consumers.

The full report and this summary are available at www.effectivehealthcare.ahrq.gov/reports/final.cfm

Background

Breast cancer is one of the most common malignancies of women, with approximately 62,000 new cases diagnosed every year in the U.S. Some breast cancers are identified by physical examination (either self-examination or performed by a physician). Population-wide screening programs that use x-ray mammography to examine asymptomatic women are also in common use, intended to detect early signs of breast cancer. If a suspicious area is seen on x-ray mammography, women are usually recalled for further examination. The results of these examinations are used to make decisions about further management: return to normal screening/ return for short-interval follow-up/refer for biopsy. In current standard practice the examinations conducted after recall usually consist of diagnostic mammography and possibly ultrasound (to identify cysts).

It is important to triage recalled women into the correct management pathway. Women with readily treatable early-stage cancers who get mistakenly triaged into “return to normal screening” may experience a significant delay in diagnosis and treatment of the cancer. However, the majority of women who are recalled for further assessment after a screening mammography do not have cancer, and significant numbers of healthy women are referred for biopsy and short-interval follow-up after recall and diagnostic mammography.

A number of non-invasive imaging technologies have been developed and proposed to be useful as part of the work-up after recall. This evidence review focuses on additional non-invasive imaging studies that can be conducted after discovery of a possible abnormality on screening mammography or physical examination, studies intended to guide patient management decisions. In other words, these studies are not intended to provide a final diagnosis as to the nature of the breast lesion; rather, they are intended to provide additional information about the nature of the lesions such that women can be more appropriately triaged into the correct management pathway.

In order for women to benefit from the information provided by imaging tests, the accuracy of the tests needs to be established. Inaccurate tests will lead to sub-optimal management decisions. This report will therefore examine the diagnostic accuracy of the tests. The accuracy of the non-invasive imaging tests will be measured in terms of sensitivity and specificity. Sensitivity is a measure of how accurately the test can identify women with cancer; specificity is a measure of how accurately the test can identify women who do not have cancer. A test with high sensitivity will rarely misclassify women with cancer as not having cancer, and a test with high specificity will rarely misclassify women without cancer as having cancer.

This report is an update of a Comparative Effectiveness Review (CER) of the same title originally published in 2006. In addition to updating the literature, the Key Questions have been revised and additional non-invasive imaging tests have been added.

Methods

Topic development and scope. The topic was selected for update by the Effective Health Care program. The Key Questions were posted for public comment. A Technical Expert Panel (TEP) was assembled to provide guidance, and a protocol for updating the review was developed and approved by AHRQ.

Patient population. The patient population of interest is the general population of women participating in routine breast cancer screening programs (including mammography, clinical examination, and self-examination) who have subsequently been referred for further evaluation of a possible breast abnormality. Populations that will not be evaluated include: women thought to be at very high risk of breast cancer due to family history or BRCA mutations; women with a personal history of breast cancer; and men.

Interventions. The non-invasive diagnostic tests to be evaluated are ultrasound (conventional B-mode grayscale, harmonic, tomography, color Doppler, and power Doppler); magnetic resonance imaging (MRI, with gadolinium-based contrast agents) with or without computer-aided diagnosis (CADx); positron emission tomography (PET, with 18-fluorodeoxyglucose [FDG]) with or without concurrent computed tomography (CT) scans; scintimammography (with technetium-99m sestamibi [MIBI]), including Breast Specific Gamma Imaging (BSGI).

Comparators. The accuracy of the non-invasive diagnostic tests will be evaluated by a direct comparison to histopathology (surgical or biopsy specimens) or to clinical follow-up, or a combination of these methods. In addition, the relative accuracy of the different tests under evaluation will be directly and indirectly compared as the evidence permits.

Outcomes. Outcomes of interest are diagnostic test characteristics; namely, sensitivity, specificity, and likelihood ratios. Because predictive values vary as the prevalence of disease changes, we did not calculate predictive values. Adverse events related to the procedures, such as discomfort and reactions to contrast agents, will also be discussed as the evidence permits.

Timing. Any duration of follow-up, from same day interventions to many years of clinical follow-up, will be evaluated.

Setting. Any care setting will be evaluated, including general hospitals, physician's offices, and specialized breast imaging centers.

Study selection. The medical literature, including PubMed and EMBASE, was searched from December 1994 through September 2010. Diagnostic cohort studies that enrolled the patient population of interest, and used current generation scanners and protocols of the non-invasive imaging technologies of interest, were selected. Case-control studies were not included, nor were meeting presentations or very small (<10 patients) studies. Data was abstracted from the included studies.

Strength of evidence. The strength of evidence supporting each major conclusion was graded as High, Moderate, Low, or Insufficient. The grade was developed by considering four important domains: the risk of bias in the evidence base (internal validity, or quality of the studies), the consistency of the findings, the precision of the results, and the directness of the evidence.

Data analysis. The data were meta-analyzed using a bivariate mixed-effects binomial regression model. The summary likelihood ratios and Bayes' theorem were used to calculate the post-test probability of having a benign or malignant lesion. In cases where a bivariate binomial model could not be fit, we meta-analyzed the data using two random-effects models, one for sensitivity and one for specificity. Heterogeneity in the data was explored with meta-regressions using standard methodology.

Peer review and public comment. Concurrent with Peer Review, a draft of the report will be posted for public comment.

Results

Magnetic resonance imaging (MRI). We identified 41 studies of MRI that studied a total of 3,882 patients with 4,202 suspicious breast lesions. We combined the data reported by all 41 studies into a bivariate binomial mixed-effects model. The summary sensitivity was 91.7% (95% CI: 88.5 to 94.1%) and the summary specificity was 77.5% (95% CI: 71.0 to 82.9%). The estimate of accuracy was judged to be supported by a Moderate to Low strength of evidence (Low for the estimate of specificity due to the wide confidence interval). The dataset was very heterogeneous (I^2 = 98.4%). We explored the heterogeneity with meta-regression and found that the prevalence of disease in the study population and whether or not the image readers were blinded was statistically significantly correlated with the results. Subgroup analyses found that MRI was less sensitive (84.0% vs. 91.7%) for evaluation of microcalcifications.

The probability that a woman actually does have cancer even after a finding of "benign" on MRI depends on her probability of having cancer before undergoing the test. Bayes' theorem and the summary likelihood ratios indicate that if a woman with an estimated 5 to 10% chance of having cancer undergoes MRI and has a finding of "benign" she will then have an estimated 1% chance of having cancer; a woman with an estimated 20% chance of having cancer who has a finding of "benign" on MRI will then have an estimated 3% chance of having cancer; and a woman with an estimated 50% chance of having cancer who has a finding of "benign" on MRI will then have an estimated 10% chance of having cancer.

Positron emission tomography (PET). We identified seven studies of PET and one study of PET/CT that met our inclusion criteria. The studies of stand-alone PET included 308 women with 403 suspicious breast lesions. We combined the data reported by the seven studies of PET into a bivariate binomial mixed-effects model. The summary sensitivity was 83.0% (95% CI: 73.0 to 89.0%) and the summary specificity was 74.0% (95% CI: 58.0 to 86.0%). The estimate of

accuracy was judged to be supported by a Low strength of evidence. The dataset contained moderate heterogeneity (I^2 = 64.0%). We explored the heterogeneity with meta-regression and did not identify any possible causes. Subgroup analyses found that PET was more sensitive for evaluation of palpable lesions.

The probability that a woman actually does have cancer even after a finding of "benign" on PET depends on her probability of having cancer before undergoing the test. Bayes' theorem and the summary likelihood ratios indicate that if a woman with an estimated 5% chance of having cancer undergoes PET and has a finding of "benign" she will then have an estimated 1% chance of having cancer; a woman with an estimated 20% chance of having cancer who has a finding of "benign" on PET will then have an estimated 6% chance of having cancer; and a woman with an estimated 50% chance of having cancer who has a finding of "benign" on PET will then have an estimated 19% chance of having cancer.

Scintimammography. We identified 10 studies of scintimammography and one study of BSGI that met our inclusion criteria. The studies included a total of 1,064 suspicious lesions. We combined the data reported by all 11 studies into a bivariate binomial mixed-effects model. The summary sensitivity was 84.7% (95% CI: 78.0 to 89.7%) and the summary specificity was 77.0% (95% CI: 64.7 to 85.9%). The estimate of accuracy was judged to be supported by a Low strength of evidence. The dataset was very heterogeneous (I^2 = 93.0%). We explored the heterogeneity with meta-regression and did not identify any possible causes.

The probability that a woman actually does have cancer even after a finding of "benign" on scintimammography depends on her probability of having cancer before undergoing the test. Bayes' theorem and the summary likelihood ratios indicate that if a woman with an estimated 5% chance of having cancer undergoes scintimammography and has a finding of "benign" she will then have an estimated 1% chance of having cancer; a woman with an estimated 20% chance of having cancer who has a finding of "benign" on scintimammography will then have an estimated 5% chance of having cancer; and a woman with an estimated 50% chance of having cancer who has a finding of "benign" on scintimammography will then have an estimated 17% chance of having cancer.

Ultrasound. We identified a total of 31 diagnostic cohort studies of ultrasound. Of these, there were 21 studies of B-mode grayscale ultrasound, six studies of color Doppler ultrasound, and seven studies of power Doppler ultrasound. We combined the data reported by these studies into bivariate binomial mixed-effects models. For B-mode grayscale, summary sensitivity was 92.4% (95% CI: 84.6 to 96.4%) and the summary specificity was 75.8% (95% CI: 60.8 to 86.3%); for color Doppler, summary sensitivity was 88.5% (95% CI: 74.4 to 95.4%) and summary specificity was 76.4% (95% CI: 61.7 to 86.7%); for power Doppler, summary sensitivity was 70.8% (95% CI: 47 to 86.6%) and summary specificity was 72.6% (95% CI: 59.9 to 82.5%). These estimates of accuracy were all judged to be supported by a Low strength of evidence. The datasets were heterogeneous. We explored the heterogeneity of the largest dataset (21 studies of B-mode) with meta-regression and found that whether the studies blinded the image readers and accounted for inter-reader differences were statistically significantly associated with the results.

The probability that a woman actually does have cancer even after a finding of "benign" on ultrasound depends on her probability of having cancer before undergoing the test. Bayes' theorem and the summary likelihood ratios indicate that if a woman with an estimated 5 to 10% chance of having cancer undergoes B-mode grayscale ultrasound and has a finding of "benign"

she will then have an estimated 1% chance of having cancer; a woman with an estimated 20% chance of having cancer who has a finding of "benign" on B-mode grayscale ultrasound will then have an estimated 2% chance of having cancer; and a woman with an estimated 50% chance of having cancer who has a finding of "benign" on B-mode grayscale ultrasound will then have an estimated 9% chance of having cancer.

Discussion

According to the American College of Radiology, the threshold of suspicion of malignancy at which management of women changes is 2%. After recall and workup, women with a suspicion of malignancy greater than 2% are generally recommended to undergo tissue sampling of some kind (biopsy), and women with a lower suspicion of malignancy are triaged into imaging management pathways. We used the 2% threshold to explore the clinical usefulness of the various non-invasive imaging technologies as add-ons to the current standard of care; namely, if a woman was recalled for evaluation after a screening mammography, and received standard of care workup vs. standard of care workup plus the non-invasive imaging technology, would use of the non-invasive imaging technology be likely to alter the recommendations for care after the workup?

For all of the technologies evaluated in this assessment, only women with a very low suspicion of malignancy on standard of care workup might be expected to experience a change in management decisions as a result of additional non-invasive imaging. A woman with an approximate 2 to 20% suspicion of malignancy who has benign findings on one of the non-invasive imaging technologies could have her suspicion of malignancy drop below the 2% threshold, and therefore she might be assigned to imaging management rather than tissue sampling management.

Therefore, if the 2% threshold is chosen, the use of non-invasive imaging in addition to standard workup may be clinically useful for women with a low (2 to 20%) suspicion of malignancy. When choosing which non-invasive imaging technology to use for this purpose, diagnostic B-mode grayscale ultrasound and MRI appear to be more accurate than PET, scintimammography, or the other types of ultrasound (e.g., Doppler) that were evaluated in this comparative effectiveness review.

Changes since 2006. This CER is an update of a CER finalized in 2006. The updated results are, in general, very similar to the findings of the 2006 report. For MRI, in 2006 we found that the sensitivity was 92.5% and the specificity was 75.5%; the updated evidence base supported estimates of 91.7% sensitivity and 77.5% specificity. In both reports, MRI was found to be less sensitive (approximately 85%) for evaluation of microcalcifications than for evaluation of lesions in general. For PET, in 2006 we found that the sensitivity was 82.2% and the specificity was 78.3%; the updated evidence base supported estimates of 83.0% sensitivity and 74.0% specificity. In the updated report we attempted to evaluate the accuracy of PET/CT, but only one study that met the inclusion criteria was identified.

For scintimammography, the updated evidence base identified a sensitivity of 84.7%, much higher than the sensitivity estimate from 2006 of 68.7%. Specificity was estimated at 84.8% in 2006, and at 77.0% in the update; however, the confidence intervals around the updated estimate of specificity are wide. It is possible that improvements in the technology in the last few years improved the sensitivity of the technique.

For ultrasound, in 2006 we only evaluated a relatively small set of studies of B-mode grayscale ultrasound, and estimated a sensitivity of 86.1% and a specificity of 66.4%. The update included a significantly expanded evidence base on B-mode grayscale ultrasound, and identified a sensitivity of 92.4% and specificity of 75.8%. In the update we included numerous other types of ultrasound, including power and color Doppler ultrasound, that were not studied in the 2006 report.

Remaining Issues

Further studies on the diagnostic accuracy of the non-invasive technologies evaluated are unlikely to cause a substantive change in the nature of the conclusions. The conclusions were for the most part rated as being supported by Low strength of evidence due to the imprecision of the estimates (wide confidence intervals around the estimates of accuracy); the publication of additional diagnostic accuracy studies are likely to increase the precision of the estimates of accuracy. There was also considerable heterogeneity (inconsistency) in the majority of the evidence bases, which contributed to the Low strength of evidence rating. Most likely the heterogeneity was due to slight differences in imaging methodology or patient populations across studies; future research intended to tease out factors affecting the accuracy of imaging may be helpful to the clinician.

Future research efforts should focus on studies that report the impact of the use of non-invasive imaging on patient-oriented outcomes such as quality of life. The diagnostic thresholds that trigger invasive, aggressive diagnostic testing and treatment should also be studied in the context of the addition of non-invasive imaging to standard protocols. Current standard of care results in large numbers of healthy women undergoing invasive diagnostic procedures, and many women may be undergoing unnecessarily aggressive treatment for small early-stage breast cancers. The diagnostic thresholds in current use are intended to reduce the rate of missed cancers to its absolute minimum, which by necessity causes a severe loss of specificity; and are also intended to partially compensate for diagnostic inaccuracy of tests in current use. The addition of non-invasive imaging to standard protocols may allow for the use of different diagnostic thresholds. It is possible that women at a low risk of malignancy may benefit tremendously from monitoring with non-invasive imaging rather than going straight to biopsy. The thresholds used in clinical practice to trigger implementation of aggressive diagnostic testing and treatment should be based on solid evidence derived from controlled trials and modelling studies.

Conclusions

The use of non-invasive imaging in addition to standard workup may be clinically useful for women with a low (2 to 20%) suspicion of malignancy. When choosing which non-invasive imaging technology to use for this purpose, diagnostic B-mode grayscale ultrasound and MRI appear to more accurate than PET, scintimammography, or Doppler ultrasound.

Table 1. Summary of key findings

Technology	Summary Sensitivity	Summary Specificity	Strength of Evidence
B-mode grayscale 2D ultrasound	92.4% (84.6 to 96.4%)	75.8% (60.8 to 86.3%)	Low
MRI	91.7% (88.5 to 94.1%)	77.5% (71.0 to 82.9%)	Moderate (sensitivity) to Low (specificity)
Scintimammography	84.7% (78.0 to 89.7%)	77.0% (64.7 to 85.9%)	Low
PET	83.0% (73.0 to 89.0%)	74.0% (58.0 to 86%)	Low

Evidence Report

Chapter 1. Introduction

Background

Breast Cancer

Breast cancer is the second most common malignancy of women.[1] The American Cancer Society estimates that in the U.S. in 2009, 62,280 women were diagnosed with new cases of *in situ* cancer, 192,730 women were newly diagnosed as having invasive breast cancer, and there were 40,480 deaths due to this disease.[2] In the general population, the cumulative risk of being diagnosed with breast cancer by age 70 is estimated to be 6% (lifetime risk of 13%).[3,4]

The most common type of breast cancer, accounting for over 85% of cases diagnosed, is ductal carcinoma.[5] Ductal carcinoma arises within the ducts of the breast from the cells lining the ducts. Early stage breast cancer confined to the inside of the duct is referred to as ductal carcinoma *in situ* (DCIS). Later stages of ductal carcinoma that have invaded or broken through the walls of the ducts into nearby tissues may be referred to as invasive or infiltrating ductal carcinoma. Cases of invasive ductal carcinoma that are found to be well-differentiated specific subtypes (such as mucinous, medullary, tubular, or papillary) are much rarer than the common "otherwise not specified" type of invasive ductal carcinoma.

Another type of invasive carcinoma is lobular carcinoma. Lobular carcinoma is similar to ductal carcinoma, first arising in the terminal ducts of the lobules and then invading through the walls of the ducts and invading nearby tissues. Other rare types of potentially life-threatening breast tumors include papillary carcinoma, inflammatory breast cancer, and sarcomas, among others.[5]

A number of different breast lesions have been described that, while not malignant, are believed to predispose to the development of invasive breast carcinomas. These lesions include atypical ductal hyperplasia (ADH), papillary lesions, radial scars, atypical lobular hyperplasia (ALH), and lobular carcinoma in situ (LCIS).[6] However, the most commonly reported breast abnormalities diagnosed after screening are benign: benign fibrocystic changes, cysts, and benign fibroadenomas.

Breast Cancer Diagnosis

Breast cancer is usually first detected by feeling a lump on physical examination (either self-examination or an exam conducted by a health practitioner) or by observing an abnormality during 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 close to 100%. The five-year survival rate for women with stage IV (cancer that has spread beyond the breast) is only 27%.[7] Because early breast cancer is asymptomatic, the only way to detect it is through screening of asymptomatic women. Mammography is a widely accepted and used method for breast cancer screening.[8-10] Meta-analyses of large clinical trials have demonstrated that mammography screening reduces breast cancer mortality.[11,12]

Mammography uses x-rays to examine the breast for clusters of microcalcifications, circumscribed and dense masses, masses with indistinct margins, architectural distortion compared with the contralateral breast, or other abnormal structures. The United States Preventive Services Task Force (USPSTF) has recently recommended routine screening

mammography every two years for women aged 50 to 74, with decisions to screen women under the age of 50 made on an individual basis.[10] After identification of a possible abnormality on screening mammography or physical examination, women typically undergo additional imaging studies (diagnostic mammography and/or ultrasound) and a physical examination. If these studies suggest the abnormality may be malignant, a biopsy of the suspicious area may be recommended.

The American College of Radiology has created a standardized system for reporting the results of mammography, the Breast Imaging Reporting and Data System (BI-RADS®).[13-15] There are seven categories of assessment, each with an accompanying clinical management recommendation:

0 Need additional imaging evaluation and/or prior mammograms for comparison

1 Negative

2 Benign finding

3 Probably benign finding. Initial short interval follow-up suggested.

4 Suspicious abnormality. Biopsy should be considered.

5 Highly suggestive of malignancy. Appropriate action should be taken.

6 Known biopsy-proven malignancy. Appropriate action should be taken.

Non-invasive breast imaging tests have multiple uses, including image-guidance of biopsy procedures, searching for multifocal lesions in a woman diagnosed with or at high risk of breast cancer, and screening women at high risk of breast cancer. This evidence review specifically focuses only on the use of non-invasive imaging studies that can be conducted after the discovery of a possible abnormality on screening mammography or physical examination-studies intended to guide patient management decisions. In other words, these studies are not intended to provide a final diagnosis as to the nature of the breast lesion; rather, they are intended to provide additional information about the nature of the lesion such that women can be appropriately triaged into "biopsy/watchful waiting/return to normal screening intervals" care pathways.

It is important to accurately triage women into the correct care pathway. Women with readily treatable breast cancers who get incorrectly triaged into "return to normal screening care pathways" may experience a significant delay in diagnosis and treatment of the cancer. However, the majority of women who are recalled for further assessment after a screening mammogram do not have cancer. Elmore et al. estimated that the cumulative risk for a woman having a false-positive finding on screening mammography is close to 50% after ten years of yearly screenings.[16] In addition, diagnostic mammography performed after a mammographic screening recall often leads to identification of a "probably benign" lesion. Women with "probably benign" lesions are usually referred for short-interval repeat mammography examinations. If an available non-invasive diagnostic test could assist clinicians in evaluating women recalled for further investigation after mammographic screening, namely, in assisting in accurately distinguishing between "benign," "probably benign," and "probably not benign" lesions, then many women could avoid frequent repeat mammography exams, with their attendant discomfort, inconvenience, x-ray exposure, and emotional distress.

The majority of women who traditionally have been referred for biopsy also do not have cancer. For example, Lacquement et al. examined a series of 668 women who underwent biopsy, and reported that only 23% of these women were diagnosed with breast cancer after biopsy.[17] Exposing large numbers of women who do not have cancer to invasive procedures may be considered an undesirable medical practice. In conclusion, many women could benefit from highly accurate non-invasive tests that can assist in guiding decisions about appropriate care after discovery of a possible breast abnormality.

The ultimate goal of this comparative effectiveness review is to provide information about the accuracy of non-invasive imaging technologies. This information may be useful to clinicians in deciding when it is clinically appropriate to use various types of non-invasive technologies to evaluate breast abnormalities. It is reasonable to assume that no non-invasive technology will achieve an accuracy equivalent to or better than biopsy. However, it is also reasonable to assume that non-invasive technologies are safer than invasive biopsy methods, and therefore some women may benefit from the use of particular non-invasive technologies.

In order for clinicians to decide upon the "clinical appropriateness" of a test for any particular patient, an accurate estimate of the accuracy 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 be at high risk of breast cancer (due to carrying BRCA1 and BRCA2 mutations or having a very strong family history of breast cancer) 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 review. Instead, we will focus on the use of non-invasive imaging technology for women from the general population who present with an abnormal finding by screening mammography or physical examination. We will also (if the evidence permits) examine the influence of age; the size and morphological characteristics of the lesion; and other key clinical risk factors on the accuracy of the non-invasive imaging methods.

Non-invasive Imaging

Non-invasive imaging technologies generally fall into two primary groups: technologies that examine the anatomy, or physical structure, of the breast; and technologies that detect abnormal metabolic patterns. Some non-invasive imaging technologies are slightly invasive in that they require the infusion or injection of a tracer or contrast agent; and some technologies expose patients to radiation. Each of the non-invasive technologies considered in this review is briefly introduced in the *Results* section of this report.

Conceptual Framework

The analytical framework (Figure 1) demonstrates the links between patients, tests, interventions, and outcomes. The numbers on the diagram refer to the Key Questions (see next section) and their placement in Figure 1 illustrates the many links separating the Key Questions from the patient-oriented outcomes. Fryback and Thornbury have proposed a six-level model of assessing diagnostic efficacy.[18] Level 1 is analytic validity; Level 2 is diagnostic accuracy; Level 3 is diagnostic thinking; Level 4 is impact on choice of treatment; Level 5 is patient-important outcomes; and Level 6 is societal impact. Demonstration of efficacy at each lower level is logically necessary, but not sufficient, to assure efficacy at higher levels. Patients and health-care providers are generally most interested in studies that evaluate the impact of diagnostic tests on Level 5, patient-oriented outcomes. However, studies that directly link diagnostic tests to patient-oriented outcomes are expensive, time consuming, and difficult to

conduct. In the absence of direct evidence, the effect of diagnostic tests on patient-oriented outcomes can be estimated by creating indirect chains of evidence by evaluating other levels. This systematic review is primarily concerned with Level 2, the diagnostic accuracy of various non-invasive methods of evaluating breast abnormalities. Our literature searches did not identify any relevant studies that reported the impact of the diagnostic tests on patient-oriented outcomes.

Figure 1. Analytical framework

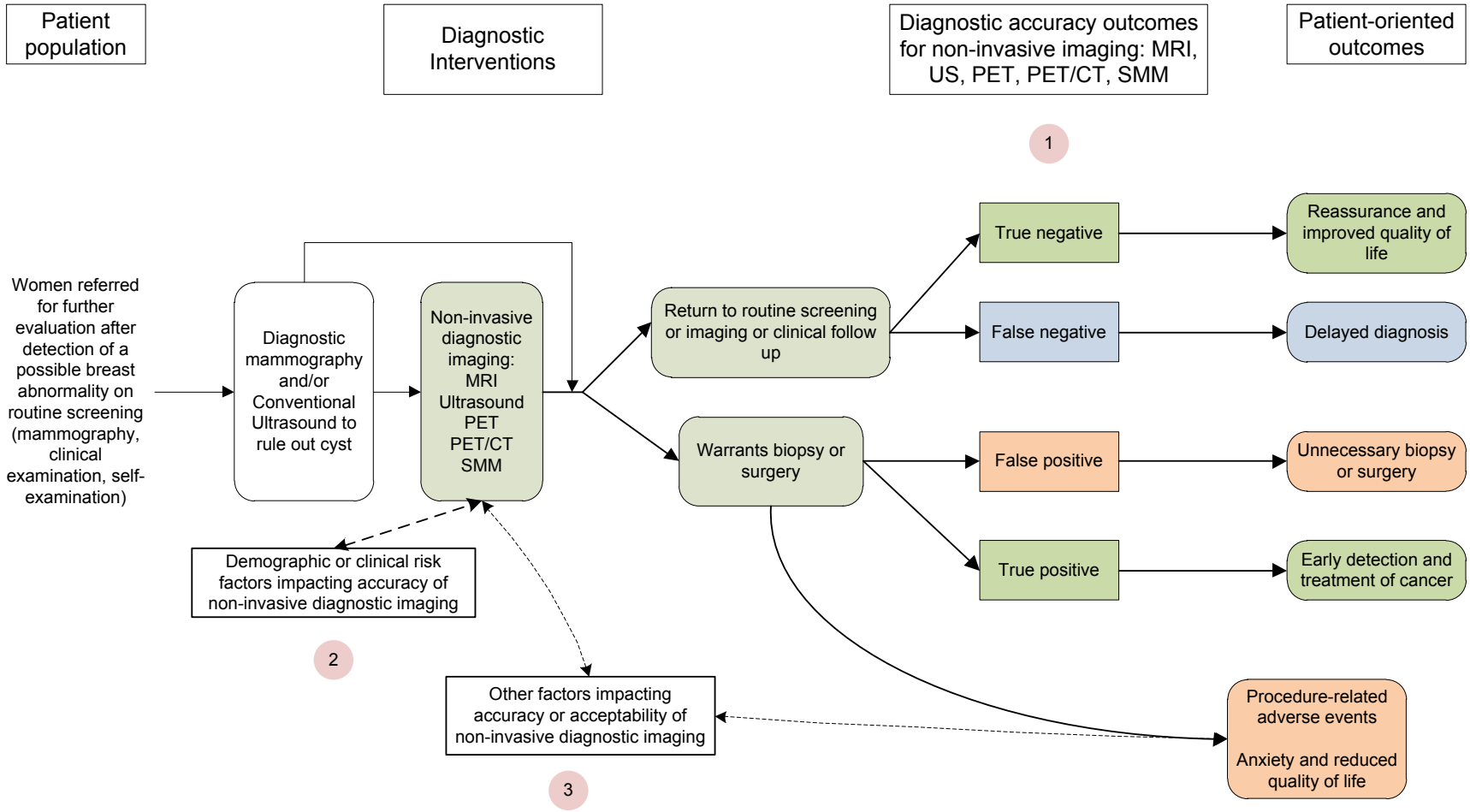


Figure 1 depicts the Key Questions within the context of the patient population, diagnostic tests, subsequent interventions, and outcomes. In general, the figure illustrates how the use of additional non-invasive imaging tests may affect decisions about patient management, and how such decisions may impact patient outcomes. The Key Questions are depicted within the figure as numbers inside circles. Abbreviations: MRI = magnetic resonance imaging; PET = positron emission tomography; CT = computed tomography; SMM = scintimammography.

Diagnostic Test Characteristics

No diagnostic test is perfect. Studies of test performance compare test results on a group of individuals, some of whom have the disease and some of whom do not. Each individual undergoes the experimental test as well as a second reference test to determine "true" disease status. The relationship between the diagnostic test results and disease status is described using diagnostic test characteristics. It is important that the reference test is very accurate in measuring "true" disease status, or else the performance of the experimental diagnostic test will be poorly estimated.

Sensitivity and specificity. The results of the experimental and reference standard test and their relationship are commonly presented as two-by-two (2x2) tables (see Table 2). From the 2x2 table, sensitivity and specificity are readily calculated:

Sensitivity = TP/(TP+FN)

Specificity = TN/(FP+TN)

Table 2. Example of a 2x2 table

		Disease	
		Present	Absent
Test Results	Positive	True positives (TP)	False positives (FP)
	Negative	False negatives (FN)	True negatives (TN)

Sensitivity and specificity are test properties that are useful when deciding whether to use the 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).

Predictive values and likelihood ratios. To make sense of a diagnostic investigation, a clinician needs to be able to make an inference regarding the probability that a patient has the disease in question according to the result obtained from the test. Sensitivity and specificity do not directly provide this information. The predictive values and likelihood ratios can also be directly calculated from a 2x2 table:

Positive predictive value = TP/(TP+FP)

Negative predictive value = TN/(FN+TN)

Positive likelihood ratio = (TP/(TP+FN))/(FP/(FP+TN))

Negative likelihood ratio = (FN/(TP+FN))/(TN/(FP+TN))

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. Predictive values describe the probabilities

that positive or negative results are correct for an individual patient. However, predictive values depend on the prevalence of disease in the population. A study that enrolled a patient population with a disease prevalence of 70% may report a positive predictive value of 80%. If a clinician tests a patient from a population with a disease prevalence of 70%, and the test comes back positive, the clinician knows the patient has an 80% chance of having the disease in question. However, if the patient comes from a population with a disease prevalence of 20%, the clinician cannot apply the results of the study directly to this patient.

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 more clinically useful diagnostic test performance measure. 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 are independent of prevalence and therefore can be directly applied in the clinic to update an individual's estimated chances of disease according to their test result. Likelihood ratios can be used in Bayes' theorem to calculate post-test odds of having a disease from the pre-test suspicion of the patient's odds of having that disease. Clinicians may be familiar with simple nomograms that allow a direct visualization of post-test chances of disease given a positive or negative test result, without the need to go through the tedious calculations of Bayes' theorem; see, for example, the interactive form of the nomogram provided by the Center for Evidence-based Medicine at http://www.cebm.net.

Scope and Key Questions

This systematic review was commissioned by the Agency for Healthcare Research and Quality (AHRQ) to address the following Key Questions:

1. What is the accuracy (expressed as sensitivity, specificity, predictive values, and likelihood ratios) of non-invasive tests for diagnosis of breast cancer in women referred for further evaluation after identification of a possible breast abnormality on routine screening (mammography and/or clinical or self-detection of a palpable lesion)?

   The non-invasive tests to be evaluated are:

   Ultrasound (conventional B-mode, color Doppler, power Doppler, tissue harmonics, and tomography)

   Magnetic resonance imaging (MRI) with breast-specific coils and gadolinium-based contrast agents, with or without computer-aided diagnosis (CADx)

   Positron emission tomography (PET) with 18-fluorodeoxyglucose (FDG) as the tracer, with or without concurrent computed tomography (CT) scans

   Scintimammography (SMM) with technetium-99m sestamibi (MIBI) as the tracer, including Breast Specific Gamma Imaging (BSGI)

2. Are there demographic (e.g., age) and clinical risk factors (e.g., morphologic characteristics of the lesion) that affect the accuracy of the tests considered in Question 1?

3. Are there other factors and considerations (e.g., safety, care setting, patient preferences, ease of access to care) that may affect the accuracy or acceptability of the tests considered in Questions 1 and 2?

This report is an update of a Comparative Effectiveness Review (CER) of the same title originally published in 2006. The Key Questions have been revised and additional diagnostic tests have been added to the list of tests to be evaluated. The 2006 version of the CER only evaluated B-mode ultrasound, MRI (without CADx), PET (without CT), and full-body scintimammography.

Chapter 2. Methods

Topic Development

In response to Section 1013 of the Medicare Modernization Act, AHRQ requested an update of the evidence report Effectiveness of Non-Invasive Diagnostic Tests for Breast Abnormalities.[19] The original report was finalized in February 2006. Due to technological advances and continuing innovation in the fields of non-invasive imaging, the conclusions of the original report are possibly no longer relevant to current clinical practice. Consequently, the topic was selected for update. The Scientific Resource Center (SRC) for the AHRQ Effective Health Care Program recruited a technical expert panel (TEP) to give input on key steps including the selection and refinement of the questions to be examined. The expert panel membership is provided in the Appendices.[1]

Upon AHRQ approval, the draft Key Questions were posted for public comment. After receipt and consideration of the public commentary, the Key Questions were finalized and submitted to AHRQ for approval. These Key Questions are presented in the Scope and Key Questions section of the Introduction.

The EPC created a work plan for developing the evidence report. The process consisted of working with AHRQ, the SRC, and the TEP to outline the report's objectives, performing a comprehensive literature search, abstracting data, constructing evidence tables, synthesizing the data, and submitting the report for peer review.

In designing the study questions and methodology at the outset of this report, the EPC consulted several technical and content experts. Broad expertise and perspectives were sought. Divergent and conflicted opinions are common and perceived as healthy scientific discourse that results in a thoughtful, relevant systematic review. Therefore, in the end, study questions, design and/or methodologic approaches do not necessarily represent the views of individual technical and content experts.

The topic development procedure employed the "PICOTS" approach; namely, carefully and clearly defining the Patients, the Intervention(s), the Comparator(s), the Outcomes, the Timing of follow-up, and the Setting of care.[20]

Patients. The patient population of interest is the general population of women participating in routine breast cancer screening programs (including mammography, clinical examination, and self-examination). Populations that will not be evaluated include: women thought to be at very high risk of breast cancer due to family history or BRCA mutations; women with a personal history of breast cancer; and men.

Interventions. The non-invasive diagnostic tests to be evaluated are:

- Ultrasound (conventional B-mode, harmonic, tomography, color Doppler, and power Doppler)
- Magnetic resonance imaging (MRI) with breast-specific coils and gadolinium-based contrast agents, with or without computer-aided diagnosis (CADx)

[1] Appendices and Evidence Tables cited in this report are provided electronically at http://www.ahrq.gov/downloads/pub/evidence/pdf/XXX/XXX.pdf.

- Positron emission tomography (PET) with 18-fluorodeoxyglucose (FDG) as the tracer, with or without concurrent computed tomography (CT) scans
- Scintimammography with technetium-99m sestamibi (MIBI) as the tracer, including Breast Specific Gamma Imaging (BSGI).

Technologies that were proposed for evaluation but, after discussion by the TEP, were not included, are: elastography; molecular breast imaging; scintimammography using tracers other than MIBI; PET using tracers other than FDG; digital tomosynthesis mammography; computer-aided diagnostic x-ray mammography; breast thermography; electrical impedance tomography; and optical breast imaging. The primary reason that the TEP decided to not include these technologies in the current CER was concern that there is currently insufficient robust evidence about each of these technologies to warrant their inclusion in a systematic review.

Comparators. The accuracy of the non-invasive imaging tests was evaluated by a direct comparison to histopathology (biopsy or surgical specimens) or to clinical follow-up, or a combination of these methods. In addition, the relative accuracy of the different tests under evaluation was evaluated by directly and indirectly comparing the tests (as the reported evidence permitted).

Outcomes. Outcomes of interest are diagnostic test characteristics, namely, sensitivity, specificity, and likelihood ratios. Adverse events related to the procedures, such as discomfort and reactions to contrast agents, will be discussed in the answer to Key Question 3.

Timing. Any duration of follow-up, from same-day interventions to many years of clinical follow-up, was be evaluated.

Setting. Any care setting was acceptable, including general hospitals, physician's offices, and specialized breast imaging centers.

Search Strategy

The medical literature was searched from December 1994 through September 2010. The full strategy is provided in Appendix A. In brief, we searched 10 external and internal databases, including PubMed and EMBASE, for clinical trials addressing the Key Questions. To supplement the electronic searches, we also examined the bibliographies/reference lists of included studies, recent narrative reviews, and scanned the content of new issues of selected journals and selected relevant gray literature sources.

Study Selection

We selected the studies that we consider in this report using *a priori* inclusion criteria. Some of the criteria we employed are geared towards ensuring that we used only the most reliable evidence. Other criteria were developed to ensure that the evidence is not derived from atypical patients or interventions, and/or outmoded technologies.

Studies of diagnostic test performance compare results of the experimental test to a reference test. The reference test is intended to measure the "true" disease status of each patient. It is important that the results of the reference test be very close to the truth, or the performance of the experimental test will be poorly estimated. For the diagnosis of breast cancer, the "gold standard" reference test is open surgical biopsy. However, an issue with the use of open surgical biopsy as the reference standard in large cohort studies of screening-detected breast

abnormalities is the difficulty of subjecting women with probably benign lesions to open surgical biopsy. Furthermore, restricting the evidence base to studies that used open surgery as the reference standard for all enrolled subjects would eliminate the majority of the evidence. Therefore, we have chosen to use a combination of clinical and radiologic follow-up as well as core-needle biopsy and open surgical biopsy as the reference standard for our analysis, although we acknowledge that this decision may cause our analysis to over-estimate the accuracy of the non-invasive tests.[21]

We used the following formal criteria to determine which studies would be included in our analysis. Many of our inclusion criteria were intended to reduce the potential for spectrum bias. Spectrum bias refers to the fact that diagnostic test performance is not constant across populations with different spectrums of disease. For example, patients presenting with severe symptoms of disease may be easier to diagnose than asymptomatic patients in a screening population; and a diagnostic test that performs well in the former population may perform poorly in the latter population. The results of our analysis are intended to apply to a general population of women participating in routine breast cancer screening programs (mammography, clinical examination, and self-examination programs) and therefore many of our inclusion criteria are intended to eliminate studies that enrolled populations of women at very high risk of breast cancer due to family history, or populations of women at risk of recurrence of a previously diagnosed breast cancer.

1. The study must have directly compared the test of interest to core-needle biopsy, open surgery, or clinical follow-up of the same patient.

   Although it is possible to estimate diagnostic accuracy from a two-group trial, the results of such indirect comparisons must be viewed with great caution. Diagnostic cohort studies, wherein each patient acts as her own control, are the preferred study design for evaluating the accuracy of a diagnostic test.[22] Studies may have performed biopsy procedures on all patients, or may have performed biopsy on some patients and followed the other patients with clinical examination and mammograms. Fine-needle aspiration of solid lesions is not an acceptable reference standard for the purposes of this assessment.[23-26]

   Retrospective cohort studies that enrolled all or consecutive patients were considered acceptable for inclusion. However, retrospective case-control studies and case reports were excluded. Retrospective case-control studies have been shown to overestimate the accuracy of diagnostic tests, and case reports often report unusual situations or individuals that are unlikely to yield results that are applicable to general practice.[21,22] Retrospective case studies (studies that selected cases for study on the basis of the type of lesion diagnosed) were also excluded because the data such studies report cannot be used to accurately calculate the overall diagnostic accuracy of the test.[21]

2. The studies must have used current generation scanners and protocols of the selected technologies only. Other non-invasive breast imaging technologies are out of the scope of this assessment.

   Studies of outdated technology and experimental technology are not relevant to current clinical practice. Definitions of "outdated technology" and "current technology" were developed through discussions with experts in relevant fields. Definitions of "current technology to be included" are defined in Table 3.

Table 3. Non-invasive current technologies to be evaluated

Technology	Cut-off Publication Date (to present) to Exclude Outdated Technology	Other Inclusion Criteria
Ultrasound	1994	
Magnetic resonance imaging (MRI)	2000	Must have used specific breast coils with at least 8 channels, and used gadolinium-based contrast agents
Computer Aided Detection (CAD) MRI	2005	Must have used specific breast coils with at least 8 channels, and used gadolinium-based contrast agents. CAD systems must be FDA approved for diagnostic breast cancer use, and are defined as stand-alone third-party packages that may be added to standard MRI systems to assist interpretation of the images.
Positron emission tomography (PET)	2000	FDG as the PET tracer
Combined PET/computed tomography (CT) systems	2000	FDG as the PET tracer
Scintimammography (SMM)	2005	Includes breast specific gamma imaging (BSGI) and also single photon emission tomography (SPECT); only studies that used sestamibi, also called MIBI, also called Technetium-99m sestamibi, as the tracer.

3. The study enrolled female human subjects.

   Animal studies or studies of "imaging phantoms" are outside the scope of the report. Studies of breast cancer in men are outside the scope of the report. However, studies of primarily female patients that enrolled one or two men were considered acceptable.

4. The study must have enrolled patients referred for the purpose of primary diagnosis of a breast abnormality detected by routine screening (mammography and/or physical examination).

   Studies that enrolled women who were referred for evaluation after discovery of a possible breast abnormality by screening mammography or routine physical examination were included. Studies that enrolled subjects that were undergoing evaluation for any of the following purposes were excluded as being out of scope of the report: screening of asymptomatic women; breast cancer staging; evaluation for a possible recurrence of breast cancer; monitoring response to treatment; evaluation of the axillary lymph nodes; evaluation of metastatic or suspected metastatic disease; or diagnosis of types of cancer

other than primary breast cancer. Studies that enrolled patients from high-risk populations such as BRCA1/2 mutation carriers, or patients with a strong family history of breast cancer, are also out of scope. If a study enrolled a mixed patient population and did not report data separately, it was excluded if more than 15% of the subjects did not fall into the "primary diagnosis of women at average risk presenting with an abnormality detected on routine screening" category.

5. Study must have reported test sensitivity, specificity, negative or positive predictive values, or sufficient data to calculate these measures of diagnostic test performance; or (for Key Question 3) reported factors that affected the accuracy of the non-invasive test being evaluated.

   Other outcomes are beyond the scope of this report.

6. Fifty percent or more of the subjects must have completed the study.

   Studies with extremely high rates of attrition are prone to bias and were excluded.

7. Study must be published in English.

   Moher et al. have demonstrated that exclusion of non-English language studies from meta-analyses has little impact on the conclusions drawn. Juni et al found that non-English studies typically were of lower methodological quality and that excluding them had little effect on effect size estimates in the majority of meta-analyses they examined. Although we recognize that in some situations exclusion of non-English studies could lead to bias, we believe that the few instances in which this may occur do not justify the time and cost typically necessary for translation of studies to identify those of acceptable quality for inclusion in our reviews.[27,28]

8. Study must be published as a peer-reviewed full article. Meeting abstracts were not 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.[29,30] *In addition, it is not uncommon for abstracts that are published as part of conference proceedings to have inconsistencies when compared to the final publication of the study, or to describe studies that are never published as full articles.*[31-35]

9. The study must have enrolled 10 or more individuals per arm.

   The results of very small studies are unlikely to be applicable to general clinical practice. Small studies are unable to detect sufficient numbers of events for meaningful analyses to be performed, and are at risk of enrolling unique individuals.

10. When several sequential reports from the same patients/study are available, only outcome data from the 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 more recent report.

The abstracts of articles identified by the literature searches were screened for possible relevance in duplicate by four analysts. All exclusions at the abstract level were approved by the lead research analyst. The full-length articles of studies that appeared relevant at the abstract

level were then obtained and three research assistants examined the articles to see if they met the inclusion criteria. All exclusions were approved by the lead research analyst. The excluded articles and primary reason for exclusion are shown in the Appendices.

Data Abstraction

Standardized data abstraction forms were created and data were entered by each reviewer into the SRS© 4.0 database (see Appendices). Three research assistants abstracted the data. The first fifty articles were abstracted in duplicate. All conflicts were resolved by the lead research analyst.

Study Quality Evaluation

We used an internal validity rating scale for diagnostic studies to grade the quality (internal validity) of the evidence base (see Appendices). This instrument is based on a modification of the QUADAS instrument with reference to empirical studies of design-related bias in diagnostic test studies.[21,36] Each question in the instrument addresses an aspect of study design or conduct that can help to protect against bias. Each question can be answered "yes," "no," or "not reported," and each is phrased such that an answer of "yes" indicates that the study reported a protection against bias on that aspect.

Responses to the questions in the quality assessment instrument for each study are presented in the Evidence Tables in Appendix C.

Strength of Evidence Assessment

We applied a formal grading system that conforms with the EPC Methods Guide for Comparative Effectiveness Reviews recommendations on grading the strength of evidence.[37,38]

The overall strength of evidence supporting each major conclusion was graded as High, Moderate, Low, or Insufficient. The grade was developed by considering four important domains: the risk of bias in the evidence base, the consistency of the findings, the precision of the results, and the directness of the evidence.

The risk of bias (internal validity) of each individual study was rated as being Low, Medium, or High; and the risk of bias of the aggregate evidence base supporting each major conclusion was similarly rated as being Low, Medium, or High. We used our inclusion/exclusion criteria to eliminate studies with designs known to be prone to bias from the evidence base. Namely, case reports, case-control studies, and retrospective studies that did not enroll all or consecutive patients were not included for analysis. Because we eliminated all studies with a High risk of bias from the evidence base, we consider the remaining evidence base to have either a Low or Medium risk of bias.

We initially used an internal validity rating instrument for diagnostic studies to grade the internal validity of the individual studies (see section above Study Quality Evaluation). However, after we had conducted meta-regressions investigating the correlation between key individual items on the quality rating instrument and the results reported by the studies (see Appendix D for details), we consistently found that the majority of the items on the instrument had no statistically significant correlation with the reported results (with one exception, discussed below). We therefore concluded that the quality instrument was not adequately capturing the potential for bias of the studies in our sample after eliminating retrospective case-control studies

and case reports. Unlike studies of interventions, diagnostic cohort studies are quite simple in design - one group of patients acting as their own controls. As long as all enrolled patients receive both the diagnostic test and the reference standard test, opportunities for bias (due to study design or conduct) to affect the results are limited. As mentioned above, we eliminated all studies with a High risk of bias due to their study design from the evidence base. We did not identify any obvious design flaws in the remaining studies that suggested they were at Medium risk of bias; therefore, we rated all of the included studies, and the aggregate evidence bases, as being at Low risk of bias.

Meta-regressions did identify a statistically significant correlation between blinding of image readers to patient clinical information and the reported results of studies of MRI and ultrasound. Studies that blinded image readers to patient clinical information generally reported the blinded image readers had less accurate findings. It may, therefore, be that lack of blinding is a design flaw that is biasing the results. However, an alternative interpretation, which we favor, is that blinding image readers to patient clinical information is an artificial construct that will rarely if ever occur in clinical practice; therefore, non-blinded studies are generating an estimate of accuracy that is closer to the "real" accuracy that can be obtained in clinical practice. The majority of the studies are either non-blinded or did not specifically state whether they were blinded, leading us to believe that our aggregate pooled summary estimate of accuracy is close to the "real" accuracy of the technologies as used in routine clinical practice.

We rated the consistency of conclusions supported by meta-analyses with the statistic I^2.[39,40] Datasets that were found to have an I^2 of less than 50% were rated as being "Consistent"; 50% or greater were rated as being "Inconsistent"; and datasets for which I^2 could not be calculated (e.g., a single study) were rated as "Consistency Unknown".

For qualitative direct comparisons between different diagnostic tests, we rated conclusions as consistent if the effect sizes were all in the same direction. For example, when comparing the accuracy of ultrasound without a contrast agent to the accuracy of ultrasound with a contrast agent, if the estimates of sensitivity of the individual studies are consistently higher for studies that used a contrast agent, then the evidence base would be rated as "consistent."

We defined a "precise" estimate of sensitivity or specificity as one for which the upper AND lower bound of the 95% confidence interval was no more than 5 points away from the summary estimate; for example, sensitivity 98% (95% CI: 97 to 100%) would be a precise estimate of sensitivity, whereas sensitivity 95% (95% CI: 88 to 100%) would be an imprecise estimate of sensitivity. Precision could be rated separately for summary estimates of sensitivity and specificity for each major conclusion.

For qualitative direct comparisons between different diagnostic tests, the conclusion is "Precise" if the confidence intervals around the summary estimates being compared do not overlap. We did not derive any formal conclusions (or formally rate the strength of evidence for any speculative statements) about indirect comparisons between different diagnostic tests.

According to the *Methods Guide*,[37]

> *The rating of directness relates to whether the evidence links the interventions directly to health outcomes.*

For studies of diagnostic test accuracy, the evidence should always be rated as "Indirect" because the outcome of test accuracy is indirectly related to health outcomes. However, the

Key Questions in this particular comparative effectiveness review do not ask about the impact of test accuracy on health outcomes. We therefore did not incorporate the "Indirectness" of the evidence into the overall rating of strength of evidence for these Key Questions because they did not ask about health outcomes.

Overall Rating of Strength of Evidence

The initial rating is based on the risk of bias. If the evidence base has a Low risk of bias, the initial strength of evidence rating is High; if the evidence base has a Moderate risk of bias, the initial strength of evidence rating is Moderate; if the evidence base has a High risk of bias, the initial strength of evidence rating is Low. For this particular comparative effectiveness review, as explained above, the rating of risk of bias was Low for all evidence bases, and therefore the initial strength of evidence rating is High. The remaining two domains are used to up- or down-grade the initial rating as per the following flow charts:

Consistent, Precise: High

Inconsistent, Precise: Moderate

Consistent, Imprecise: Moderate

Inconsistent, Imprecise: Low

"Consistency Unknown", Precise: Low

"Consistency Unknown", Imprecise: Insufficient

Evidence bases judged to be too small to support an evidence-based conclusion (e.g., one or two small studies) were simply rated "Insufficient" without formally considering the various domains. Further details about grading the strength of evidence may be found in the *Evidence* Tables section of the Appendices.

Applicability

The issue of applicability was chiefly addressed by excluding studies that enrolled patient populations that were not a general population of asymptomatic women participating in routine breast cancer screening programs. We defined the population of interest as women at average risk of breast cancer participating in routine breast cancer screening programs (including mammography, clinical examination, and self-examination). We excluded studies that enrolled women thought to be at very high risk of breast cancer due to personal history, family history, or known carriers of BRCA mutations.

Data Analysis and Synthesis

The majority of studies reported data on a per-lesion rather than a per-patient basis, and therefore we analyzed the data on a per-lesion basis assuming that statistical assumptions about data independence were not being violated. Because the number of lesions was usually very similar to the number of patients (i.e., the vast majority of patients only had one lesion) we do not believe that this assumption will have a significant impact on the results.

We performed a standard diagnostic accuracy analysis. For the diagnostic accuracy analysis,

- true negatives were defined as lesions diagnosed as benign on imaging that were found to be benign by the reference standard;

- false negatives were defined as lesions diagnosed as benign on imaging that were found to be malignant (invasive or in situ) by the reference standard;
- true positives were defined as lesions diagnosed as malignant (invasive or in situ) on imaging that were found to be malignant (invasive or in situ) on the reference standard
- false positives were defined as lesions diagnosed as malignant that were found to be benign on the reference standard.

We meta-analyzed the data reported by the studies using a bivariate mixed-effects binomial regression model as described by Harbord et al.[41] All such analyses were computed by the STATA 10.0 statistical software package using the "midas" command.[42] The summary likelihood ratios and Bayes' theorem were used to calculate the post-test probability of having a benign or malignant lesion. In cases where a bivariate binomial regression model could not be fit, we meta-analyzed the data using a random-effects model and the software package Meta-Disc.[43] Meta-regressions were also performed with the STATA software and the "midas" command. We did not assess the possibility of publication bias because statistical methods developed to assess the possibility of publication bias in treatment studies have not been validated for use with studies of diagnostic accuracy.[44,45]

Diagnostic tests all have a trade-off between minimizing false-negative and minimizing false-positive errors. False-positive errors that occur during breast screening diagnostic workups are not considered to be as clinically relevant as false-negative errors. Women who experience a false-positive error will be sent for unnecessary procedures, and may suffer anxiety and reduced quality of life. However, women who experience a false-negative error may die from a delayed cancer diagnosis.

Likelihood ratios can be used along with Bayes' theorem to directly compute an individual woman's risk of actually having a malignancy following a diagnosis on imaging. However, each individual woman's post-test risk varies by her pre-test risk of malignancy. Simple nomograms are available for in-office use that allow clinicians to directly read individual patients' post-test risk off a graph without having to go through the tedium of calculations. Predictive value is another commonly used measure of errors; however, negative and positive predictive values are specific to specific populations of women. Predictive values vary by the prevalence of disease in each specific population and should not be applied to other populations with different prevalences of disease. For this reason, we have avoided the use of predictive values in this systematic review.

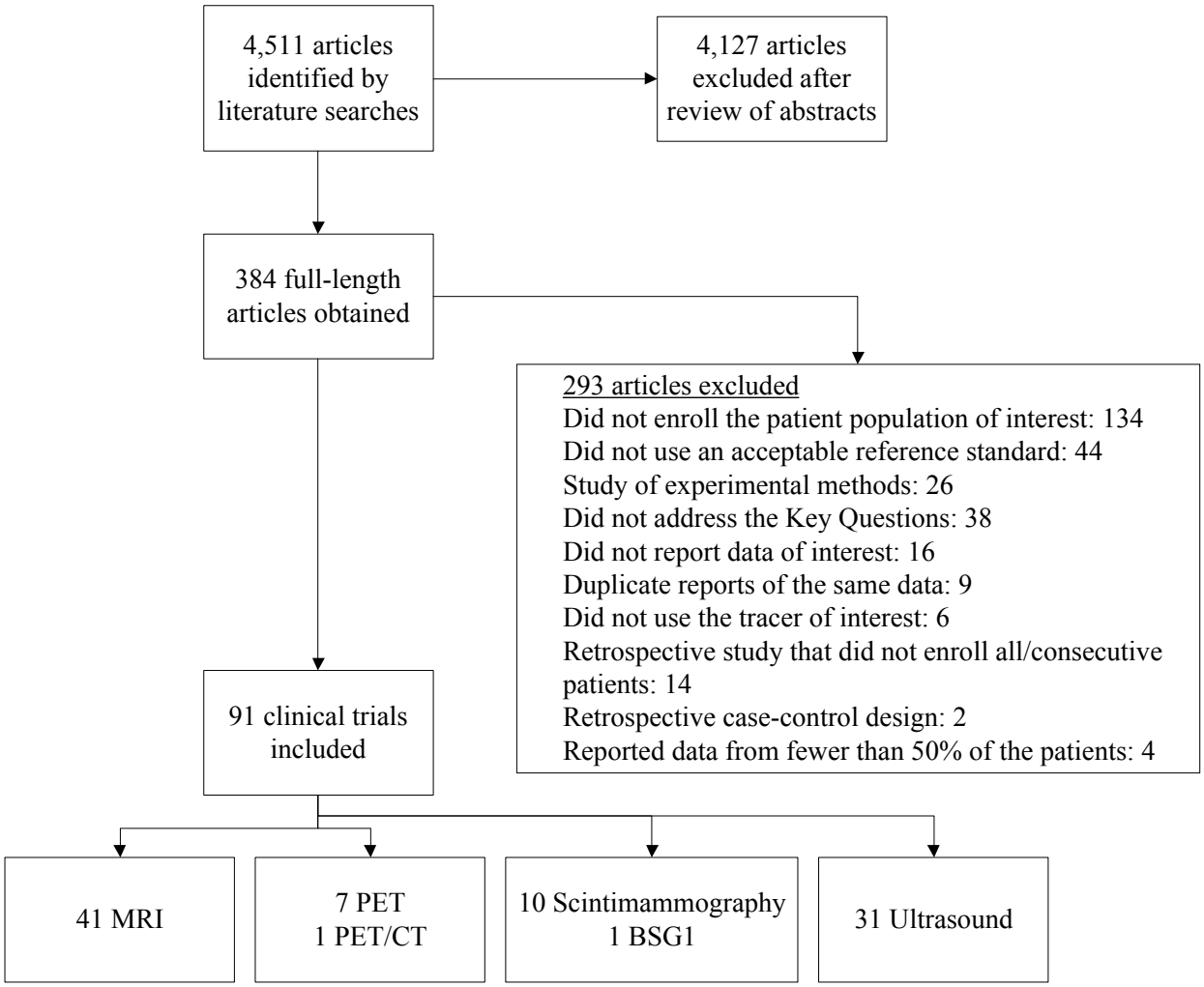
Peer Review and Public Commentary

A draft of the completed report will be sent to the peer reviewers, the representatives of AHRQ, and the SRC. The draft report will be posted to a Web site for public comment. In response to the comments of the peer reviewers and the public, revisions will be made to the evidence report, and a summary of the comments and their disposition will be submitted to AHRQ. Peer reviewer comments on a preliminary draft of this report will be considered by the EPC in preparation of the final report. Synthesis of the scientific literature presented here does not necessarily represent the views of individual reviewers.

Chapter 3. Results

Our literature searches identified a total of 4,511 possible articles. After review of the abstracts, we selected 384 for further review as full-length articles to determine whether they met the inclusion criteria. The study selection process is summarized in Figure 2. Full details of excluded articles and reasons for exclusion are shown in the Appendices.[2] The included articles are described throughout this *Results* section. We have organized the *Results* section by type of non-invasive test rather than by Key Question.

Figure 2. Study selection process



[2] Appendices and Evidence Tables cited in this report are provided electronically at http://www.ahrq.gov/downloads/pub/evidence/pdf/XXX/XXX.pdf.

Magnetic Resonance Imaging (MRI)

Background

Technology. MRI systems use strong magnetic fields and radiofrequency energy to translate hydrogen nuclei distribution in tissues into computer-generated images of the structure of the interior of the breast. MRI does not expose patients to radiation. However, the procedure is not completely non-invasive because often contrast agents are infused to improve the resolution of the images.

MRI systems are usually described primarily in terms of strength of the magnet, in the unit Tesla (T). Systems in commercial use for breast imaging usually vary from 0.5T to 3.0T. In general, increasing the strength of the magnet increases the spatial resolution of the images. MRI systems that use field strengths below 1.0T are usually open gantries and are primarily used for patients who cannot be accommodated inside the bores of higher field strength magnets due to claustrophobia. An additional reason for the use of open gantry systems is that MRI-guided invasive procedures, such as biopsies, are much easier to perform than in closed systems.[46]

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.[47] The dedicated breast coils allow the patient to lie prone with her breasts in close proximity to 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).[48,49] Coils are described by the number of channels they contain. In general, increasing the number of channels improves the signal to noise ratio.[48,50,51] Eight-channel breast coils are considered standard equipment for breast MRI examinations.

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 have been observed.[52]

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 prosthetic cardiac valves.[53] Also, respiratory motion can be a problem, although when the patient is prone the effect is reduced.[53] Interpretation of the images is a subjective procedure that requires specialized training.[54,55] Computer-based tools to partially automate the interpretation procedure are available and may reduce subjectivity and decrease time required for image interpretation.[56]

The use of contrast agents for MRI breast examinations is considered standard procedure. Gadolinium-based paramagnetic contrast agents accumulate in the vascular system and can aid in visualizing tumors by highlighting areas containing a dense blood vessel network. There are currently five slightly different gadolinium-based contrast agents in common clinical use: gadobenate dimeglumine, gadopentetate dimeglumine, gadodiamide, gadoteridol, and gadoversetamide.[57] These agents differ slightly in molecular structure; all, however, consist of

the heavy metal gadolinium bound to a chelating molecule.[58] Different agents may have different imaging properties.[59,60] When using conventional gadolinium contrast agents, the exact dose used does not appear to be particularly relevant to image quality when used in the normal range (0.1 to 0.2 mmol/kg). 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 considered indicative of malignancy.[54] In premenopausal women, the normal parenchyma can demonstrate enhancement that can decrease the specificity of breast MRI studies.[61,62] The amount of enhancement depends on the stage in the menstrual cycle. Therefore, in order to ensure accurate results, an MRI study should if possible be performed during the second week of the menstrual cycle when proliferative changes are at their lowest level.

For the purposes of this assessment, only MRI conducted on a 0.5 to 3T system using 8-channel breast coils and gadolinium-based contrast agents has been considered. These requirements were selected because they describe the systems and methods currently considered to be "standard practice" for breast imaging; other systems and methods would be unlikely to produce results that would be applicable to current clinical practice.[63]

Patient Safety and Comfort. A number of well-known safety 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.[64-69] Precautions are 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. No adverse effects have been conclusively identified in association with the magnetic fields to which patients are exposed during routine MRI scanning.[70-73] Therefore, so long as routine precautions are followed, breast MRI can be considered a safe exam for most patients.

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.[53] 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.

Gadolinium-based contrast agents are generally considered to be very safe.[74,75] However, in 2007, FDA requested that manufacturers include a new warning on the labeling of all gadolinium-based contrast agents which are used to enhance MRI.[57] The new labeling warns that the use of these agents increases the risk of development of nephrogenic systemic fibrosis (NSF) in patients with pre-existing acute or chronic severe renal insufficiency or renal dysfunction due to recent liver transplantation or hepatorenal syndrome.[76-78] NSF is a progressive, disabling, and potentially fatal disorder that leads to deposition of excessive connective tissue in the skin and internal organs. The condition was previously unknown; the typical patient is a middle-aged individual with severe renal disease who first exhibits skin changes 2 to 4 weeks after undergoing an MRI examination that used gadolinium-based contrast agents.[78]

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.[79]

Findings from 2006 Review

Our CER from 2006 included 19 prospective diagnostic cohort studies of MRI (published between 1991 and 2004) that studied a total of 2181 suspicious breast lesions.[59,80-97]

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% and the specificity was 72.4%. For lesions with microcalcifications, our analysis found that the sensitivity of MRI was 85.9% and the specificity was 75.5%.

Evidence Base

Our literature searches identified 41 diagnostic cohort studies of MRI (published 2000 through 2009) that studied a total of 3882 patients with 4202 suspicious breast lesions.[86-91,93-127] The majority of the studies used 1.5T magnets (33 studies) and gadopentetic acid enhancement (26 studies). The studies and patients are described in detail in the Appendices, and listed at the end of this subsection on MRI in Table 5.

Question 1. What is the accuracy of MRI for diagnosis of breast cancer in women referred for further evaluation after identification of a possible breast abnormality on routine screening (mammography and/or clinical or self-detection of a palpable lesion)?

We combined the data reported by all 41 studies into a bivariate binomial mixed model. The data were extremely heterogeneous (I^2 = 98.4%). The summary sensitivity of MRI for all lesions was 91.7% (95% CI: 88.5 to 94.1%) and the summary specificity was 77.5% (95% CI: 71.0 to 82.9%). These summary estimates are fairly similar to our 2006 estimates of the accuracy of MRI (at the mean threshold the sensitivity was 92.5%, and the specificity was 72.4%).

We investigated the heterogeneity with meta-regression. The variables investigated were: the strength of the magnet, the type of contrast agent used, whether the study enrolled all/consecutive patients or not, whether the study was prospective in design or not, whether all diagnoses were verified by histopathology or not, whether any financial conflicts of interest from the funding source existed or not, whether the study was multi- or single-centered, whether readers were blinded to clinical information or not, whether the study accounted for inter-reader differences or not, the geographical setting of the study, whether the study was clearly affected by spectrum bias or not, and the prevalence of disease. The prevalence of disease in the study population and whether or not readers were blinded to clinical information were both found to be statistically significantly correlated with the accuracy data reported by the studies (p = 0.02 and 0.03, respectively). However, in subgroup analyses there was a correlation between blinding of readers and prevalence of disease. Graphical analysis of prevalence of disease by accuracy failed to reveal any consistent pattern; therefore it is possible that the correlation between prevalence of disease and accuracy is an artifact caused by the correlation between blinding and enrollment of a population with a higher prevalence of disease. Studies that reported they had blinded readers

to clinical information had a lower sensitivity than non-blinded studies (86.8% vs. 93.9%) but approximately the same specificity (74.7% vs. 78.0%).

Question 2. Are there demographic (e.g., age) and clinical risk factors (e.g., morphologic characteristics of the lesion) that affect the accuracy of the tests considered in Question 1?

Two studies reported the accuracy of MRI by patient age.[88,97] One of these two studies (Bluemke et al.[97]) investigated the relative accuracy by premenopausal status vs. post-menopausal status of the patients, and reported virtually no difference in either sensitivity or specificity between groups. The other study (Imbriaco et al.[88]) reported the accuracy of MRI for women 50 years of age and older vs. younger women, and found that MRI was more sensitive (100% vs. 92.9%) in younger women, but had virtually the same specificity (75.0%) in both age groups.

Eight of the studies enrolled patients who had been referred for further investigation after identification of microcalcifications on mammography.[86,97,107,109,110,112,121,127] When combined in a bivariate mixed-effects model the data from these eight studies had very low heterogeneity (I^2 = 3.86%). The summary sensitivity of MRI for microcalcifications was 84.0% (79.5 to 88.3%) and the summary specificity was 79.4% (71.5 to 85.6%). The summary sensitivity of MRI for evaluation of microcalcifications is considerably lower than the sensitivity of MRI for evaluation of any/all lesions (84.0% vs. 91.7%). The specificity for microcalcifications is approximately the same (79.4% vs. 77.5%). Two studies also directly compared the sensitivity of MRI for evaluation of microcalcifications vs. other types or all types of lesions (Bluemke et al.[97] and Van Goethem et al.[127]) and reported similar results: the sensitivity of MRI for evaluation of microcalcifications is approximately 85%, which is considerably lower than the sensitivity of MRI for evaluation of all/other types of lesions; whereas the specificity of MRI for evaluation of microcalcifications is approximately 77%, which may be slightly higher than the specificity of MRI for evaluation of all/other types of lesions.

Two studies evaluated the accuracy of MRI for dense breasts vs. all or non-dense breasts (Bluemke et al.[97] and Wiberg et al.[91]), and reported virtually no difference in the accuracy of MRI for evaluation of these different categories of breast tissue.

One study enrolled only patients with lesions classified as BIRADS 3 before investigation by MRI (Gokalp and Topal[111]); however, only one enrolled patient was found to have a malignancy and therefore the patient population is too small to draw conclusions about the accuracy of MRI for probably benign lesions.

One study each investigated the accuracy of MRI for lesions broken down by palpable vs. non-palpable (Bluemke et al.[97]) and large lesion vs. small lesion (Imbracio et al.[88]).

Question 3. Are there other factors and considerations that may affect the accuracy or acceptability of MRI?

One study reported the accuracy of MRI images interpreted with and without a Computer Aided Diagnosis (CAD) software system.[99] The study reported virtually no difference in either sensitivity (77.4% vs. 78.9%) or specificity (73.2% vs. 73.2%) with or without CAD assistance.

Previously Published Systematic Reviews

We identified three systematic reviews of the use of MRI to evaluate women with prior clinical findings that suggest the possibility of breast cancer; two were published prior to the release of the 2006 version of this report. The methods and conclusions of these reviews are summarized in Table 4. The authors of two of the 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; the authors of the third review (Peters et al.) did not speculate on the clinical utility of MRI.[128]

Table 4. Other published technology assessments of MRI

Study	Methods	Conclusions
Peters et al. 2008[128]	Systematic review of the literature on the diagnostic performance of contrast-enhanced MRI for breast lesions. The review included studies published 1985 through 2005, and identified 44 studies of 3101 women who had both MRI and breast biopsies. Summary ROC was fitted, and bivariate analyses were performed.	The summary sensitivity of MRI was 90% (95% CI: 88 to 92%), and the specificity was 72% (95% CI: 67% to 77%). Meta-regressions found that the prevalence of cancer in the population being studied affected the accuracy, as did the criteria used to identify lesions as malignant.

Table 4. Other published technology assessments of MRI (continued)

Study	Methods	Conclusions
The Blue Cross/Blue Shield Technology Evaluation Program, published in 2002[129] and then updated in 2004[130]	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[131]	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 were combined meta-analytically using the method of Littenburg and Moses.[132] 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.

Conclusion

We found that the summary sensitivity of MRI for all lesions is 91.7% (95% CI: 88.5 to 94.1%) and the summary specificity is 77.5% (95% CI: 71.0 to 82.9%). The data are inconsistent (namely, demonstrated signficant heterogeneity in our statistical model), but the estimate of sensitivity is precise, therefore the strength of evidence supporting the estimate of the sensitivity of MRI is moderate. The estimate of specificity is imprecise, and therefore the strength of evidence supporting the estimate of specificity of MRI is low.

The only patient or lesion "factor" that was found to affect the accuracy of MRI and that had sufficient evidence to support a conclusion was the consistent finding that the sensitivity of MRI for evaluation of microcalcifications is considerably lower than the sensitivity of MRI for evaluation of any/all lesions. The strength of evidence supporting this conclusion was rated as high.

To aid in interpretation of these findings, we used Bayes' theorem and the summary likelihood ratios for MRI used to evaluate lesions in general and to evaluate lesions with microcalcifications (see Table 7 and Table 8). These calculations suggest that MRI examinations may not be clinically useful for patients who are strongly suspected to have either malignant or benign breast lesions after mammography, because the input provided by the MRI examinations would probably not affect the suspicion of malignancy sufficiently to alter clinical decisions about management of the patient (e.g., recommendations for biopsy vs. follow-up). However, the results of an MRI examination conducted on patients suspected of being at a real but low risk of having a malignant lesion after mammography may be useful for both patients and clinicians when making decisions about management.

These calculations suggest that MRI examinations may be clinically useful for some women, because the input provided by the MRI examination may affect the suspicion of malignancy, and therefore may alter clinical decisions about management of the patient (e.g., recommendations for biopsy vs. follow-up).

Table 5. Included studies: magnetic resonance imaging (MRI)

Study	MRI Methods Studied	Design*	N Patients
Akita et al. 2009[98]	1.5T gadodiamide	Diagnostic cohort study	50
Baltzer et al. 2009[99]	1.5T gadopentetic acid CAD assistance vs. not	Prospective diagnostic cohort	329
Hara et al. 2009[100]	1.5T gadodiamide	Diagnostic cohort study	103
Kim et al. 2009[101]	1.5T gadopentetic acid	Diagnostic cohort study	249
Lo et al. 2009[102]	3T gadopentetic acid	Prospective diagnostic cohort	31
Imbracio et al. 2008[103]	1.5T gadopentetic acid	Prospective diagnostic cohort	44

Table 5. Included studies: magnetic resonance imaging (MRI) (continued)

Study	MRI Methods Studied	Design*	N Patients
Pediconi et al. 2008[104]	1.5T gadopentetic acid vs. gadobenic acid	Prospective diagnostic cohort	47
Vassiou et al. 2009[105]	1.5T gadopentetic acid	Prospective diagnostic cohort	69
Brem et al. 2007[106]	1.5T gadopentetic acid	Diagnostic cohort study	23
Cilotti et al. 2007[107]	1.5T gadoteridol	Retrospective	55
Pediconi et al. 2007[108]	1.5T gadobenic acid	Prospective diagnostic cohort	164
Zhu et al. 2007[109]	1.5T gadodiamide	Retrospective	52
Bazzocchi et al. 2006[110]	1.0 or 1.5 T gadoteridol	Prospective diagnostic cohort	174
Gokalp and Topal 2006[111]	1.5T gadopentetic acid	Prospective diagnostic cohort	43
Kneeshaw et al. 2006[112]	1.5T gadopentetic acid	Prospective diagnostic cohort	88
Ricci et al. 2006[113]	1.5T gadobenic acid	Prospective diagnostic cohort	48
Pediconi et al. 2005[114]	1.5T gadobenic acid	Prospective diagnostic cohort	36
Pediconi et al. 2005[115]	1.5T gadopentetic acid vs. gadobenic acid	Prospective diagnostic cohort	26
Wiener et al. 2005[116]	1.5 T gadopentetic acid	Prospective diagnostic cohort	65
Bluemke et al. 2004[97]	1.5T gadopentetic acid	Prospective diagnostic cohort	960
Huang et al. 2004[96]	1.5T gadodiamide	Prospective diagnostic cohort	50
Bone et al. 2003[95]	1.5T gadopentetic acid	Prospective diagnostic cohort	97
Daldrup-Link et al. 2003[117]	1.5T gadopentetic acid	Prospective diagnostic cohort	14
Heinisch et al. 2003[94]	1.0T gadopentetic acid	Prospective diagnostic cohort	36
Walter et al. 2003[93]	1.0T gadopentetic acid	Prospective diagnostic cohort	40
Guo et al. 2002[118]	1.5T gadopentetic acid	Retrospective diagnostic cohort	52

Table 5. Included studies: magnetic resonance imaging (MRI) (continued)

Study	MRI Methods Studied	Design*	N Patients
Kelcz et al. 2002[119]	1.5T gadodiamide	Prospective diagnostic cohort	62
Schedel et al. 2002[120]	1.5T gadopentetic acid	Diagnostic cohort study	65
Trecate et al. 2002[121]	1.5T gadopentetic acid	Prospective diagnostic cohort	28
Wiberg et al. 2002[91]	1.5T gadopentetic acid	Prospective diagnostic cohort	93
Brix et al. 2001[90]	1.5T gadopentetic acid	Prospective diagnostic cohort	14
Cecil et al. 2001[89]	1.5T gadopentetic acid	Diagnostic cohort study	37
Furman-Haran et al. 2001[122]	1.5T gadodiamide	Prospective diagnostic cohort	40
Imbriaco et al. 2001[88]	0.5T gadopentetic acid	Prospective diagnostic cohort	49
Malich et al. 2001[87]	1.5T gadopentetic acid	Diagnostic cohort study	94
Nakahara et al. 2001[86]	0.5T gadopentetic acid	Retrospective review of patients with microcalcifications on mammogram	40
Torheim et al. 2001[123]	1.5T gadodiamide	Prospective diagnostic cohort	127
Wedegartner et al. 2001[124]	1.0T gadopentetic acid	Prospective diagnostic cohort	53
Yeung et al. 2001[125]	1.5T gadopentetic acid	Diagnostic cohort study	30
Kvistad et al. 2000[126]	1.5T gadodiamide	Prospective diagnostic cohort	130
Van Goethem et al. 2000[127]	NR T gadopentetic acid	Retrospective review of patients with microcalcifications or a problem after clinical examination/mammogram/US	75

* At times it was difficult to determine if a study was prospective or retrospective, and in those cases we defaulted to simply calling it a "diagnostic cohort study".

Table 6. Magnetic resonance imaging (MRI) accuracy

	N Studies	N Lesions	Summary Sensitivity (95% CI)	Summary Specificity (95% CI)	Strength of Evidence
MRI, overall	41	3,882	91.7% (88.5 to 94.1%)	77.5% (71.0 to 82.9%)	Moderate (sensitivity)/ Low (specificity)
MRI, lesions with microcalcifications	8	692	84.3% (79.5 to 88.3%)	79.4% (71.5 to 85.6%)	High (sensitivity), Moderate (specificity)
MRI, dense breasts vs. others	2	935	Results were inconsistent	Results were inconsistent	Insufficient
MRI, lesions classified as BIRADS 3 before MRI imaging	1	56	100.0% (20.8 to 99.2%)	96.4% (87.5 to 98.9%)	Insufficient
MRI, palpable lesions vs. non-nonpalpable lesions	1	821	MRI is more sensitive for palpable lesions	MRI is more specific for non-palpable lesions	Insufficient
MRI, small lesions vs. larger lesions	1	53	MRI is more sensitive for larger lesions	MRI is more specific for larger lesions	Insufficient
MRI, readers blinded vs. not	41	3,882	Sensitivity is lower if readers are blinded to patient clinical information	Specificity is not affected	Moderate
MRI, CAD assistance vs. not	1	451	Sensitivity is not affected	Specificity is not affected	Insufficient
MRI, patient age	2	874	Results were inconsistent	Specificity is not affected	Insufficient

Table 7. Clinical interpretations of magnetic resonance accuracy: benign finding on MRI

Pre-test Probability of the Lesion Being Malignant	Post-test Probability of the Lesion Being Malignant Despite a Finding of "Benign" on the MRI Exam	
	Lesions in General	Lesions with Microcalcifications
1%	0% (0 to 0%)	0% (0 to 0%)
5%	1% (0 to 1%)	1% (0% to 1%)
10%	1% (1 to 2%)	2% (2 to 3%)
20%	3% (2 to 4%)	5% (4 to 6%)
30%	5% (3 to 6%)	8% (6 to 10%)
40%	7% (5 to 9%)	12% (9 to 15%)
50%	10% (7 to 13%)	16% (13 to 21%)
60%	14% (11 to 18%)	23% (18 to 28%)
70%	20% (16 to 26%)	31% (26 to 38%)
80%	31% (24 to 38%)	44% (37 to 51%)
90%	50% (42 to 57%)	64% (57 to 70%)

Table 8. Clinical interpretations of MRI accuracy: malignant finding on MRI

Pre-test Probability of the Lesion Being Malignant	Post-test Probability of the Lesion Being Malignant After a Finding of "Malignant" on the MRI Exam	
	Lesions in General	Lesions with Microcalcifications
1%	4% (3 to 5%)	4% (3 to 5%)
5%	18% (14 to 22%)	18% (13 to 23%)
10%	31% (26 to 37%)	31% (25 to 39%)
20%	50% (44 to 57%)	51% (42 to 59%)
30%	64% (57 to 69%)	64% (56 to 71%)
40%	73% (67 to 78%)	73% (66 to 79%)
50%	80% (76 to 84%)	80% (75 to 85%)
60%	86% (82 to 89%)	86% (81 to 90%)
70%	90% (88 to 93%)	91% (87 to 93%)
80%	94% (93 to 95%)	94% (92 to 96%)
90%	97% (97 to 98%)	97% (96 to 98%)

Positron Emission Tomography (PET)

Background

Technology. Positron emission tomography (PET) is a nuclear imaging modality that uses radioactive tracers to provide images of metabolic processes. Several different radiopharmaceuticals can be used in PET imaging. The tracer most commonly used is ^{18}F-fluorodeoxyglucose (FDG). ^{18}F (fluorine-18) is a positron-emitting radionuclide, and this assessment will focus exclusively on PET scans that used FDG as a tracer. Fluorodeoxyglucose is a glucose analog that accumulates in tissue in proportion to the tissue's metabolic activity. Rapidly dividing tumor cells metabolize large amounts of glucose. The uptake of the radioactive tracer FDG can be monitored by PET and provide images of regional glucose metabolism. Areas of elevated metabolism, which may be tumor cells, can be visualized on the PET images.

When performing a PET scan, a small amount of 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. 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. The performance of the different cameras may vary. However, it is not clear how clinically relevant these differences are with respect to the accuracy of breast imaging.[133]

Other factors may also affect the quality of the breast image acquired through a PET scan. In general, longer image acquisition times will improve the image quality of any PET scan.[133] However, other factors such as patient movement, comfort, and workflow suggest that acquisition times be kept to minimum. The optimum time depends on the characteristics of the detector, with dedicated breast cameras requiring the least amount of time (four to five minutes) and whole body scanners requiring the most time (45 to 60 minutes) to acquire the full image.[133]

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 ("attenuation correction"). In breast imaging some operators believe that attenuation correction is essential for tumor localization and quantification of uptake.[134]

The standardized uptake value (SUV), which is the mean tracer activity detected normalized for the injected dose of tracer and body weight, is dependent on the image reconstruction algorithm.[135] The reconstruction algorithm is manufacturer dependent. Therefore, diagnostic performance of breast PET imaging may vary across manufacturers. Diagnostic performance may also vary depending on study-specific factors such as FDG uptake time, patient motion, size of the lesion, histology of lesion, patient weight, blood glucose level, patient position, spatial resolution, and interpretation of the breast image.[134,136,137]

According to Rosen et al., stand-alone whole-body PET scanners for oncology indications are rapidly becoming obsolete.[138] Combined computed tomography (CT)/PET systems are increasingly available and currently account for almost all of the new whole-body PET installations. These systems allow images of metabolism and anatomy to be obtained at the same time. The combined machine uses x-rays to generate 3D anatomical images (CT scanning) upon

which the PET images of metabolism can be overlaid on a computer workstation. In this report, whole-body scanners that combine PET with CT and stand-alone PET scanners will be considered as separate technologies.

Patient Safety and Comfort. Using a typical dose for a whole-body scan, the effective radiation dose delivered during a typical PET study is 19 μSv/MBq (the value depends on how often the patient voids). This translates to 7.6 mSv for a typical 400 MBq whole-body PET exam. The use of a combined CT/PET scanner also exposes the patient to x-rays. A typical abdominal CT scan exposes the body to approximately 10 mSv, for a total of around 18 mSv for a single PET/CT study.[139] Studies of atomic-bomb survivors and radiation workers have found a significant increase in the risk of cancer after exposure to as little as 20 mSv.[139] Therefore, radiation dose from PET/CT scans may be a health concern. Following the exam, the short half-life of ^{18}F means that additional precautions, such as avoiding public transportation, are not necessary.[140]

The intravenous administration of any pharmaceutical could lead to an adverse reaction. In a retrospective analysis of 81,801 administrations of PET radiopharmaceuticals, the number of serious adverse reactions reported was zero.[141] Therefore, PET radiopharmaceuticals can be considered safe. 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 15 minutes to an hour, depending on the coverage of the study. No significant patient comfort issues have been reported.

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.[142]

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.

Findings from 2006 Review

In the 2006 version of this CER, we included eight prospective diagnostic cohort studies of 226 breast lesions that were examined by whole-body PET scanning[90,93,94,143-147] and one study of 50 patients that compared whole-body PET scanning to PET imaging with a gamma camera.[148] We found that for suspicious lesions in general, 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% and the specificity was 78.3%. There were no or insufficient data to come to any conclusions about the use of PET to evaluate any sub-populations of patients. Finally, we found that whole-body PET scanning was more accurate than

gamma camera PET imaging for ruling out breast cancer. No studies of dedicated breast PET scanners met the inclusion criteria.

Evidence Base

Our literature searches identified seven diagnostic cohort studies of 18-fluorodeoxyglucose PET that met our inclusion criteria[90,93,94,143,149-151] and one study of the diagnostic value of dual-time point FDG-PET/CT.[103] All of the studies used a whole-body PET scanner.

The included studies enrolled 398 patients who were all women with suspicious lesions detected by physical exam, mammography, or ultrasound. Overall, a total of 403 lesions were detected. One of the studies excluded patients with lesions smaller than 1.0 cm (Brix et al.[90]). Patients ranged in age from 21 to 91, and reported mean ages ranged from 48.3 to 58.0, suggesting that the patient populations studied are younger than the typical breast cancer population. In all seven studies, final diagnosis was established through biopsy or surgery. One study also clinically followed patients who were diagnosed as benign at biopsy (Kaida et al. 2008[149]). The included studies are listed in Table 9 at the end of this subsection on PET, and are described in detail in the Appendices.

The single included study of PET/CT enrolled a total of 44 patients with 55 suspicious breast lesions detected by physical examination, mammography, or ultrasound.[103] No studies of dedicated breast PET scanners met the inclusion criteria.

Question 1. What is the accuracy of PET for diagnosis of breast cancer in women referred for further evaluation after identification of a possible breast abnormality on routine screening (mammography and/or clinical or self-detection of a palpable lesion)?

Seven studies reported results for 403 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), summarized in Table 9. When combined in a mixed-effects bivariate model, the summary sensitivity of PET for all lesions was 83.0% (95% CI: 73.0 to 89.0%), and the summary specificity was 74.0% (95% CI: 58 to 86%), findings that are virtually identical to our estimates in the 2006 CER. However, the data were found to contain significant heterogeneity (I^2 = 64.0%), indicating substantial variability across the study results. The observed heterogeneity could not be explained through meta-regression using the following covariates: position (prone versus supine), mostly palpable lesions (>75% versus <75% or not reported), and blinded to patient clinical information (versus not blinded or not reported).

Because the PET data are inconsistent and imprecise, we rated the strength of evidence supporting the estimate of accuracy as “low.”

The study of PET/CT was a single-center study that enrolled a total of 44 patients with 55 suspicious breast lesions detected by physical examination, mammography, or ultrasound.[103] PET scanning was performed at two time points. The first acquisition (Time 1) occurred immediately after an initial whole-body PET scan, and the second one (Time 2) occurred three hours after the first. At both time points, the images of the breast were acquired in the prone position. The CT data were used for attenuation correction, and images were reconstructed using a standard iterative algorithm.

The authors reported that dual-time point PET/CT (Time 2) demonstrated a sensitivity of 80% and specificity of 100% compared to a sensitivity of 62% and specificity of 100% for single time-point PET/CT. The authors concluded that malignant lesions showed a significant increase in FDG over time compared to benign lesions.

Question 2. Are there demographic (e.g., age) and clinical risk factors (e.g., morphologic characteristics of the lesion) that affect the accuracy of the tests considered in Question 1?

In three of the seven studies that addressed Question 1, the majority (>75.0%) of the women presented with palpable breast lesions— Kiada et al.[149]: 88.0% palpable, Schirrmeister et al.[151]: 76.0%, and Yutani et al.[143]: 93.0% palpable. Because there were only three studies, we could not fit the data in a bivariate model. Instead, we pooled the reported sensitivities and specificities in random-effects meta-analyses. However, the data were heterogeneous (I^2 = 68.0% and I^2 = 54.6% for sensitivity and specificity, respectively), indicating substantial variability among the study results. With only three studies, we did not attempt to explore possible reason(s) for the heterogeneity. The overall sensitivity for primarily palpable lesions is higher than that for all seven studies considered under Key Question 1 (86.5% vs. 83.0%), but the specificity is lower (64.2% vs. 74.0%).

One study directly compared images acquired when patients were in prone position to images of the same patients in supine position.[149] In this study by Kaida et al. 2008, 118 women with 122 lesions suspected of having breast cancer underwent whole-body PET in the supine position immediately followed by prone breast PET imaging. According to the results reported in the study, the sensitivity and specificity of images in the supine position were 83.0% and 50.0%, respectively. The sensitivity and specificity of images in the prone position were 96.0% and 50.0%, respectively.

One study, Yutani et al. 2000, reported results separately for patients with BIRADS 5, lesions 1.5 cm or larger, and who were younger than 65.[143] The authors reported that PET was more sensitive for larger lesions, but the specificity was unchanged; and for the other factors, the accuracy of PET was virtually the same as for PET for all patients.

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

None of the seven studies on stand-alone PET scanning or the one study on PET with CT reported information that addressed this question.

Previously Published Systematic Reviews

We identified two systematic reviews of PET for differential diagnosis of breast lesions. The 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.[152,153] The review included 13 articles published before March 2001. A more recent review was written by Escalona et al. and published in 2010.[154] This review included 16 studies of PET for diagnosis of breast lesions published before February 2007.

Sampson et al. performed a meta-analysis using a random-effects model, and selected a point on the summary ROC 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.[152,153]

Escalona et al. conducted a narrative discussion of the included studies and their findings. The authors concluded that "FDG-PET does not appear to be sufficiently accurate to be used in isolation for ruling out the presence of a primary tumour."[154]

Conclusion

We found that the summary sensitivity of PET for all lesions is 83.0% (95% CI: 73.0 to 89.0%) and the summary specificity is 74.0% (95% CI: 58.0 to 86.0%). The data are, however, inconsistent and imprecise, therefore the strength of evidence supporting the estimate of the accuracy of PET is low.

There was insufficient data reported by the studies to conclude much about the impact of various factors on the accuracy of PET. PET may be equally accurate for evaluation of palpable lesions as for evaluation of lesions in general, but only three studies reported information about palpable lesions only.

To aid in interpretation of these findings, we used Bayes' theorem and the summary likelihood ratios for PET used to evaluate lesions in general (see Table 11 and Table 12). These calculations suggest that PET examinations may not be clinically useful for patients who are strongly suspected to have either malignant or benign breast lesions after mammography, because the input provided by the PET examinations will probably not affect the suspicion of malignancy sufficiently to alter clinical decisions about management of the patient (e.g., recommendations for biopsy vs. follow-up). However, the results of an PET examination conducted on patients suspected of being at a real but low risk of having a malignant lesion after mammography may be useful for both patients and clinicians when making decisions about management.

These calculations suggest that PET examinations may be clinically useful for some women, because the input provided by the PET examination may affect the suspicion of malignancy, and therefore may alter clinical decisions about management of the patient (e.g., recommendations for biopsy vs. follow-up).

Table 9. Included studies: PET and PET/CT

Study	PET Methods Studied	Study Design*	Number of Patients
Imbriaco et al. 2008[103]	PET/CT	Diagnostic cohort study	44
Kaida et al. 2008[149]	Whole body PET	Prospective cohort study	118
Buchmann et al. 2007[150]	Whole body PET	Prospective cohort study	29
Hienisch et al. 2003[94]	Whole body PET	Prospective cohort study	36
Walter et al. 2003[93]	Whole body PET	Prospective cohort study	44
Brix et al. 2001[90]	Whole body PET	Prospective cohort study	14
Schirrmeister et al. 2001[151]	Whole body PET	Prospective cohort study	117
Yutani et al. 2000[143]	Whole body PET	Prospective cohort study	40

* At times it was difficult to determine if a study was prospective or retrospective, and in those cases we defaulted to simply calling it a "diagnostic cohort study".

Table 10. PET accuracy

Category	N Studies	N Lesions	Summary Sensitivity (95% CI)	Summary Specificity (95% CI)	Strength of Evidence
PET	7	403	83.0% (73.0 to 89.0%)	74.0% (58.0 to 86%)	Low
PET/CT	1	55	80% (63 to 89%)	100% (63 to 100%)	Insufficient
PET, palpable lesions	3	275	86.5% (81.4 to 90.7%)	64.2% (49.8 to 76.9%)	Low
PET, prone vs. supine	1	122	PET performed in the prone position is more sensitive	Patient position did not affect specificity of PET	Insufficient
PET, BIRADS 5 lesions	1	26	93% (76.5% to 99.1%)	100.0% (15.7% to 84.3%)	Insufficient
PET, large lesions	1	27	79.4% (62.1% to 91.3%)	100.0% (2.5% to 100.0%)	Insufficient
PET, patients younger than age 65	1	25	78.1% (60.0% to 90.7%)	100.0% (15.8% to 100.0%)	Insufficient

Table 11. Clinical interpretations of PET accuracy: benign finding on PET

Pre-test Probability of the Lesion Being Malignant	Post-test Probability of the Lesion Being Malignant Despite a Finding of "Benign" on the PET Exam
	Lesions in General
1%	0% (0 to 0%)
5%	1% (1 to 2%)
10%	3% (2 to 4%)
20%	6% (4 to 8%)
30%	9% (6 to 14%)
40%	14% (9 to 20%)
50%	19% (13 to 27%)
60%	26% (18 to 36%)
70%	36% (26 to 46%)
80%	49% (38 to 60%)
90%	68% (57 to 77%)

Table 12. Clinical interpretations of PET accuracy: malignant finding on PET

Pre-test Probability of the Lesion Being Malignant	Post-test Probability of the Lesion Being Malignant After a Finding of "Malignant" on the PET Exam
	Lesions in General
1%	3% (2 to 5%)
5%	14% (9 to 22%)
10%	26% (17 to 38%)
20%	44% (32 to 57%)
30%	58% (45 to 70%)
40%	68% (56 to 78%)
50%	76% (66 to 84%)
60%	83% (74 to 89%)
70%	88% (82 to 93%)
80%	93% (88 to 96%)
90%	97% (94 to 98%)

Scintimammography

Background

Technology. Scintimammography (SMM) 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}technetium-sestamibi (MIBI), and this assessment will only evaluate studies that used MIBI as the tracer. MIBI has a strong affinity for breast tumors, but may also accumulate in areas of inflammation or infection.[155] A method of improving visualization of tumor tissue specifically is "double phase" SMM, in which two sets of images, one acquired immediately after administration of the tracer, and one approximately 30 minutes later, are acquired and compared. Gamma cameras used for scintimammography are designed to perform either planar imaging or single photon emission tomography (SPECT). In planar imaging, each imaged point represents the superimposition of all materials in front and behind it over-laid into a two-dimensional image. This causes objects that are perpendicular to the image to appear shortened.[156] SPECT is a technique that uses multiple camera heads and computer processing to create a three-dimensional representation of the administered radiopharmaceutical taken up by tissue.

Scintimammography with MIBI may have limited spatial resolution for demonstrating cancers with diameters smaller than 10 mm.[157-159] The sensitivity of scintimammography has also been reported to be affected by type of tumor, size of tumor, and the phase of the menstrual cycle.[160] Scintimammography has been reported to be unaffected by the presence of a breast implant or by the density of the breast tissue.[160]

Breast specific gamma imaging (BSGI) is an offshoot of scintimammography. In 1999, Dilon Technologies received FDA 510(k) clearance for a BSGI camera. Their current product, the Dilon 6800®, is purported to overcome the obstacles of traditional scintimammography by providing a high resolution image with a small field of view. Specifically, the manufacturer claims it can identify very early stage cancers, about 1 mm in size; is not affected by breast density; can differentiate benign from malignant lesions; and is smaller than traditional gamma imaging systems, allowing for easy portability from site to site.[161]

Patient Safety and Comfort. Intravenous injection of MIBI has been associated with very few reported adverse reactions.[162] A case of a patient without a past history of allergies, who developed a rash following administration of MIBI, has been reported in the literature.[163] Another study reports, in addition to rash development, patients experiencing a strange taste following injection of MIBI.[164]

Other than removal of all clothing and jewelry above the waist, no special preparation is required of patients undergoing a scintimammography imaging study. Compared to other breast imaging procedures, scintimammography imaging takes longer to perform – forty minutes or more.[165] During a typical study, the patient is placed in a prone position with the breast to be imaged hanging down.[166] Although taut compression of the breast to be imaged is not required, prevention of cross-talk may require compression of the opposite breast.[158,167]

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.[142]

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.

Findings from 2006 Review

Forty-four diagnostic cohort studies published in 45 manuscripts met our inclusion criteria.[81,84,85,88,95,118,143,145,168-204] 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%. For palpable lesions and suspicious breast lesions in general, there was unexplained heterogeneity in the data, and therefore summary diagnostic test characteristics were not calculated.

Evidence Base

Our literature searches identified a total of 11 studies of 1,064 patients that met the inclusion criteria for Key Question 1. One study evaluated BSGI;[106] another tested planar and SPECT imaging combined;[205] five studies assessed double-phase scintimammography;[101,206-209] and the remaining four studies assessed planar imaging.[210-213] These studies are described in detail in the Appendices, and are listed at the end of this subsection on scintimammography in Table 14.

Question 1. What is the accuracy of scintimammography for diagnosis of breast cancer in women referred for further evaluation after identification of a possible breast abnormality on routine screening (mammography and/or clinical or self-detection of a palpable lesion)?

When all 11 studies were combined in the analysis, regardless of imaging technique(s) used, the summary sensitivity of SMM for all lesions was 84.7% (78.0 to 89.7%) and the summary specificity was 77.0% (95% CI: 64.7 to 85.9%). We also meta-analyzed the data reported by the nine included studies that used standard SMM (planar and double-phase imaging) by fitting a bivariate mixed-effects model. The summary sensitivity of standard SMM for all lesions was 84% (95% CI: 76% to 89%) and the summary specificity was 79% (95% CI: 63% to 89%), approximately the same as for the full dataset. In 2006, we found that the sensitivity of scintimammography was 68.7% and the specificity was 84.8%. Improvements in technology and techniques since then, such as the development of double-phase imaging, may explain the improved accuracy in the more recent studies.

There was a great deal of heterogeneity (I^2 = 93%) in the reported data. We were unable to identify with meta-regression any study- related characteristics that explained this heterogeneity, such as consecutive enrollment of patients, blinding of the diagnostic test reader to patient history/other clinical information, and use of the gold standard (biopsy) as the reference standard.

Question 2. Are there demographic (e.g., age) and clinical risk factors (e.g., morphologic characteristics of the lesion) that affect the accuracy of the tests considered in Question 1?

Two studies evaluated only patients with palpable breast masses,[206,211] one study evaluated only patients with non-palpable breast masses,[212] and one study evaluated only patients with microcalcifications detected on x-ray mammography.[210] With so few studies reporting on each category, evidence-based conclusions are difficult to support.

None of the studies reported outcomes by patient demographics or any other clinical risk factors that may have affected the accuracy of SMM.

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

None were identified.

Previously Published Systematic Reviews

We identified two decision/cost effectiveness analyses and four systematic reviews of the use of scintimammography to evaluate women after a positive mammography exam. The majority of these analyses were published prior to publication of most of the studies included in the present report. The findings of these reports are briefly summarized in Table 13. The accuracy of scintimammography reported by all four systematic reviews is very similar to our findings—a summary sensitivity of approximately 85%. Most of the systematic reviews reported a slightly higher (approximately 85%) specificity than our finding of approximately 80% specificity, but the confidence interval around our estimate of 80% is wide (imprecise estimate).

Table 13. Other published technology assessments of scintimammography

Study	Methods	Conclusions
Hussain and Buscombe 2006[214]	A meta-analysis of trials of scintimammography for diagnosis of breast cancer was performed. Studies that included more than 100 patients published since 1997 were identified and included.	The overall sensitivity was 85% and the specificity was 84%.
Liu et al. 2005[215]	A systematic review and meta-analysis intended to determine the value of scintimammography in diagnosing primary breast cancer. The authors of the review excluded the bulk of the published literature on the basis of “poor quality.”	The overall sensitivity was 86% and specificity was 80% for diagnosis of breast cancer by scintimammography; these numbers dropped to 69% for diagnosis of non-palpable lesions

Table 13. Other published technology assessments of scintimammography (continued)

Study	Methods	Conclusions
Medical Advisory Secretariat, Ontario Ministry of Health 2003[162]	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 were combined meta-analytically using the method of Littenburg and Moses.[132]	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.
Liberman et al. 2003[216]	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.
Allen et al. 2000[217]	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.

Table 13. Other published technology assessments of scintimammography (continued)

Study	Methods	Conclusions
Hillner 1997[218]	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 follow-up 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.

Conclusion

The estimates of the accuracy of various types of scintimammography, along with a rating of the strength of evidence supporting the accuracy estimate, are summarized in Table 15. We found that the summary sensitivity of scintimammography for all lesions was 84.7% (95% CI: 78.0 to 89.7%) and the summary specificity was 77.0% (64.7 to 85.9%). The data are, however, inconsistent and imprecise, therefore the strength of evidence supporting the estimate of the accuracy of scintimammography is low.

There was insufficient data reported by the studies to conclude much about the impact of patient demographics, clinical risk factors, lesion types, or other various factors on the accuracy of scintimammography.

To aid in interpretation of these findings, we used Bayes' theorem and the summary likelihood ratios for scintimammography used to evaluate lesions in general (see Table 16 and Table 17). These calculations suggest that scintimammography examinations may not be clinically useful for patients who are strongly suspected to have either malignant or benign breast lesions after mammography, because the input provided by the examinations will probably not affect the suspicion of malignancy sufficiently to alter clinical decisions about management of the patient (e.g., recommendations for biopsy vs. follow-up). However, the results of a scintimammography examination conducted on patients suspected of being at a real but low risk of having a malignant lesion after mammography may be useful for both patients and clinicians when making decisions about management.

These calculations suggest that scintimammography examinations may be clinically useful for some women, because the input provided by the scintimammography examination may affect

the suspicion of malignancy, and therefore may alter clinical decisions about management of the patient (e.g., recommendations for biopsy vs. follow-up).

Table 14. Included studies: scintimammography

Study	Scintimammography Methods Studied	Design*	N Patients
Grosso et al. 2009[210]	Planar scintimammography with patient supine and prone	Prospective diagnostic cohort	283
Habib et al. 2009[206]	Double-phase scintimammography with patients supine and prone	Prospective diagnostic cohort	22
Kim et al. 2009[101]	Double-phase scintimammography	Prospective diagnostic cohort	249
Kim et al. 2008[207]	Double-phase scintimammography	Prospective diagnostic cohort	75
Wang et al. 2008[211]	Planar scintimammography	Prospective diagnostic cohort	55
Brem et al. 2007[106]	BSGI	Diagnostic cohort	33
Gommans et al. 2007[212]	Planar scintimammography	Prospective diagnostic cohort	101
Kim et al. 2007[208]	Double-phase scintimammography	Prospective diagnostic cohort	78
Schillaci et al. 2007[213]	Planar scintimammography	Prospective diagnostic cohort	53
Pinero et al. 2006[209]	Double phase scintimammography	Prospective diagnostic cohort	88
Mathieu et al. 2005[205]	SPECT	Retrospective chart review	37

* At times it was difficult to determine if a study was prospective or retrospective, and in those cases we defaulted to simply calling it a "diagnostic cohort study".

Table 15. Scintimammography accuracy

Category	N Studies	N Lesions	Summary Sensitivity (95% CI)	Summary Specificity (95% CI)	Strength of Evidence
Scintimammography, any	11	1,064	84.7% (78.0 to 89.7%)	77.0% (64.7 to 85.9%)	Low
Scintimammography, double-phase planar	5	502	84.6% (73.2 to 91.7%)	72.8% (59.2 to 83.1%)	Low
Scintimammography, planar	4	492	81.5% (74.3 to 87.3%)	82.1% (77.6 to 86.0%)	Low
Scintimammography, BSGI	1	33	88.9% (51.8 to 99.7%)	70.8% (48.9 to 87.4%)	Insufficient
Scintimammography, SPECT	1	37	95.0% (75.1 to 99.9%)	70.6% (44.0 to 89.7%)	Insufficient
Scintimammography, palpable lesions	2	77	85.0% (73.4 to 92.9%)	90.5% (80.4 to 96.4%)	Insufficient
Scintimammography, nonpalpable lesions	1	101	82.2% (67.9 to 92.0%)	92.9% (82.7 to 98.0%)	Insufficient
Scintimammography, microcalcifications	1	283	78.1% (60.0 to 90.7%)	82.5% (77.2 to 87.0%)	Insufficient

Table 16. Clinical interpretations of scintimammography accuracy: benign finding on scintimammography

Pre-test Probability of the Lesion Being Malignant	Post-test Probability of the Lesion Being Malignant Despite a Finding of "Benign" on the SC Exam
	Lesions in General
1%	0% (0 to 0%)
5%	1% (1 to 2%)
10%	2% (2 to 3%)
20%	5% (3 to 6%)
30%	8% (6 to 11%)
40%	12% (9 to 16%)
50%	17% (13 to 22%)
60%	23% (18 to 29%)
70%	32% (25 to 39%)
80%	44% (36 to 52%)
90%	64% (56 to 71%)

Table 17. Clinical interpretations of scintimammography accuracy: malignant finding on scintimammography

Pre-test Probability of the Lesion Being Malignant	Post-test Probability of the Lesion Being Malignant After a Finding of "Malignant" on the Scintimammography Exam
	Lesions in General
1%	4% (2 to 5%)
5%	17% (10 to 26%)
10%	29% (21 to 39%)
20%	48% (37 to 59%)
30%	61% (51 to 71%)
40%	71% (61 to 79%)
50%	79% (71 to 85%)
60%	85% (78 to 89%)
70%	90% (85 to 93%)
80%	94% (91 to 96%)
90%	97% (96 to 98%)

Ultrasound

Background

Technology. 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.[219]

The most commonly used type of ultrasound (conventional, or regular, ultrasound) may be referred to as B-mode gray-scale ultrasound.[220] The contrast resolution of conventional ultrasound depends on the tranducer's frequency. All modern breast imaging applications employ high frequency transducers (7 MHz or higher). Ultrasound images obtained by B-mode gray-scale imaging use differences in the brightness of the image (caused by different ways the ultrasound waves reflect and absorb off tissue interfaces) to examine the internal anatomy of the breast.[220] The echoes of the sound waves are combined to form two-dimensional images of the structure of the interior of the breast. Malignant breast lesions generally appear darker on the images than the surrounding normal tissues, and often have ill-defined borders.[221-223]

One of the known problems with B-mode ultrasound is that interpretation of the images is primarily done by visually inspecting the image. Differences in human perception and utilization of different features for use in diagnosis cause variability in diagnosis and reader-dependent variations in the accuracy of diagnosis.[221] Computer-aided diagnosis (CAD) systems are under development to address this problem. CAD systems are designed to detect patterns in images that are suggestive of malignancy, and to draw the readers' attention to the areas of suspicion.

Compound imaging is a variant on B-mode imaging that is intended to reduce the "noise" in the image and thus improve the image quality.[221] Compound imaging takes multiple ultrasound views from different angles and combines the many views into a single two-dimensional image.

Another variant on B-mode ultrasound is harmonic imaging. B-mode ultrasound waves develop harmonics (multiples of the transmission frequency) as they pass through breast tissue. Digital encoding can be used by computers to construct images from the harmonic frequencies.[220] Harmonic images generally have improved resolution and fewer artifacts than regular B-mode ultrasound.[221]

Doppler ultrasound uses ultrasound to evaluate blood flow through vessels. The speed of blood flow can be evaluated by observing changes in the pitch of the reflected sound waves (the Doppler effect). Malignant masses often exhibit increased rates and amounts of blood flow (increased vascularity) in comparison to benign tissues.[221] Doppler imaging can also be performed with microbubble contrast agents that enhance imaging of blood vessels.[221] Two primary types of Doppler imaging exist, color and power. Color Doppler imaging encodes the mean Doppler frequency shifts at particular locations in various colors, whereas power Doppler imaging encodes the power of the signal (extent of the Doppler effect) at particular locations in various colors.[224] Color Doppler therefore detects the velocity of the blood cells while power Doppler detects the amount of blood present.[224]

Ultrasound tomography uses ultrasound to acquire multiple images of the breast from different angles, and uses a computer to develop a 3D image of the structure of the interior of the

breast. We intended to include ultrasound tomography in this systematic review, but did not identify any studies that met the inclusion criteria.

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.[225] Because taut compression is not required, ultrasound is generally painless. 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.[226]

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.

Findings from 2006 Review

In the 2006 version of this CER, we included eight prospective diagnostic cohort studies of 5,348 breast lesions that were examined by B-mode gray-scale ultrasound.[80,87,227-232] 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 was judged to be moderate, indicating a small chance that publication of new evidence could substantially change these estimates.

Evidence Base

Our literature searches identified 31 diagnostic cohort studies of various types of ultrasound published between 1994 and 2009.[87,105,113,180,209,228-230,233-255] These studies included a total of 8,642 patients with 9,044 breast lesions. The included studies are listed in Table 18 at the end of this subsection on ultrasound, and are described in detail in the Appendices.

Question 1. What is the accuracy of ultrasound for diagnosis of breast cancer in women referred for further evaluation after identification of a possible breast abnormality on routine screening (mammography and/or clinical or self-detection of a palpable lesion)?

B-mode 2D grayscale. Twenty-one studies of 8,199 lesions addressed the accuracy of B-mode 2D grayscale.[105,113,228-230,233-240,242-244,247,249,252,254,255] We combined the reported data in a bivariate binomial model. The summary sensitivity of B-mode 2D grayscale ultrasound for all lesions was 92.4% (95% CI: 84.6 to 96.4%) and the summary specificity was 75.8% (60.8 to 86.3%); there was, however, considerable heterogeneity in the data (I^2 = 99.6%). In our 2006 assessment, we found that for suspicious lesions in general, the sensitivity of B-mode ultrasound examination was 86.1%, considerably lower than the findings of the current update; and we also found in 2006 that the specificity was 66.4%, lower than the 75.8% specificity of the current update. The 2006 version of the report included only a small subset of the evidence base included in the current update.

We conducted meta-regressions to explore the heterogeneity in the data. The variables we investigated were: whether the studies accounted for inter-reader differences; whether the studies blinded image readers to clinical information or not; whether all diagnoses were verified by histopathology or not; whether a prospective design was used; whether the study was funded by a source without a financial interest in the results or not; whether the study enrolled consecutive/ all patients; the geographical location of the study; what type(s) of breast lesions were enrolled in the study; and the prevalence of disease in the study. Two of these variables, whether the studies accounted for inter-reader differences, and whether the studies blinded image readers to clinical information or not, were statistically significantly associated with the results (p = 0.01 and 0.03, respectively). Subgroup analyses found that studies that had blinded image readers to clinical information had a higher sensitivity (96.6% vs. 87.0%) but a much lower specificity (59.5% vs. 85.1%) than unblinded studies. Studies that had accounted for inter-reader differences had a similar sensitivity (93.4% vs. 93.0%) but a much lower specificity (52.7% vs. 90.1%) than studies that did not account for inter-reader differences.

B-mode 2D grayscale, contrast enhanced. Only two studies of a total of 154 breast lesions reported on the accuracy of B-mode 2D grayscale contrast-enhanced ultrasound compared to non-contrast enhanced.[113,234] Contrast enhancement was reported to increase the sensitivity (97.5% vs. 82.7%) but to not dramatically affect the specificity (76.7% vs. 74.0%).

B-mode 3D grayscale. Only one study of 150 breast lesions, Cho et al., reported on the accuracy of B-mode 3D grayscale ultrasound.[239]

Color doppler ultrasound. Six studies of a total of 718 lesions reported on the accuracy of color Doppler ultrasound.[180,245,249,250,252,253] We combined the data reported by these studies in a bivariate binomial model. The summary sensitivity of color doppler ultrasound for all lesions was 88.5% (95% CI: 74.4 to 95.4%) and the summary specificity was 76.4% (95% CI: 61.7 to 86.7%). There was considerable heterogeneity in the data (I^2 = 95.2%). Exclusion of data from two studies that enrolled only patients with palpable lesions[180,252] from the bivariate model did not affect the results. There were too few studies of color Doppler to perform full meta-regressions.

Color doppler ultrasound, contrast-enhanced. Two studies of 146 lesions compared the accuracy of contrast-enhanced color Doppler to non-enhanced color Doppler.[245,250] Contrast-enhancement was found to slightly increase the sensitivity (97.8% vs. 95.7%) and to dramatically increase the specificity (90.7% vs. 55.6%).

Color doppler ultrasound directly compared to B-mode grayscale ultrasound. Two studies directly compared the accuracy of color Doppler ultrasound to B-mode grayscale ultrasound.[249,252] Color Doppler was found to have a higher sensitivity (74.0% vs. 53.1%) but a lower specificity than B-mode ultrasound (84.0% vs. 96.3%).

Power Doppler ultrasound. Nine studies of a total of 614 lesions reported on the accuracy of power Doppler ultrasound.[233,240,241,243,246-248,250,251] We combined the data in a bivariate binomial model. The summary sensitivity of power Doppler ultrasound for all lesions was 70.8% (95% CI: 47.5 to 86.6%) and the summary specificity was 72.6% (95% CI: 59.9 to 82.5%). There was considerable heterogeneity in the data (I^2 = 97.4%).

Power Doppler ultrasound, contrast enhanced. Seven studies of 403 lesions reported on the accuracy of contrast-enhanced power Doppler ultrasound.[240,241,243,247,248,250,251] When we combined the data in a bivariate binomial model, the summary sensitivity for all lesions was 89.3% (95% CI: 52.4 to 98.4%) and the summary specificity was 70.4% (95% CI: 55.4 to 82.0%). There was considerable heterogeneity in the data (I^2 = 87.5%).

Power Doppler ultrasound directly compared to B-mode grayscale ultrasound. Four studies of 248 lesions directly compared the accuracy of power Doppler ultrasound to B-mode grayscale ultrasound.[233,240,243,247] Power Doppler was found to have a lower sensitivity (54.7% vs. 87.7%) but a higher specificity (79.4% vs. 50.7%) than B-mode grayscale ultrasound in these four direct comparisons.

Power Doppler ultrasound directly compared to color Doppler ultrasound. One study directly compared the accuracy of power Doppler, with and without contrast-enhancement, to color Doppler, with and without contrast-enhancement.[250] This study reported that all four methods had a 100% sensitivity, but specificity for contrast-enhanced methods was much higher than for non-contrast-enhanced methods.

Tissue Harmonics. Only one study of 91 lesions reported on the accuracy of tissue harmonic ultrasound methods.[236]

Question 2. Are there demographic (e.g., age) and clinical risk factors (e.g., morphologic characteristics of the lesion) that affect the accuracy of the tests considered in Question 1?

None were identified.

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

None were identified.

Previously Published Systematic Reviews

Flobbe et al. published a decision analysis model comparing different strategies for managing patients presenting with palpable breast masses in 2004.[256] Their decision model was based entirely on data from a single clinical study they previously authored (Flobbe et al.[257]). 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. cannot 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.

Conclusion

The estimates of the accuracy of the various types of ultrasound, along with a rating of the strength of evidence supporting the accuracy estimate, are summarized in Table 19. We intended to evaluate ultrasound tomography, but did not identify any relevant studies that met the inclusion criteria.

Qualitative indirect and direct comparisons between different types of ultrasound imaging were also performed. B-mode grayscale ultrasound was found to be more sensitive than either power or color Doppler imaging (conclusion supported by a Low strength of evidence). Color Doppler imaging was more accurate (both more sensitive and more specific) than power Doppler imaging (conclusion supported by a Low strength of evidence). In general, contrast-enhancement was found to improve the accuracy of all types of ultrasound imaging (conclusion supported by a Low strength of evidence). However, in actual clinical practice, it is unlikely that Doppler imaging would be used in isolation; most likely Doppler imaging and B-mode imaging would be performed by the same operator during the same procedure, and the image reader would incorporate information from all of the types of imaging into the diagnosis. There is insufficient data available to reach conclusions about the accuracy of combined ultrasound modalities.

We were unable to identify any patient demographics, clinical risk factors, or other factors that affected the accuracy of the various types of ultrasound imaging. Most of the studies did not enroll women found to have obvious cysts, and therefore our findings do not apply to women who clearly have cystic lesions on ultrasound imaging.

To aid in interpretation of these findings, we used Bayes' theorem and the summary likelihood ratios for the three primary types of ultrasound imaging (see Table 20 and Table 21). These calculations suggest that ultrasound examinations may not be clinically useful for patients who are strongly suspected to have either malignant or benign breast lesions after mammography, because the input provided by the ultrasound examinations will probably not affect the suspicion of malignancy sufficiently to alter clinical decisions about management of the patient (e.g., recommendations for biopsy vs. follow-up). However, the results of a grayscale B-mode ultrasound examination conducted on patients suspected of being at a real but low risk of having a malignant lesion after mammography may be useful for both patients and clinicians when making decisions about management.

These calculations suggest that ultrasound examinations may be clinically useful for some women, because the input provided by the examination may affect the suspicion of malignancy,

and therefore may alter clinical decisions about management of the patient (e.g., recommendations for biopsy vs. follow-up).

Table 18. Included studies: ultrasound

Study	US Methods Studied	Design*	N Patients
Gokalp et al. 2009[233]	B-mode 2D grayscale, power Doppler, and combination of both methods	Prospective diagnostic cohort	49
Vassiou et al. 2009[105]	B-mode 2D grayscale	Prospective diagnostic cohort	69
Liu et al. 2008[234]	B-mode 2D grayscale, with and without contrast (with Sono Vue (Bracco, Italy)), and combination of both methods	Diagnostic cohort study	108
Vade et al. 2008[235]	B-mode 2D grayscale	Retrospective chart review	20
Cha et al. 2007[236]	B-mode 2D grayscale and tissue harmonic imaging	Prospective diagnostic cohort	88
Chala et al. 2007[237]	B-mode 2D grayscale	Retrospective chart review	203
Zhi et al. 2007[238]	B-mode 2D grayscale	Diagnostic cohort study	232
Cho et al. 2006[239]	B-mode 2D and 3D grayscale	Prospective diagnostic cohort	141
Pinero et al. 2006[209]	Combination power Doppler and color Doppler using a contrast agent (Levovist [Schering AG, Berlin, Germany])	Prospective diagnostic cohort	88
Ricci et al. 2006[113]	B-mode grayscale with and without contrast (with Sono Vue [Bracco, Italy]); also compared US to MRI	Prospective diagnostic cohort	48
Forsberg et al. 2004[240]	B-mode 2D grayscale and power Doppler, with and without contrast (Levovist or Optison)	Diagnostic cohort study	55
Meyberg-Solomayer et al. 2004[228]	B-mode 2D gray-scale	Prospective diagnostic cohort	65
Ozdemir et al. 2004[241]	Power Doppler, with or without contrast (Levovist)	Prospective diagnostic cohort	80
Chen et al. 2003[242]	B-mode 2D gray scale	Prospective diagnostic cohort	32
Kook and Kwag 2003[243]	B-mode US and power Doppler, with and without contrast (Levovist)	Prospective diagnostic cohort	36
Marini et al. 2003[244]	B-mode 2D grayscale	Diagnostic cohort study	238

Table18. Included studies: ultrasound (continued)

Study	US Methods Studied	Design*	N Patients
Caruso et al. 2002[245]	Color Doppler with and without contrast (Levovist)	Prospective diagnostic cohort	36
Koukouraki et al. 2001[180]	Color Doppler	Prospective diagnostic cohort	116
Malich et al. 2001[87]	Combination of B-mode, power Doppler, and color Doppler; also compared US to MRI	Diagnostic cohort study	94
Milz et al. 2001[246]	Power Doppler	Prospective diagnostic cohort	102
Reinikainen et al. 2001[247]	B-mode US and power Doppler, with and without contrast (Levovist)	Prospective diagnostic cohort	63
Moon et al. 2000[248]	Power Doppler, with and without contrast (Levovist)	Prospective diagnostic cohort	69
Blohmer et al. 1999[249]	B-mode 2D gray-scale and color Doppler	Prospective diagnostic cohort	200
Chao et al. 1999[229]	B-mode 2D gray-scale	Prospective diagnostic cohort	3,050
Schroeder et al. 1999[250]	Power and color Doppler, with and without contrast (Levovist)	Prospective diagnostic cohort	92
Albrecht et al. 1998[251]	Power Doppler, with or without contrast (EchoGen)	Prospective diagnostic cohort	20
Wilkens et al. 1998[252]	B-mode 2D gray-scale and color Doppler	Diagnostic cohort study	53
Buadu et al. 1997[253]	Color Doppler	Diagnostic cohort study	114
Stavros et al. 1995[254]	B-mode 2D gray-scale	Prospective diagnostic cohort	622
Ciatto et al. 1994[255]	B-mode 2D gray scale	Prospective diagnostic cohort	2,079
Perre et al. 1994[230]	B-mode 2D gray-scale	Prospective diagnostic cohort	380

* At times it was difficult to determine if a study was prospective or retrospective, and in those cases we defaulted to simply calling it a "diagnostic cohort study".

Table 19. Ultrasound accuracy: accuracy of different types of ultrasound

Type of Ultrasound	N Studies	N Lesions	Summary Sensitivity (95% CI)	Summary Specificity (95% CI)	Strength of Evidence
B-mode grayscale 2D	21	8,199	92.4% (84.6 to 96.4%)	75.8% (60.8 to 86.3%)	Low
B-mode grayscale 2D contrast enhanced	2	154	97.5% (91.4 to 99.7%)	76.7% (65.4 to 85.8%)	Low
B-mode grayscale 3D	1	150	98.3% (91.1 to 100.0%)	70.0% (59.4 to 79.2%)	Insufficient
Color Doppler	6	718	88.5% (74.4 to 95.4%)	76.4% (61.7 to 86.7%)	Low
Color Doppler contrast enhanced	2	146	97.8% (92.4 to 99.7%)	90.7% (79.7 to 96.9%)	Low
Power Doppler	9	614	70.8% (47.5 to 86.6%)	72.6% (59.9 to 82.5%)	Low
Power Doppler contrast enhanced	7	403	89.3% (52.4 to 98.4%)	70.4% (55.4 to 82.0%)	Low
Tissue harmonics	1	91	96.7% (82.8 to 99.9%)	62.3% (49.0 to 74.4%)	Insufficient

Table 20. Clinical interpretations of ultrasound accuracy: benign finding on ultrasound

Pre-test Probability of the Lesion Being Malignant	Post-test Probability of the Lesion Being Malignant Despite a Finding of "Benign" on the Ultrasound Exam		
	B-mode Grayscale 2D Ultrasound	Power Doppler Ultrasound	Color Doppler Ultrasound
1%	0% (0 to 0%)	0% (0 to 1%)	0% (0 to 0%)
5%	1% (0 to 1%)	2% (1 to 4%)	1% (0 to 2%)
10%	1% (1 to 2%)	4% (2 to 8%)	2% (1 to 3%)
20%	2% (1 to 5%)	9% (5 to 16%)	4% (2 to 7%)
30%	4% (2 to 8%)	15% (9 to 24%)	6% (3 to 12%)
40%	6% (3 to 12%)	21% (13 to 33%)	9% (5 to 17%)
50%	9% (5 to 17%)	29% (18 to 43%)	13% (7 to 24%)
60%	13% (7 to 23%)	38% (25 to 53%)	18% (10 to 32%)
70%	19% (10 to 32%)	48% (34 to 63%)	26% (14 to 42%)
80%	29% (16 to 45%)	62% (47 to 75%)	38% (22 to 56%)
90%	47% (31 to 65%)	78% (66 to 87%)	57% (39 to 74%)

Table 21. Clinical interpretations of ultrasound accuracy: malignant finding on ultrasound

Pre-test Probability of the Lesion Being Malignant	Post-test Probability of the Lesion Being Malignant After a Finding of "Malignant" on the Ultrasound Exam		
	B-mode Grayscale 2D Ultrasound	Power Doppler Ultrasound	Color Doppler Ultrasound
1%	4% (2 to 7%)	3% (2 to 4%)	4% (2 to 6%)
5%	17% (5 to 11%)	12% (9 to 16%)	17% (11 to 24%)
10%	30% (20 to 42%)	22% (17 to 29%)	29% (21 to 40%)
20%	49% (36 to 62%)	39% (32 to 47%)	48% (38 to 62%)
30%	62% (49 to 73%)	53% (45 to 61%)	62% (51 to 72%)
40%	72% (60 to 81%)	63% (56 to 71%)	71% (62 to 80%)
50%	79% (69 to 86%)	72% (66 to 78%)	79% (71 to 86%)
60%	85% (77 to 91%)	80% (74 to 84%)	85% (78 to 90%)
70%	90% (84 to 94%)	86% (82 to 89%)	90% (85 to 93%)
80%	94% (90 to 96%)	91% (88 to 94%)	94% (91 to 96%)
90%	97% (95 to 98%)	96% (94 to 97%)	97% (96 to 98%)

Comparative Accuracy

We identified three studies that directly compared PET and MRI[90,93,94] and one study that directly compared PET/CT and MRI.[103] There was no consistent pattern of relative accuracy across the three studies that directly compared PET and MRI. Imbracio et al. directly compared the diagnostic accuracy of PET/CT and MRI in the same set of patients.[103] MRI was more sensitive but less specific than PET/CT in diagnosing breast lesions in this study.[103] An indirect comparison of the summary accuracy estimates from the other sections of this report suggests that MRI is more sensitive than PET, but the two imaging methods have approximately the same specificity.

We identified two studies that directly compared B-mode grayscale ultrasound to MRI,[105,113] and one study that compared a combination of several Doppler ultrasound methods to MRI.[87] All three studies found that MRI was more sensitive than ultrasound for diagnosing breast lesions (results for specificity were inconsistent across studies). An indirect comparison of the summary accuracy estimates from the other sections of this report suggest that the two imaging methods are of approximately equal accuracy.

We identified one study that directly compared scintimammography to a combination of several Doppler ultrasound methods[209] that found the two methods were approximately equally accurate, with a slightly higher sensitivity for scintimammography. Indirect comparisons of the summary accuracy estimates from the other sections of this report suggest that ultrasound may be slightly more sensitive than scintimammography.

We identified one study[101] that directly compared scintimammography and MRI, and found MRI to be more sensitive but less specific than scintimammography. An indirect comparison of the summary accuracy estimates from the other sections of this report concurs with the direct comparison conclusion. We also identified one study[106] that directly compared MRI to a variant of scintimammography (BSGI) with similar findings (MRI more sensitive but less specific than BSGI).

The summary estimates of accuracy of each modality are shown in Table 22. The data suggest that ultrasound and MRI are more accurate than PET or scintimammography for evaluation of suspicious breast lesions. Because the evidence supporting these comparisons is, for the most part, indirect in nature, and not reported in sufficient detail to support statistical testing, we have refrained from drawing any solid evidence-based conclusions about comparisons across technologies; at best, the evidence suggests that certain technologies may be more accurate than others for evaluation of breast lesions.

Table 22. Summary accuracy results

Technology	N Studies	N Lesions	Summary Sensitivity (95% CI)	Summary Specificity (95% CI)	Post-test Probability of "Malignancy"[a]	Strength of Evidence
B-mode grayscale 2D	21	8,199	92.4% (84.6 to 96.4%)	75.8% (60.8 to 86.3%)	2% (1 to 5%)	Low
MRI	41	3,882	91.7% (88.5 to 94.1%)	77.5% (71.0 to 82.9%)	3% (2 to 4%)	Moderate (sensitivity)/ Low (specificity)
Scintimammography	11	1,064	84.7% (78.0 to 89.7%)	77.0% (64.7 to 85.9%)	5% (3 to 6%)	Low
PET	7	403	83.0% (73.0 to 89.0%)	74.0% (58.0 to 86%)	6% (4 to 8%)	Low

[a] post-test probability of a lesion being "malignant" after a benign finding on the test for a typical woman with an estimated 20% chance of having a malignant lesion.

Chapter 4. Discussion

After identification of a possible abnormality on screening mammography or physical examination, women typically undergo additional imaging studies (diagnostic mammography) and a physical examination. If these studies suggest the abnormality may be malignant, a biopsy of the suspicious area may be recommended. This evidence review focuses on the non-invasive imaging studies conducted after the discovery of a possible abnormality on screening mammography or physical examination- studies intended to guide patient management decisions. In other words, these studies are not intended to provide a final diagnosis as to the nature of the breast lesion; rather, they are intended to provide additional information about the nature of the lesion such that women can be appropriately triaged into "biopsy," "watchful waiting," or "return to normal screening intervals" care pathways.

According to the American College of Radiology, the threshold of suspicion at which management of women changes is 2%.[258] After recall and workup, women with a suspicion of malignancy greater than 2% are generally advised to undergo tissue sampling of some kind (biopsy), and women with a lower suspicion of malignancy are triaged into imaging pathways. We used the 2% threshold to explore the clinical usefulness of the various non-invasive imaging technologies as add-ons to the current standard of care, namely, if a woman was recalled for evaluation after a screening mammography, and received standard of care workup vs. standard of care workup plus the non-invasive imaging technology, would the use of the non-invasive imaging technology be likely to alter the recommendations for care after the workup?

For all of the technologies evaluated in this assessment, only women with a low suspicion of malignancy on standard of care workup might be expected to experience a change in management decisions as a result of additional non-invasive imaging. A woman with an approximate 2 to 20% suspicion of malignancy who has benign findings on one of the non-invasive imaging technologies could have her suspicion of malignancy drop below the 2% threshold, and therefore she might be assigned to imaging management rather than tissue sampling management. Clinicians can estimate the risk of malignancy by using patient age, family and personal history details, details of the mammographic images, and results of physical examination.[259,260] For example, Wiratkapun et al. recently reported that while 20% of a large cohort of women classified as BI-RADS 4 after diagnostic mammography were subsequently diagnosed with breast cancer, by using risk factors and details of the mammographic images the women could be classified into sub-categories that had cancer prevalences that ranged from as low as 9% to as high as 57%.[259]

Therefore, if the 2% threshold is chosen, the use of non-invasive imaging in addition to standard workup may be clinically useful for women with a low (2 to 20%) suspicion of malignancy. When choosing which non-invasive imaging technology to use for this purpose, diagnostic B-mode grayscale ultrasound and MRI appear to more accurate than PET, scintimammography, or the other types of ultrasound (Doppler) that were evaluated in this comparative effectiveness review.

Non-invasive imaging appears to be an acceptable option for many women. Liang et al. invited a series of women referred for breast biopsy to undergo an additional mammographic exam, MRI, and scintimammography before the biopsy.[261] The women reported that MRI and scintimammography were much more comfortable than mammography, and that they would rather have additional non-invasive tests, even if they had to pay extra money out of pocket,

instead of proceeding to immediate biopsy (assuming the results of the non-invasive tests were very accurate).

Changes since 2006

This CER is an update of a CER finalized in 2006. The updated results are, in general, very similar to the findings of the 2006 report. For MRI, in 2006 we found that the sensitivity was 92.5% and the specificity was 75.5%; the updated evidence base supported estimates of 91.7% sensitivity and 77.5% specificity. In both reports, MRI was found to be less sensitive (approximately 85%) for evaluation of microcalcifications than for evaluation of lesions in general. For PET, in 2006 we found that the sensitivity was 82.2% and the specificity was 78.3%; the updated evidence base supported estimates of 83.0% sensitivity and 74.0% specificity. In the updated report we attempted to evaluate the accuracy of PET/CT, but only one study that met the inclusion criteria was identified.

For scintimammography, the updated evidence base identified a sensitivity of 84.7%, much higher than the sensitivity estimate from 2006 of 68.7%. Specificity was estimated at 84.8% in 2006, and at 77.0% in the update; however, the confidence intervals around the updated estimate of specificity are wide. It is possible that improvements in the technology in the last few years improved the sensitivity of the technique.

For ultrasound, in 2006 we only evaluated a relatively small subset of studies of B-mode grayscale ultrasound, and estimated a sensitivity of 86.1% and a specificity of 66.4%. The update included a significantly expanded evidence base on B-mode grayscale ultrasound, and identified a sensitivity of 92.4% and specificity of 75.8%. In the update we included numerous other types of ultrasound, including power and color Doppler ultrasound, that were not studied in the 2006 report.

Limitations of the Evidence Base

The majority of conclusions were rated as supported by "Low" strength of evidence. The evidence bases were rated as Low rather than Moderate or High due primarily to the heterogeneity of the results (inconsistency). All of the evidence bases were found to contain significant heterogeneity, and exploratory meta-regressions did not identify satisfactory explanations for the heterogeneity.

Another limitation of the evidence base is that most of the studies included only patients who had been referred for biopsy or surgery. Therefore the patient population under study does not contain a good representation of patients thought to be at sufficiently low risk of malignancy that additional imaging would be considered rather than immediate biopsy. In addition, little information was reported about different patient subgroups, making it difficult to address Key Questions 2 and 3.

Applicability

We used inclusion criteria intended to restrict the evidence base to only those studies that included the population of interest: women of average baseline risk after discovery of a suspicious lesion on routine screening. However, the patient populations studied had much higher prevalences of cancers than would be expected if the populations were actually representative of the patient population of interest. The prevalence of cancers in the general population sent for breast biopsy (in the USA) has been reported to be around 23%.[17] The

population of interest includes not only those women who will be referred for biopsy, but should also include women who will be referred for short interval follow-up, and therefore the expected prevalence of cancers in the population of interest should be lower than 20%. However, the prevalence of cancers in the included studies was 25.8% for ultrasound, 54.5% for MRI, 56% for scintimammography, and 75.9% for PET. One reason for the elevated prevalence is that the studies generally attempted to use the "gold standard" reference to verify diagnoses (histopathology), and therefore many of the studies only enrolled patients who subsequently underwent biopsy or surgery. An additional possible reason for the elevated prevalence of disease is the fact that many of the studies were conducted in non-USA locations, where the prevalence of cancers in populations sent for biopsy has been reported to be 60 to 70%.[262]

The patient populations studied are therefore not truly representative of the patient population of interest. It is possible that the accuracy estimates we derived from these studies do not apply to women thought to be at sufficiently low risk of malignancy that additional imaging would be considered rather than immediate biopsy.

Possible Impact of Key Assumptions on the Conclusions

The key assumption made was that the "reference standard," a combination of biopsy, open surgery, and patient followup, was 100% accurate. Open surgery has been reported to have a false-negative rate of approximately 1 to 2%.[263] Biopsy and patient followup have error rates higher than open surgery. Therefore some of the reference standard diagnoses were almost certainly incorrect. However, the errors should consist of a low rate of both false-negatives and false-positives, which should not systematically bias the results in any one direction. It seems unlikely that our estimates of diagnostic accuracy are significantly different from the "true" accuracy solely due to errors made by the reference standard diagnoses.

Future Research Needed

The strength of the evidence supporting the conclusions in this assessment was in general rated as "low" primarly due to imprecise estimates of accuracy (wide confidence intervals) and/or inconsistencies across studies (heterogeneity). While further studies on the diagnostic accuracy of the non-invasive technologies evaluated are unlikely to substantially change the conclusions, the publication of additional diagnostic accuracy studies may increase the precision of the estimates of accuracy, and provide enough additional information to allow productive exploration into the causes of the heterogeneity.

One primary shortcoming in the current evidence base is the lack of evidence for specific subgroups of lesion types. For example, while we were able to determine the accuracy of MRI for patients presenting with microcalcifications, due to lack of evidence we were unable to determine the accuracy of PET, ultrasound, or scintimammography for patients presenting with microcalcifications. We had also hoped to be able to study the impact of variations in MRI methodology on the accuracy, but the many variations of imaging methods in use and the inconsistency in reporting across studies precluded any such analysis. Also, due to lack of evidence we were unable to determine the impact of patient characteristics such as age on the accuracy of the various imaging methods. Future diagnostic accuracy studies that report data for specific subgroups of patients or directly compare different imaging methods would be helpful in addressing these unanswered questions. Studies of new technologies, and improvements in current technologies, are of course essential. For example, the use of computer-aided diagnosis software (CADx) to help interpret MRI images is a technology that is rapidly diffusing, yet there is little clinical evidence available on the impact of CADx on MRI accuracy.

Future research efforts should also be turned to studies that report the impact of the use of non-invasive imaging on management decisions and patient-oriented outcomes. The ideal design for these studies would be a randomized controlled trial in which one arm had non-invasive imaging and one did not; the non-invasive imagine results were used in management decisions; and the patients are followed up for long periods of time to determine the downstream impact of the use of non-invasive imaging on survival and quality of life. Admittedly such studies may be logistically difficult to conduct. The diagnostic thresholds that trigger invasive, aggressive diagnostic testing and treatment should also be studied in the context of the addition of non-invasive imaging to standard protocols. Current standard of care results in large numbers of healthy women undergoing invasive diagnostic procedures, and many women may be undergoing unnecessarily aggressive treatment for small early-stage breast cancers.[16,17,264,265] The diagnostic thresholds in current use are intended to reduce the rate of missed cancers to its absolute minimum, which by necessity causes a severe loss of specificity; and are also intended to partially compensate for diagnostic inaccuracy of tests in current use. The addition of non-invasive imaging to standard protocols may allow for the use of different diagnostic thresholds. It is possible that women at a low risk of malignancy may benefit from monitoring with non-invasive imaging rather than going straight to biopsy. The thresholds used in clinical practice to trigger implementation of aggressive diagnostic testing and treatment should be based on solid evidence derived from controlled trials and modelling studies.

References and Included Studies

(Note: There is a separate set of references at the end of Appendix E whose reference numbers are different from those in the text of the report.)

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. American Cancer Society (ACS). Cancer facts & figures 2009. Atlanta (GA): American Cancer Society (ACS); 2009. 72 p. Also available: http://www.cancer.org/downloads/STT/500809web.pdf.

3. Surveillance Epidemiology and End Results. SEER stat fact sheets - cancer of the breast. [internet]. Bethesda (MD): National Cancer Institute (NCI); [accessed 2007 Feb 5]. [3 p]. Available: http://seer.cancer.gov/statfacts/html/breast_print.html.

4. Feuer EJ, Wun LM. DEVCAN: probability of developing or dying of cancer software, version 4.1. [internet]. Bethesda (MD): National Cancer Institute (NCI); 1999 [accessed 2002 May 28]. [1 p]. Available: http://canques.seer.cancer.gov/cgi-bin/cq_submit?dir-devcan1999&db=1&rpt=TAB&sel=.

5. Harvey JA. Unusual breast cancers: useful clues to expanding the differential diagnosis. Radiology 2007 Mar;242(3):683-94.

6. Levine P, Simsir A, Cangiarella J. Management issues in breast lesions diagnosed by fine-needle aspiration and percutaneous core breast biopsy. Am J Clin Pathol 2006 Jun;125 Suppl:S124-34.

7. American Cancer Society (ACS). Cancer facts & figures 2008. Atlanta (GA): American Cancer Society (ACS); 2008. 72 p. Also available: http://www.cancer.org/downloads/STT/2008CAFFfinalsecured.pdf.

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. U.S. Preventive Services Task Force. Screening for breast cancer: U.S. Preventive Services Task Force recommendation statement. Ann Intern Med 2009 Nov 17;151(10):716-26, W-236.

11. Nelson HD, Tyne K, Naik A, Bougatsos C, Chan B, Nygren P, Humphrey L. Screening for Breast Cancer: Systematic Evidence Review Update for the US Preventive Services Task Force. Ann Intern Med 2009 Nov.

12. Gotzsche PC, Nielsen M. Screening for breast cancer with mammography. In: Cochrane Database of Systematic Reviews [internet]. Issue 4. Hoboken (NJ): John Wiley & Sons, Ltd.; 2009 [accessed 2010 Oct 8]. [Art. No.: CD001877]. Available: DOI: 10.1002/14651858.CD001877.pub3.

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

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

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

16. Elmore JG, Barton MB, Moceri VM, et al. Ten-year risk of false positive screening mammograms and clinical breast examinations. N Engl J Med 1998 Apr 16;338(16):1089-96.

17. Lacquement MA, Mitchell D, Hollingsworth AB. Positive predictive value of the Breast Imaging Reporting and Data System. J Am Coll Surg 1999 Jul;189(1):34-40.

18. Fryback DG, Thornbury JR. The efficacy of diagnostic imaging. Med Decis Making 1991 Apr-Jun;11(2):88-94.

19. Bruening W, Launders J, Pinkney N, et al. Effectiveness of noninvasive diagnostic tests for breast abnormalities (Prepared by ECRI Evidence-based Practice center under Contract No. 290-02-0019). Rockville (MD): Agency for Healthcare Research and Quality (AHRQ); 2006. 130 p. (Comparative effectiveness review; no. 2). Also available: http://effectivehealthcare.ahrq.gov/repFiles/BrCADx%20Final%20Report.pdf.

20. Counsell C. Formulating questions and locating primary studies for inclusion in systematic reviews. Ann Intern Med 1997 Sep 1;127(5):380-7.

21. Lijmer JG, Mol BW, Heisterkamp S, et al. Empirical evidence of design-related bias in studies of diagnostic tests. JAMA 1999 Sep 15;282(11):1061-6.

22. 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.

23. Kolko DJ, Anderson L, Campbell M. Sensory preference and overselective responding in autistic children. J Autism Dev Disord 1980 Sep;10(3):259-71.

24. Vetrani A, Fulciniti F, Di Benedetto G, et al. 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.

25. 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.

26. Ljung BM, Drejet A, Chiampi N, et al. Diagnostic accuracy of fine-needle aspiration biopsy is determined by physician training in sampling technique. Cancer 2001 Aug 25;93(4):263-8.

27. Moher D, Pham B, Klassen TP, et al. What contributions do languages other than English make on the results of meta-analyses? J Clin Epidemiol 2000 Sep;53(9):964-72.

28. Juni P, Holenstein F, Sterne J, et al. Direction and impact of language bias in meta-analyses of controlled trials: empirical study. Int J Epidemiol 2002 Feb;31(1):115-23.

29. 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.

30. 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.

31. Dundar Y, Dodd S, WIlliamson P, et al. Case study of the comparison of data from conference abstracts and full-text articles in health technology assessment of rapidly evolving technologies: does it make a difference? Int J Technol Assess Health Care 2006 Jul;22(3):288-94.

32. 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.

33. Scherer RW, Langenberg P. Full publication of results initially presented in abstracts. In: Cochrane Library [Cochrane methodology review]. Issue 2. Oxford: Update Software; 2001 [accessed 2001 Apr 23]. [35 p]. Available: http://www.cochrane.org/index.htm.

34. 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 Jun-Jul;20(6):1173-7.

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

36. Whiting P, Rutjes AW, Reitsma JB, et al. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. BMC Med Res Methodol 2003 Nov 10;3(1):25. Also available: http://www.biomedcentral.com/content/pdf/1471-2288-3-25.pdf.

37. Owens DK, Lohr KN, Atkins D, et al. Grading the strength of a body of evidence when comparing medical interventions-Agency for Healthcare Research and Quality and the Effective Health Care Program. J Clin Epidemiol 2010 May;63(5):513-23.

38. Owens D, Lohr K, Atkins D, et al. Grading the strength of a body of evidence when comparing medical interventions. In: Methods guide for comparative effectiveness reviews. Rockville (MD): Agency for Healthcare Research and Quality (AHRQ); 2009 Jul. p. 19. Also available: http://effectivehealthcare.ahrq.gov/ehc/products/122/328/2009_0805_grading.pdf.

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-analyses. BMJ 2003 Sep 6;327(7414):557-60.

41. Harbord RM, Deeks JJ, Egger M, et al. A unification of models for meta-analysis of diagnostic accuracy studies. Biostatistics 2007 Apr;8(2):239-51.

42. STATA statistics/data analysis. MP parallel edition. College Station (TX): StataCorp; 1984-2007. Single user Stata for Windows. Also available: http://www.stata.com.

43. Zamora J, Abraira V, Muriel A, et al. Meta-DiSc: a software for meta-analysis of test accuracy data. BMC Med Res Methodol 2006;6:31.

44. Deeks J, Macaskill P, Irwig L. Detecting publication bias in systematic reviews of diagnostic test accuracy [abstract O-009]. In: Cochrane Collaboration [database online]. Oxford (UK): Cochrane Collaboration; 2004 [accessed 2005 May 25]. [1 p]. Available: http://www.cochrane.org/colloquia/abstracts/ottawa/O-009.htm.

45. Deeks J, Macaskill P, Irwig L. By how much does publication bias affect the results of systematic reviews of diagnostic test accuracy? (abstract O-010). In: Cochrane Collaboration [database online]. Oxford (UK): Cochrane Collaboration; 2004 [accessed 2005 May 25]. [2 p]. Available: http://www.cochrane.org/colloquia/abstracts/ottawa/O-010.htm.

46. 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.

47. 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.

48. 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.

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

50. 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.

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

52. 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.

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

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

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

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

57. MedWatch. Information on Gadolinium-Containing Contrast Agents. [internet]. Rockville (MD): U.S. Food and Drug Administration; 2009 Aug 13 [accessed 2010 Mar 17]. Available: http://www.fda.gov/Drugs/DrugSafety/PostmarketDrugSafetyInformationforPatientsandProviders/ucm142882.htm.

58. 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.

59. 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.

60. National Cancer Data Base (NCDB) [database online]. Chicago (IL): American College of Surgeons (ACS); 1996-1998. [accessed 1999 Mar 5]. Relative survival of breast cancer patients by AJCC stage of disease at dignosis, 1985-1990 cases [table]. [3 p]. Available: http://www.facs.org/about_college/acsdept/cancer_dept/programs/ncdb/breastcancer4.html.

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

62. 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.

63. Morris EA, Bassett LW, Berg WA, Birdwell RL, Brenner RJ, Comstock CE, DeBruhl ND, D'Orsi CJ, Evans WP III, Everson LI, Harvey JA, Herman CR, Jong RA, Kaplan SS, Liberman L, Mahoney MC, Mendelson EB, Parikh JR, Rabinovitch RA, et al. ACR practice guideline for the performance of contrast-enhanced magnetic resonance imaging (MRI) of the breast. [online publication]. Reston (VA): American College of Radiology (ACR); 2008. 7 p. Also available: http://www.acr.org/SecondaryMainMenuCategories/quality_safety/guidelines/breast/mri_breast.aspx.

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

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

66. 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.

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

68. 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.

69. 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.

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

71. 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.

72. 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.

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

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

75. 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.

76. Centers for Disease Control and Prevention (CDC). Nephrogenic fibrosing dermopathy associated with exposure to gadolinium-containing contrast agents--St. Louis, Missouri, 2002-2006. MMWR Morb Mortal Wkly Rep 2007 Feb 23;56(7):137-41.

77. Marckmann P, Skov L, Rossen K, et al. Nephrogenic systemic fibrosis: suspected causative role of gadodiamide used for contrast-enhanced magnetic resonance imaging. J Am Soc Nephrol 2006 Sep;17(9):2359-62.

78. Grobner T. Gadolinium--a specific trigger for the development of nephrogenic fibrosing dermopathy and nephrogenic systemic fibrosis. Nephrol Dial Transplant 2006 Apr;21(4):1104-8.

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

80. 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 Nov-Dec;9(6):232-40.

81. 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.

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

83. 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.

84. 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.

85. 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.

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

87. 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.

88. 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.

89. 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.

90. Brix G, Henze M, 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.

91. 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.

92. 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 2002 Oct;104(4):262-72.

93. 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.

94. 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.

95. 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.

96. 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.

97. 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.

98. Akita A, Tanimoto A, Jinno H, et al. The clinical value of bilateral breast MR imaging: is it worth performing on patients showing suspicious microcalcifications on mammography? Eur Radiol 2009 Sep;19(9):2089-96.

99. Baltzer PA, Freiberg C, Beger S, et al. Clinical MR-mammography: are computer-assisted methods superior to visual or manual measurements for curve type analysis? A systematic approach. Acad Radiol 2009 Sep;16(9):1070-6.

100. Hara M, Watanabe T, Okumura A, et al. Angle between 1 and 4 min gives the most significant difference in time-intensity curves between benign disease and breast cancer: analysis of dynamic magnetic resonance imaging in 103 patients with breast lesions. Clin Imaging 2009 Sep-Oct;33(5):335-42.

101. Kim IJ, Kim YK, Kim SJ. Detection and Prediction of Breast Cancer Using Double Phase Tc-99m MIBI Scintimammography in Comparison with MRI. Onkologie 2009 Oct;32(10):556-60.

102. Lo GG, Ai V, Chan JK, et al. Diffusion-weighted magnetic resonance imaging of breast lesions: first experiences at 3 T. J Comput Assist Tomogr 2009 Jan-Feb;33(1):63-9.

103. Imbriaco M, Caprio MG, Limite G, et al. Dual-time-point 18F-FDG PET/CT versus dynamic breast MRI of suspicious breast lesions. AJR Am J Roentgenol 2008 Nov;191(5):1323-30.

104. Pediconi F, Catalano C, Padula S, et al. Contrast-enhanced MR mammography: improved lesion detection and differentiation with gadobenate dimeglumine. AJR Am J Roentgenol 2008 Nov;191(5):1339-46.

105. Vassiou K, Kanavou T, Vlychou M, et al. Characterization of breast lesions with CE-MR multimodal morphological and kinetic analysis: comparison with conventional mammography and high-resolution ultrasound. Eur J Radiol 2009 Apr;70(1):69-76.

106. Brem RF, Petrovitch I, et al. Breast-specific gamma imaging with 99mTc-Sestamibi and magnetic resonance imaging in the diagnosis of breast cancer--a comparative study. Breast J 2007 Sep-Oct;13(5):465-9.

107. Cilotti A, Iacconi C, Marini C, et al. Contrast-enhanced MR imaging in patients with BI-RADS 3-5 microcalcifications. Radiol Med 2007 Mar;112(2):272-86.

108. Pediconi F, Catalano C, Padula S, et al. Contrast-enhanced magnetic resonance mammography: does it affect surgical decision-making in patients with breast cancer? Breast Cancer Res Treat 2007 Nov;106(1):65-74.

109. Zhu J, Kurihara Y, Kanemaki Y, et al. Diagnostic accuracy of high-resolution MRI using a microscopy coil for patients with presumed DCIS following mammography screening. J Magn Reson Imaging 2007 Jan;25(1):96-103.

110. Bazzocchi M, Zuiani C, Panizza P, et al. Contrast-enhanced breast MRI in patients with suspicious microcalcifications on mammography: results of a multicenter trial. AJR Am J Roentgenol 2006 Jun;186(6):1723-32.

111. Gokalp G, Topal U. MR imaging in probably benign lesions (BI-RADS category 3) of the breast. Eur J Radiol 2006 Mar;57(3):436-44.

112. Kneeshaw PJ, Lowry M, Manton D, et al. Differentiation of benign from malignant breast disease associated with screening detected microcalcifications using dynamic contrast enhanced magnetic resonance imaging. Breast 2006 Feb;15(1):29-38.

113. Ricci P, Cantisani V, Ballesio L, et al. Benign and malignant breast lesions: efficacy of real time contrast-enhanced ultrasound vs. magnetic resonance imaging. Ultraschall Med 2007 Feb;28(1):57-62.

114. Pediconi F, Catalano C, Venditti F, et al. Color-coded automated signal intensity curves for detection and characterization of breast lesions: preliminary evaluation of a new software package for integrated magnetic resonance-based breast imaging. Invest Radiol 2005 Jul;40(7):448-57.

115. Pediconi F, Catalano C, Occhiato R, et al. Breast lesion detection and characterization at contrast-enhanced MR mammography: gadobenate dimeglumine versus gadopentetate dimeglumine. Radiology 2005 Oct;237(1):45-56.

116. Wiener JI, Schilling KJ, Adami C, et al. Assessment of suspected breast cancer by MRI: a prospective clinical trial using a combined kinetic and morphologic analysis. AJR Am J Roentgenol 2005 Mar;184(3):878-86.

117. Daldrup-Link HE, Kaiser A, Helbich T, et al. Macromolecular contrast medium (feruglose) versus small molecular contrast medium (gadopentetate) enhanced magnetic resonance imaging: differentiation of benign and malignant breast lesions. Acad Radiol 2003 Nov;10(11):1237-46.

118. 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.

119. 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.

120. Schedel H, Oellinger H, Kohlschein P, et al. 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.

121. 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 May-Jun;88(3):224-33.

122. Furman-Haran E, Grobgeld D, Kelcz F, et al. Critical role of spatial resolution in dynamic contrast-enhanced breast MRI. J Magn Reson Imaging 2001 Jun;13(6):862-7.

123. Torheim G, Godtliebsen F, Axelson D, et al. Feature extraction and classification of dynamic contrast-enhanced T2*-weighted breast image data. IEEE Trans Med Imaging 2001 Dec;20(12):1293-301.

124. 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.

125. Yeung DK, Cheung HS, Tse GM. Human breast lesions: characterization with contrast-enhanced in vivo proton MR spectroscopy--initial results. Radiology 2001 Jul;220(1):40-6.

126. Kvistad KA, Rydland J, Vainio J, et al. Breast lesions: evaluation with dynamic contrast-enhanced T1-weighted MR imaging and with T2*-weighted first-pass perfusion MR imaging. Radiology 2000 Aug;216(2):545-53.

127. Van Goethem M, Biltjes IG, De Schepper AM. Indications for MR mammography. A Belgian study. JBR-BTR 2000 Jun;83(3):126-9.

128. Peters NH, Borel Rinkes IH, Zuithoff NP, et al. Meta-analysis of MR imaging in the diagnosis of breast lesions. Radiology 2008 Jan;246(1):116-24.

129. Blue Cross Blue Shield Association (BCBS). Magnetic resonance imaging of the breast: differential diagnosis of a breast lesion to avoid biopsy. Technol Eval Cent Asses Program 2002 Feb;16(15):1-29.

130. Breast MRI for detection or diagnosis of primary or recurrent breast cancer. Technol Eval Cent Asses Program Exec Summ 2004 Apr;19(1):1-9.

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

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

133. Bergmann H, Dobrozemsky G, et al. 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.

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

135. Wiley G. CTA: watershed, soon to be a flood. [internet]. Los Angeles (CA): Ascend Media; 2005 Mar [accessed 2005 May 25]. [9 p]. Available: http://www.imagingeconomics.com/library/200503-02.asp.

136. 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.

137. 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.

138. Rosen EL, Eubank WB, Mankoff DA. FDG PET, PET/CT, and breast cancer imaging. Radiographics 2007 Oct;27 Suppl 1:S215-29.

139. Brenner DJ, Hall EJ. Computed tomography--an increasing source of radiation exposure. N Engl J Med 2007 Nov 29;357(22):2277-84.

140. 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.

141. 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.

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

143. 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 Mar-Apr;24(2):274-80.

144. 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.

145. 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.

146. 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.

147. 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.

148. 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.

149. Kaida H, Ishibashi M, Fuji T, et al. Improved breast cancer detection of prone breast fluorodeoxyglucose-PET in 118 patients. Nucl Med Commun 2008 Oct;29(10):885-93.

150. Buchmann I, Riedmuller K, et al. Comparison of 99mtechnetium-pertechnetate and 123iodide SPECT with FDG-PET in patients suspicious for breast cancer. Cancer Biother Radiopharm 2007 Dec;22(6):779-89.

151. Schirrmeister H, Kuhn T, Guhlmann A, et al. Fluorine-18 2-deoxy-2-fluoro-D-glucose PET in the preoperative staging of breast cancer: comparison with the standard staging procedures. Eur J Nucl Med 2001 Mar;28(3):351-8.

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

153. 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.

154. Escalona S, Blasco JA, Reza MM, et al. A systematic review of FDG-PET in breast cancer. Med Oncol 2010 Mar;27(1):114-29.

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

156. Prendergast A, Adams J, Delgado M, et al. SPECT vs. Planar Imaging [power point]. [slide set]. Waltham (MA): Daum Communications Corp.; [accessed 2010 Mar 2]. [6 slides]. Available: http://74.125.93.132/search?q=cache:xzGe13cT8CUJ:keiserstudents.tripod.com/sitebuildercontent/sitebuilderfiles/spectvs.ppt+what+is+planar+imaging+AND+prendergast&cd=1&hl=en&ct=clnk&gl=us.

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

158. 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.

159. 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.

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

161. BSGI/Molecular breast imaging. [Web site]. Newport News (VA): Dilon Technologies; [accessed 2010 Feb 16]. [Various p]. Available: http://www.dilon.com/pages/bsgi__molecular_breast_imaging/34.php.

162. 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.

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

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

165. 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.

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

167. 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.

168. Bekis R, Derebek E, Balci P, et al. 99mTc sestamibi scintimammography. Screening mammographic non-palpable suspicious breast lesions: preliminary results. Nucl Med (Stuttg) 2004 Feb;43(1):16-20.

169. 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.

170. 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 Jul-Aug;9(4):288-94.

171. 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.

172. 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.

173. 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.

174. 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.

175. 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.

176. 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.

177. 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.

178. 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.

179. 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.

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

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

182. 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.

183. 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.

184. 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.

185. 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.

186. 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.

187. 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.

188. 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.

189. 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.

190. 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.

191. 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.

192. 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 May-Jun;17(3B):1661-5.

193. 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 May-Jun;17(3B):1599-605.

194. 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 May-Jun;17(3B):1677-81.

195. 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 May-Jun;17(3B):1695-8.

196. 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.

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

198. Scopinaro F, Ierardi M, Porfiri LM, et al. 99mTc-MIBI prone scintimammography in patients with high and intermediate risk mammography. Anticancer Res 1997 May-Jun;17(3B):1635-8.

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

200. 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.

201. 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.

202. 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.

203. 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.

204. Burak Z, Argon M, 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.

205. Mathieu I, Mazy S, Willemart B, et al. Inconclusive triple diagnosis in breast cancer imaging: is there a place for scintimammography. J Nucl Med 2005 Oct;46(10):1574-81.

206. Habib S, Maseeh-uz-Zaman, Hameed A, et al. Diagnostic accuracy of Tc-99m-MIBI for breast carcinoma in correlation with mammography and sonography. J Coll Physicians Surg Pak 2009 Oct;19(10):622-6.

207. Kim IJ, Kim SJ, Kim YK. Comparison of double phase Tc-99m MIBI and Tc-99m tetrofosmin scintimammography for characterization of breast lesions: Visual and quantitative analyses. Neoplasma 2008;55(6):526-31.

208. Kim SJ, Bae YT, Lee JS, et al. Diagnostic performances of double-phase tc-99m MIBI scintimammography in patients with indeterminate ultrasound findings: visual and quantitative analyses. Ann Nucl Med 2007 Jun;21(3):145-50.

209. Pinero A, Galindo PJ, Illana J, et al. Diagnostic efficiency of sestamibi gammagraphy and Doppler sonography in the preoperative assessment of breast lesions. Clin Transl Oncol 2006 Feb;8(2):103-7.

210. Grosso M, Chiacchio S, Bianchi F, et al. Comparison between 99mTc-sestamibi scintimammography and X-ray mammography in the characterization of clusters of microcalcifications: a prospective long-term study. Anticancer Res 2009 Oct;29(10):4251-7.

211. Wang F, Wang Z, Wu J, et al. The role of technetium-99m-labeled octreotide acetate scintigraphy in suspected breast cancer and correlates with expression of SSTR. Nucl Med Biol 2008 Aug;35(6):665-71.

212. Gommans GM, van der Zant FM, van Dongen A, et al. (99M)Technetium-sestamibi scintimammography in non-palpable breast lesions found on screening X-ray mammography. Eur J Surg Oncol 2007 Feb;33(1):23-7.

213. Schillaci O, Danieli R, Filippi L, et al. Scintimammography with a hybrid SPECT/CT imaging system. Anticancer Res 2007 Jan;27(1 B):557-62.

214. Hussain R, Buscombe JR. A meta-analysis of scintimammography: an evidence-based approach to its clinical utility. Nucl Med Commun 2006 Jul;27(7):589-94.

215. Liu M, Guo YM, Guo XJ, et al. Evaluation of 99mTc-MIBI scintimammorgraphy in the diagnosis of primary breast cancer: A meta-analysis. Chi J Evid Based Med 2005 Jul;5(7):536-42.

216. 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.

217. 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.

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

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

220. Athanasiou A, Tardivon A, Ollivier L, et al. How to optimize breast ultrasound. Eur J Radiol 2009 Jan;69(1):6-13.

221. Sehgal CM, Weinstein SP, Arger PH, et al. A review of breast ultrasound. J Mammary Gland Biol Neoplasia 2006 Apr;11(2):113-23.

222. Gokhale S. Ultrasound characterization of breast masses. Indian J Radiol Imaging 2009 Aug 1;19(3):242-7.

223. 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 - Ultrasound.

224. Martinoli C, Pretolesi F, Crespi G, et al. Power Doppler sonography: clinical applications. Eur J Radiol 1998 May;27 Suppl 2:S133-40.

225. 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.

226. 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.

227. 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.

228. 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.

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

230. 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.

231. 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.

232. van Oord JC, van der Vliet AM, Thyn CJ, et al. The value of ultrasound mammography in palpable breast masses. Rofo 1991 Jul;155(1):63-6.

233. Gokalp G, Topal U, Kizilkaya E. Power Doppler sonography: anything to add to BI-RADS US in solid breast masses? Eur J Radiol 2009 Apr;70(1):77-85.

234. Liu H, Jiang YX, Liu JB, et al. Evaluation of breast lesions with contrast-enhanced ultrasound using the microvascular imaging technique: initial observations. Breast 2008 Oct;17(5):532-9.

235. Vade A, Lafita VS, Ward KA, et al. Role of breast sonography in imaging of adolescents with palpable solid breast masses. AJR Am J Roentgenol 2008 Sep;191(3):659-63.

236. Cha JH, Moon WK, Cho N, et al. Characterization of benign and malignant solid breast masses: comparison of conventional US and tissue harmonic imaging. Radiology 2007 Jan;242(1):63-9.

237. Chala L, Endo E, Kim S, et al. Gray-scale sonography of solid breast masses: diagnosis of probably benign masses and reduction of the number of biopsies. J Clin Ultrasound 2007 Jan;35(1):9-19.

238. Zhi H, Ou B, Luo BM, et al. Comparison of ultrasound elastography, mammography, and sonography in the diagnosis of solid breast lesions. J Ultrasound Med 2007 Jun;26(6):807-15.

239. Cho N, Moon WK, Cha JH, et al. Differentiating benign from malignant solid breast masses: comparison of two-dimensional and three-dimensional US. Radiology 2006 Jul;240(1):26-32.

240. 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.

241. 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.

242. 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.

243. 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.

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

245. Caruso G, Ienzi R, Cirino A, et al. Breast lesion characterization with contrast-enhanced US. Work in progress. Radiol Med 2002 Nov-Dec;104(5-6):443-50.

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

247. 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.

248. Moon WK, Im JG, Noh DY, et al. Nonpalpable breast lesions: evaluation with power Doppler US and a microbubble contrast agent-initial experience. Radiology 2000 Oct;217(1):240-6.

249. Blohmer JU, Oellinger H, Schmidt C, et al. Comparison of various imaging methods with particular evaluation of color Doppler sonography for planning surgery for breast tumors. Arch Gynecol Obstet 1999;262(3-4):159-71.

250. 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.

251. Albrecht T, Patel N, Cosgrove DO, et al. Enhancement of power Doppler signals from breast lesions with the ultrasound contrast agent EchoGen emulsion: subjective and quantitative assessment. Acad Radiol 1998 Apr;5 Suppl 1:S195-8; discussion S199.

252. 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.

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

254. 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.

255. 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.

256. Flobbe K, Kessels AG, Severens JL, et al. Costs and effects of ultrasonography in the evaluation of palpable breast masses. Int J Technol Assess Health Care 2004 Fall;20(4):440-8.

257. 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.

258. Guidance chapter. In: Breast Imaging Reporting and Data System Atlas (BI-RADS Atlas). BI-RADS - Mammography. 4th ed. Reston (VA): American College of Radiology (ACR); 2003. p. 253-60. Also available: http://www.acr.org/SecondaryMainMenuCategories/quality_safety/BIRADSAtlas/BIRADSAtlasexcerptedtext/BIRADSMammographyFourthEdition/FollowUpandOutcomeMonitoringDoc4.aspx.

259. Wiratkapun C, Bunyapaiboonsri W, Wibulpolprasert B, Lertsithichai P. Biopsy rate and positive predictive value for breast cancer in BI-RADS category 4 breast lesions. J Med Assoc Thai 2010 Jul;93(7):830-7.

260. Venkatesan A, Chu P, Kerlikowske K, Sickles EA, Smith-Bindman R. Positive predictive value of specific mammographic findings according to reader and patient variables. Radiology 2009 Mar;250(3):648-57.

261. 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.

262. Verkooijen HM, Peeters PH, Buskens E, et al. Diagnostic accuracy of large-core needle biopsy for nonpalpable breast disease: a meta-analysis. Br J Cancer 2000 Mar;82(5):1017-21.

263. Antley CM, Mooney EE, Layfield LJ. A comparison of accuracy rates between open biopsy, cutting-needle biopsy, and fine-needle aspiration biopsy of the breast: a 3-year experience. Breast J 1998;4(1):3-8.

264. Page DL, Dupont WD, Rogers LW, et al. Intraductal carcinoma of the breast: follow-up after biopsy only. Cancer 1982 Feb 15;49(4):751-8.

265. Yen MF, Tabar L, Vitak B, et al. Quantifying the potential problem of overdiagnosis of ductal carcinoma in situ in breast cancer screening. Eur J Cancer 2003 Aug;39(12):1746-54.

List of Acronyms/Abbreviations

ADH	Atypical Ductal Hyperplasia
AHRQ	Agency for Healthcare Research and Quality
ALH	Atypical Lobular Hyperplasia
BI-RADS®	Breast Imaging Reporting and Data System
BSGI	Breast specific gamma imaging
CER	Comparative Effectiveness Review
CI	Confidence Interval
CT	Computed tomography
2D	Two dimensional
3D	Three dimensional
DCIS	Ductal Carcinoma In Situ
FDG	fluorodeoxyglucose
FN	False Negative
FP	False Positive
LCIS	Lobular Carcinoma In Situ
MIBI	Sestamibi
MRI	Magnetic Resonance Imaging
NA	Not Applicable
NR	Not Reported
PET	Positron emission tomography
SMM	Scintimammography
SPECT	Single photon emission computed tomography
TEP	Technical expert panel
TN	True Negative
TP	True Positive
UK	United Kingdom
US	Ultrasound
USA	United States of America

Glossary of Selected Terms

Definition of Terms

1. Ductal carcinoma in situ (DCIS): a type of early stage breast cancer that is confined to the breast duct in which it arose.
2. Doppler ultrasound: a method of using ultrasound to evaluate blood flow through vessels. The speed of blood flow is evaluated by observing changes in the pitch of the reflected sound waves.
3. Harmonic ultrasound: Ultrasound waves develop harmonics as they pass through breast tissue. Digital encoding can be used by computers to construct images from the harmonic frequencies.
4. Magnetic resonance imaging: A method of imaging internal anatomy by using strong magnetic fields and radiofrequency energy.
5. Positron emission tomography: A method of imaging tissues by tracking the metabolism of a positron-emitting radioactive tracer.
6. Positive likelihood ratio: the ability of the diagnostic test to accurately predict the presence of breast cancer.
7. Positive predictive value: the probability of a woman actually having breast cancer after testing positive for breast cancer. Positive predictive value = (true positives)/(true positives + false positives).
8. Negative likelihood ratio: the ability of the diagnostic test to accurately "rule out" the presence of breast cancer.
9. Negative predictive value: the probability of a woman actually not having breast cancer after testing negative for breast cancer. Negative predictive value = (true negatives)/(false negatives + true negatives).
10. Scintimammography: A method of imaging tissues by tracking the metabolism of a radioactive tracer.
11. Sensitivity: the proportion of women with breast cancer who test positive for breast cancer. Sensitivity = (true positives)/(true positives + false negatives).
12. Specificity: the proportion of women with benign lesions who test negative for breast cancer.
    Specificity = (true negatives)/(false positives + true negatives).
13. Tomography ultrasound: Multiple ultrasound images from different angles are acquired and a computer used the information to develop a three-dimensional image of the interior anatomy of the breast.
14. Ultrasound: a method of imaging anatomy by observing the reflections of high-frequency sound waves off of tissues with different acoustic properties. Conventional ultrasound is often referred to as B-mode ultrasound.

Atypical ductal hyperplasia (ADH). A condition in which the cells that line the milk ducts of the breast experience abnormal growth. The lesion itself is not malignant but may sometimes contain foci of malignant cells and women with ADH have an elevated risk of developing a malignant lesion.

Ductal carcinoma in situ (DCIS). A carcinoma of the milk ducts of the breast that is confined within the duct.

High-risk lesion. Any of a number of different types of non-cancerous lesions of the breast that have been observed to sometimes contain foci of malignant cells, and women diagnosed with these types of lesions have an elevated risk of developing a malignant lesion. Some common types of high-risk lesions include atypical ductal hyperplasia (ADH), radial scars, papillary lesions, atypical lobular hyperplasia (ALH), and lobular carcinoma in situ (LCIS).

Microcalcification. A tiny deposit of calcium visible as a bright spot on a mammogram. Tight clusters of microcalcifications may be a sign of a malignant lesion.

Negative likelihood ratio. A measure of the ability of the diagnostic test to accurately "rule out" disease. The smaller the negative likelihood ratio is, the more accurate the test is.

Palpable lesion. A breast lesion that can be felt by manual manipulation.

Sensitivity. 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).