

Evidence-based Practice Center Systematic Review Protocol

Project Title: *Comparative Effectiveness of Noninvasive Diagnostic Tests for Breast Abnormalities – An Update of a 2006 Report*

I. Background and Objectives

In response to Section 1013 of the Medicare Modernization Act, the Agency for Health Care Research and Quality (AHRQ) requested an update of the comparative effectiveness review, *Effectiveness of Noninvasive Diagnostic Tests for Breast Abnormalities*. The original report was finalized in February 2006.¹ Because of technological advances and continuing innovation in the fields of noninvasive imaging, the conclusions of the original report are possibly no longer relevant to current clinical practice.

Breast cancer is the second most common malignancy of women.² The American Cancer Society estimates that in the United States in 2009, 62,280 women were diagnosed with new cases of *in situ* breast cancer, 192,730 women were newly diagnosed as having invasive breast cancer, and 40,480 women died of this disease.² In the general population, the cumulative risk of being diagnosed with breast cancer by age 70 is estimated to be 6 percent, with a lifetime risk of 13 percent.²⁻⁴

Breast cancer is usually first detected by feeling a lump on physical examination (either self-examination or an examination 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 5-year survival rate is close to 100%. The 5-year survival rate for women with stage IV (cancer that has spread beyond the breast) is only 27 percent.³ These observations suggest that breast cancer mortality rates can be significantly reduced by identifying the disease at earlier stages. Because early breast cancer is asymptomatic, the only way to detect it is through population-wide screening. Mammography is a widely accepted method for breast cancer screening.^{2,5,6}

In mammography, x-rays are used to examine the breast for clusters of microcalcifications, circumscribed and dense masses, masses with indistinct margins, architectural distortion (in comparison to the contralateral breast), or other abnormal structures. Currently, most professional organizations recommend that all women older than 50 years of age receive an annual or biennial mammogram.^{2,6} Some professional organizations recommend that routine breast cancer screening begin earlier, at age 40, although x-ray mammographic screening is less effective in younger women.⁶ Most experts believe that regular x-ray mammographic screening of all women who are between the age of 50 and 74 can reduce mortality from breast cancer.^{2,5,6} The United States Preventive Services Task Force has recently recommended that aged 50 to 74

years receive routine screening mammography every 2 years and that decisions to screen women under the age of 50 be made on an individual basis.⁶

After a possible abnormality is identified on screening mammography or on physical examination, women typically undergo diagnostic mammography. If these radiographic studies suggest the abnormality may be malignant, a biopsy of the suspicious area may be recommended. This evidence review focuses on noninvasive imaging studies that could be conducted as part of the diagnostic workup of possible breast abnormalities and be used 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 appropriate care pathways: biopsy, "watchful waiting," or return to normal screening intervals.

The American College of Radiology has created a standardized system for reporting the results of breast imaging, the Breast Imaging Reporting and Data System (BI-RADS®).⁷⁻

⁹ 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.

It is important to accurately triage women into the correct care pathway. Women with readily treatable types of breast cancer who get incorrectly triaged into the "return to normal screening" pathway may experience a significant delay in diagnosis and treatment of their cancer, a delay that may allow the tumor to spread and become life-threatening. However, most women who are recalled for further assessment after a screening mammogram do not have cancer. Elmore et al.¹⁰ estimated that a woman's cumulative risk of having a false-positive finding on screening mammography is close to 50% after 10 years of annual screenings. 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 frequent repeat mammographic examinations. If an available noninvasive diagnostic test could assist clinicians in evaluating women who have been recalled for further investigation after mammographic screening—namely, to assist in accurately distinguishing between "benign", "probably benign", and "probably not benign" lesions—then many women could avoid frequent repeat mammographic examinations and their attendant discomfort, inconvenience, x-ray exposure, and emotional distress.

Most women who traditionally have been referred for biopsy also do not have cancer. For example, Lacquement et al.¹¹ examined a series of 668 women who underwent biopsy; only 23 percent of these women were diagnosed with breast cancer after biopsy. 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 noninvasive 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 update is to provide information about the accuracy of noninvasive imaging technologies. This information may be useful to clinicians when deciding if it is clinically appropriate to use various types of noninvasive technologies to evaluate breast abnormalities. It is reasonable to assume that none of the noninvasive technologies will achieve an accuracy equivalent to or better than biopsy. However, it is also reasonable to assume that noninvasive technologies are safer than invasive biopsy methods and, therefore, that some women may benefit from the use of particular noninvasive technologies.

For clinicians to decide whether or not a test is clinically appropriate for any patient, the diagnostic capabilities of the noninvasive tests must be known for women with a variety of demographic and clinical risk factors. Because women with a previous history of breast cancer and women known to be at high risk for the disease because they carry the *BRCA1* or *BRCA2* mutation or have a very strong family history of breast cancer are known to have a very different risk profile than the general population, we will not evaluate the use of noninvasive technologies for such women in this report. Instead, we will focus on the use of noninvasive imaging technology for women from the general population who have an abnormal finding as shown by screening mammography or a physical examination. If the evidence permits, we will also examine the influence of various factors—age; the size and morphological characteristics of the breast lesion; the presence of calcifications; the density of the breast tissue; and other key clinical risk factors—on the accuracy of the noninvasive imaging methods.

II. The Key Questions

This systematic review is an update of Comparative Effectiveness Review No. 5, which was originally published in 2006.¹ A Technical Expert Panel (TEP) was assembled to advise the Evidence-based Practice Center (EPC) investigators during the update. After discussion with the TEP, the EPC investigators revised the wording of the Key Questions and added additional diagnostic tests to the list of tests to be evaluated in the systematic review. The 2006 version of the review evaluated only B-mode ultrasound, magnetic resonance imaging (MRI) without computer-aided diagnosis, positron emission tomography (PET) without computed tomography (CT), and full-body scintimammography (see the Key Questions below for the tests to be evaluated in the updated systematic review). In addition to the tests chosen for review, the TEP discussed a number of other diagnostic tests, such as breast elastography, breast tomography, breast thermography, positron emission mammography, computer-assisted diagnostic x-ray mammography, and molecular breast imaging (MBI) but the

consensus of the TEP was that these technologies had not been studied sufficiently to warrant inclusion in the systematic review at this time.

The proposed Key Questions were posted for public comment. People who commented on the Key Questions were chiefly concerned about the need to update the 2006 review, stating that available data on the tests has not changed substantially since 2006. Other people were concerned that the conclusions of the review would not be useful for clinicians, because the current standard of care does not incorporate additional imaging into patient management after a standard diagnostic workup for a possible breast abnormality. One reviewer suggested that MBI be added to the list of tests to be evaluated. Other people commented that breast density and breast implants need to be considered when addressing Key Question 2 and that the volume of tests completed by individual facilities and radiologists should be considered when addressing Key Question 3.

The comments received on the Key Questions did not lead to revisions of the Key Questions. The EPC investigators felt that most concerns and issues raised by the reviewers should be addressed in the Results and Conclusions sections of the review and that the Key Questions should be written as posted. Only one comment—the suggestion to include MBI in the list of tests to be addressed—affected the Key Questions. The consensus of the TEP was that there is currently insufficient evidence available about MBI for clinicians to benefit from inclusion of the technology in a systematic review at this time.

Question 1

What is the accuracy (expressed as sensitivity, specificity, predictive values, and likelihood ratios) of noninvasive tests for diagnosis of breast cancer in women referred for further evaluation after a possible breast abnormality is identified on routine screening (mammography with or without clinical or self-detection of a palpable lesion)?

The non-invasive diagnostic tests to be evaluated are:

- Ultrasound (B-mode grayscale, 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.
- Positron emission tomography (PET) with fluorine-18-fluorodeoxyglucose as the tracer, with or without concurrent computed tomography (CT) scans.
- Scintimammography with technetium-99m-sestamibi as the tracer, including breast-specific gamma imaging (BSGI).

Question 2

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

Source: www.effectivehealthcare.ahrq.gov

Published Online: July 01, 2010

Question 3

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

Populations

The population of interest is women who have been referred for further evaluation after a possible breast abnormality was detected on routine screening, whether by mammography, clinical examination, or self-examination. These women would be from a general population of women who participate in screening programs. Populations that will not be evaluated include: women thought to be at very high risk of breast cancer before any abnormality is detected, such as those with an extensive family history of breast cancer or carriers of a *BRCA1* or *BRCA2* mutation; women with a personal history of breast cancer; and men.

Interventions

The noninvasive diagnostic tests to be evaluated are:

- Ultrasound (B-mode grayscale, harmonic, color Doppler, and power Doppler).
- MRI with breast-specific coils and gadolinium-based contrast agents, with or without computer-aided diagnosis
- PET with 18-fluorodeoxyglucose as the tracer, with or without concurrent computed tomography (CT) scans
- Scintimammography with technetium-99m-sestamibi as the tracer, including BSGI.

Comparators

The accuracy of the noninvasive imaging tests will be evaluated by a direct comparison to histopathology (of biopsy or surgical specimens) or to clinical followup or to a combination of these methods. In addition, the relative accuracy of the different tests will be evaluated by directly and indirectly comparing the tests as the reported evidence permits.

Outcomes

Outcomes of interest are diagnostic test characteristics, namely, sensitivity, specificity, positive predictive value, negative predictive value, 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—will be evaluated.

Setting

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

III. Analytic Framework

Figure 1. Draft Analytical Framework for the Comparative Effectiveness of Noninvasive Diagnostic Tests for Breast Abnormalities

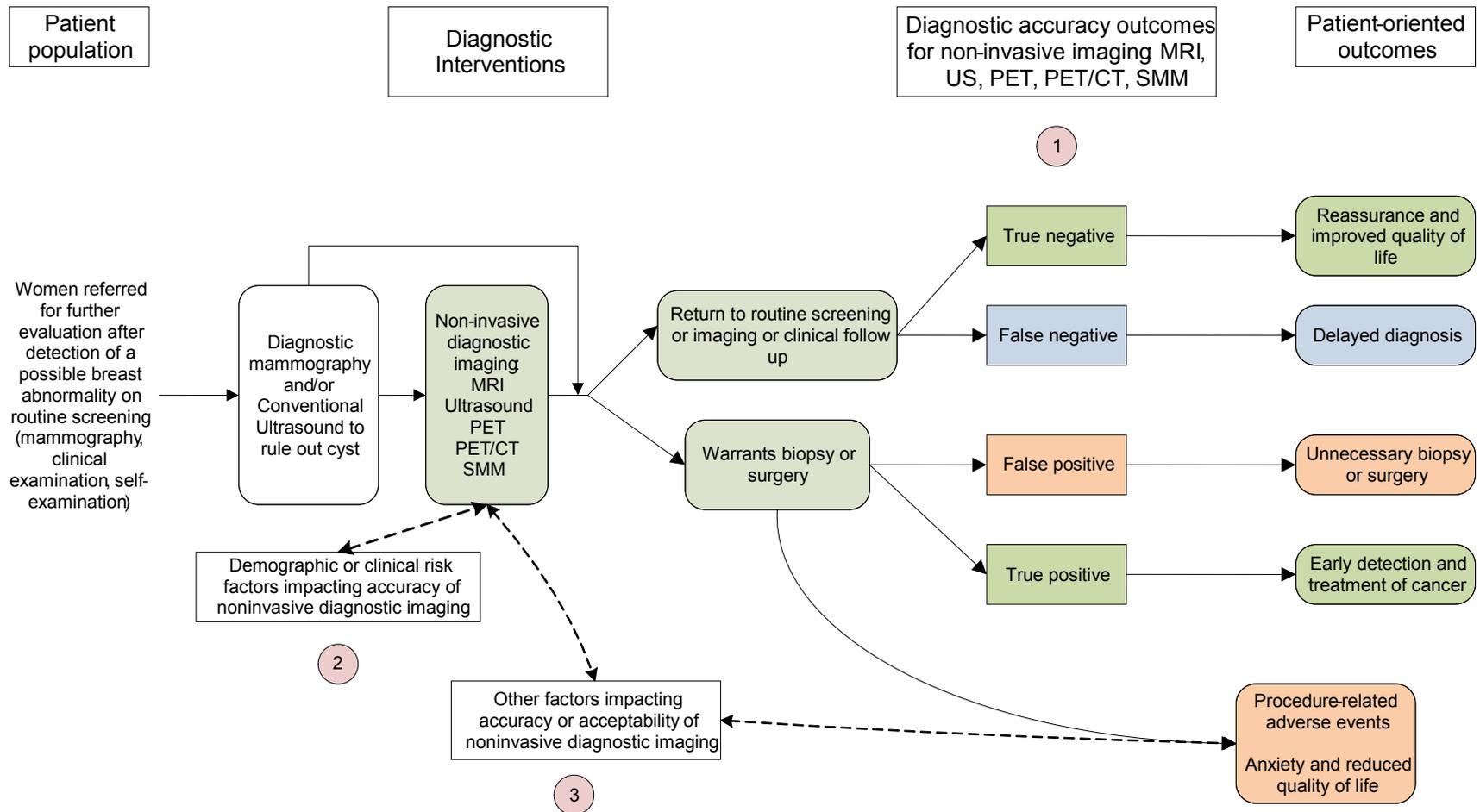


Figure 1: This figure depicts the key questions within the context of the patient population, diagnostic tests, subsequent interventions, and outcomes. In general, the figure illustrates how the use of additional non-invasive imaging tests may affect decisions about patient management, and how such decisions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

IV. Methods

A. Criteria for Inclusion/Exclusion of Studies in the Review

We will use the following formal criteria to determine which studies will be included in our analysis.

1. The study must have directly compared the test of interest to core-needle biopsy, open surgery, or patient followup in the same group of patients.

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.¹² Studies may have performed biopsy procedures on all patients, or may have performed a biopsy on some patients and followed the other patients with clinical examinations and mammograms. Fine-needle aspiration of solid lesions is not an acceptable reference standard for the purposes of this assessment.¹³⁻¹⁶

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.^{12,17} 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.¹⁷

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

Studies of outdated and experimental technologies 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 below.

Technology	Cut-off Publication Date (to Present) To Exclude Outdated Technology	Other Inclusion Criteria
Ultrasound	1994	

Technology	Cut-off Publication Date (to Present) To Exclude Outdated Technology	Other Inclusion Criteria
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 approved by the U.S. Food and Drug Administration for use as a diagnostic test for breast cancer, 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	Fluorine-18-fluorodeoxyglucose (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 single-photon emission tomography (SPECT); only studies that used sestamibi (also called MIBI or technetium-99m sestamibi) as the tracer

3. The study enrolled primarily 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 review; however, studies that enrolled primarily women and only one or two men will be included.

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

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

Source: www.effectivehealthcare.ahrq.gov

Published Online: July 01, 2010

evaluation for any of the following purposes were excluded as being outside the scope of the review: 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 carriers of the BRCA1 or BRCA2 mutations, or patients with a strong family history of breast cancer, are also outside the scope of the review. If a study enrolled a mixed patient population and did not report data separately, it was excluded if more than 15 percent of the subjects did not fall into the category of “primary diagnosis of women who are at average risk and had an abnormality detected on routine screening.”

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 noninvasive test being evaluated.

Other outcomes are beyond the scope of this review.

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 will be excluded.

7. Study must be published in English.

Moher et al. have demonstrated that excluding non-English–language studies from meta-analyses has little impact on the conclusions drawn.¹⁸ Juni et al.¹⁹ 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 excluding non-English studies could lead to bias in some situations, we believe that the risk of bias would be too low to justify the time and cost necessary to translate studies to identify those of acceptable quality for inclusion in the review.

8. Study must be published as a peer-reviewed full article.

Meeting abstracts will not be included. Published meeting abstracts have not been peer-reviewed and often do not include sufficient details about experimental methods to permit one to verify that the study was well designed.^{20,21} In addition, it is not uncommon for abstracts that are published as part of conference proceedings to have inconsistencies when they are compared to the final publication of the study or to describe studies that are never published as full articles.²²⁻²⁶

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

Source: www.effectivehealthcare.ahrq.gov

Published Online: July 01, 2010

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 of a study are available, only outcome data from the most recent report will be included. However, we will use relevant data from earlier and smaller reports if the report presented pertinent data not presented in the more recent report.

B. Searching for the Evidence: Literature Search Strategies for Identification of Relevant Studies to Answer the Key Questions

Electronic databases, including EMBASE and MEDLINE, will be searched for clinical trials that appear to address the Key Questions. Keywords included in the search strategy encompass the concepts of breast cancer, diagnosis, noninvasive imaging, and names of the specific technologies to be evaluated. The bibliographies of recent on-topic systematic and narrative reviews will be scanned for additional information. The literature searches will be updated while the report is undergoing internal review, and any key new publications identified by either the peer reviewers or the updated literature searches will be incorporated into the report before finalization.

Standardized forms for screening abstracts and articles will be developed by using the SRS[®] 4.0 database (Mobius Analytics, Ontario, Canada). The abstracts and titles of all publications identified by the literature searches will be reviewed in duplicate. Articles that appear to address the Key Questions and meet the inclusion criteria will be evaluated at the full article level by using standardized forms to evaluate each article against the inclusion and exclusion criteria described above.

C. Data Abstraction and Data Management

Standardized data abstraction forms will be created by using the SRS[®] 4.0 database. Data will be abstracted and managed also by using this database.

D. Assessment of Methodological Quality of Individual Studies

We will use an internal validity rating scale for diagnostic studies to grade the internal validity of the evidence base. This instrument is based on a modification of the QUADAS instrument with reference to empirical studies of design-related bias in diagnostic test studies.^{17,27} 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.

E. Data Synthesis

We will meta-analyze the data reported in the eligible studies by using a bivariate mixed-effects binomial regression model as described by Harbord et al.²⁸ All such analyses are computed by using the STATA 10.0 statistical software package and the “midas” command.²⁹ The summary likelihood ratios and Bayes’ theorem will be used to calculate the posttest probability of having a benign or malignant lesion. In cases where a bivariate binomial regression model cannot be fit, we will meta-analyze the data by using a random-effects model and the Meta-Disc software package.³⁰ Narrative discussion will be used to perform indirect comparisons of relative test accuracy. Meta-regressions will be performed with the “midas” software. Variables will be tested in a meta-regression to explore possible causes of heterogeneity and to address Key Questions 2 and 3 and include study quality instruments, patient and lesion characteristics, details of imaging methodology, and details of study design.

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 as the result of a delayed cancer diagnosis. The clinical relevance of the findings of the analyses will be explored by using likelihood ratios and Bayes’ theorem to directly compute an individual woman’s risk of actually having a malignancy when diagnosed by using an imaging technique. These results will be incorporated into a discussion of the possible clinical consequences of the use of noninvasive imaging.

F. Grading the Evidence for Each Key Question

We will use a formal grading system that conforms with the recommendations in the Methods Guide for Effectiveness and Comparative Effectiveness Reviews for grading the strength of evidence.^{31,32}

The overall strength of evidence supporting each major conclusion will be graded as high, moderate, low, or insufficient. The grade will be 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 in the aggregate evidence base that supports each major conclusion will be rated as being low, medium, or high. We will use our inclusion and exclusion criteria to eliminate studies that have 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 consecutive patients or did not include them in the analysis. Because we plan to eliminate all studies with a high risk of bias from the evidence base, we will consider the remaining evidence base to have either a low or medium risk of bias. We will use the internal validity rating instrument described above (see “Assessing Quality of

Source: www.effectivehealthcare.ahrq.gov

Published Online: July 01, 2010

Individual Studies”) to assess the risk of bias of each individual study and the aggregate risk of bias to rate the entire evidence base.

We will rate the consistency of conclusions supported by meta-analyses with the I^2 statistic.^{33,34} Data sets that are found to have an I^2 of less than 50 percent will be rated as being “consistent;” those with an I^2 of 50 percent or greater will be rated as being “inconsistent;” and data sets for which I^2 cannot be calculated (e.g., a single study) will be rated as “consistency unknown.” For qualitative comparisons between different diagnostic tests, we will rate conclusions as being “consistent” if the effect sizes are 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 sensitivity estimates for the individual studies are consistently higher for studies that used a contrast agent, then the evidence base will be rated as “consistent.”

For the review, we will define a “precise” estimate of sensitivity or specificity as being one for which both the upper and lower bound of the 95 percent confidence interval (95% CI) is no more than 5 points away from the summary estimate; for example, a sensitivity of 98 percent (95% CI: 97–100%) will be a precise estimate of sensitivity, whereas a sensitivity of 98 percent (95% CI: 88–100%) will be an imprecise estimate of sensitivity. Precision may be rated separately for summary estimates of sensitivity and specificity for each major conclusion. For qualitative comparisons between different diagnostic tests, the conclusion will be rated as being “precise” if the confidence intervals around the summary estimates being compared do not overlap.

For studies of diagnostic test accuracy, the evidence will be consistently rated as being “indirect,” because the outcome of test accuracy is indirectly related to health outcomes. However, the Key Questions that guide this comparative effectiveness review are not concerned about the impact of test accuracy on health outcomes. We, therefore, will not incorporate the “indirectness” of the evidence into the overall rating of strength of evidence for Key Questions that are not concerned about health outcomes.



V. References

1. Bruening W, Launders J, Pinkney N, et al. *Effectiveness of Noninvasive Diagnostic Tests for Breast Abnormalities*. Comparative Effectiveness Review No. 2 (Prepared by ECRI Evidence-based Practice Center under Contract No. 290-02-0019). Rockville (MD): Agency for Healthcare Research and Quality, February 2006. AHRQ Publication No. 06-EHC005-EF. Available at: <http://effectivehealthcare.ahrq.gov/repFiles/BrCADx%20Final%20Report.pdf>.
2. American Cancer Society Web site. Cancer Facts & Figures 2009. Available at: <http://www.cancer.org/Research/CancerFactsFigures/cancer-facts-figures-2009>. Accessed June 29, 2010.
3. National Cancer Institute Web site. Surveillance Epidemiology and End Results. SEER Stat Fact Sheets: Breast. Available at: http://seer.cancer.gov/statfacts/html/breast_print.html. Accessed June 29, 2010.
4. National Cancer Institute Web site. DevCan— Probability of Developing or Dying of Cancer. Available at: <http://srab.cancer.gov/devcan/>. Accessed June 29, 2010.
5. Rosato FE, Rosato EL. Examination techniques: roles of the physician and patient in evaluating breast diseases. In: Bland KI, Copeland EM 3rd, eds. *The breast: comprehensive management of benign and malignant diseases*. 2nd ed. Vol. 1. Philadelphia: W.B. Saunders Company; 1998. p. 615-23.
6. U.S. Preventive Services Task Force. Screening for breast cancer: U.S. Preventive Services Task Force recommendation statement. *Ann Intern Med* 2009;151:716-26.
7. Liberman L, Menell JH. Breast imaging reporting and data system (BI-RADS). *Radiol Clin North Am* 2002;40:409-30.
8. Creighton University Medical Center Web site. Diagnostic and Interventional Radiology Department: Basic Imaging: Mammography. Available at: <http://radiology.creighton.edu/mammo.htm>. Accessed June 29, 2010.
9. Obenauer S, Hermann KP, Grabbe E. Applications and literature review of the BI-RADS classification. *Eur Radiol* 2005;15:1027-36.
10. Elmore JG, Barton MB, Mocerri VM, et al. Ten-year risk of false positive screening mammograms and clinical breast examinations. *N Engl J Med* 1998;338:1089-96.
11. Lacquement MA, Mitchell D, Hollingsworth AB. Positive predictive value of the Breast Imaging Reporting and Data System. *J Am Coll Surg* 1999;189:34-40.

12. Deeks JJ. Systematic reviews of evaluations of diagnostic and screening tests. In: Egger M, Smith GD and Altman DG, eds. *Systematic reviews in health care: meta-analysis in context*. 2nd ed. London, England: BMJ Books; 2001. p. 248-82.
13. Bojia F, Demisse M, Dejane A, et al. Comparison of fine-needle aspiration cytology and excisional biopsy of breast lesions. *East Afr Med J* 2001;78:226-8.
14. 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;69:736-40.
15. 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;79:59-63.
16. 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;93:263-8.
17. Lijmer JG, Mol BW, Heisterkamp S, et al. Empirical evidence of design-related bias in studies of diagnostic tests. *JAMA* 1999;282:1061-6.
18. Moher D, Pham B, Klassen KF, et al. What contributions do languages other than English make on the results of meta-analyses? *J Clin Epidemiol*. 2000; 53:964-72.
19. 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;31:115-23.
20. Chalmers I, Adams M, Dickersin K, et al. A cohort study of summary reports of controlled trials. *JAMA* 1990;263:1401-5.
21. Neinstein LS. A review of Society for Adolescent Medicine abstracts and Journal of Adolescent Health Care articles. *J Adolesc Health Care* 1987;8:198-203.
22. 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;22:288-94.
23. 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;3:187-91.
24. Scherer RW, Langenberg P, von Elm E. Full publication of results initially presented in abstracts. *Cochrane Database Syst Rev* 2007;(2):MR000005.



25. Marx WF, Cloft HJ, Do HM, et al. The fate of neuroradiologic abstracts presented at national meetings in 1993: rate of subsequent publication in peer-reviewed, indexed journals. *AJNR Am J Neuroradiol* 1999;20:1173-7.
26. Yentis SM, Campbell FA, Lerman J. Publication of abstracts presented at anaesthesia meetings. *Can J Anaesth* 1993;40:632-4.
27. 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;3:25
28. Harbord RM, Deeks JJ, Egger M, et al. A unification of models for meta-analysis of diagnostic accuracy studies. *Biostatistics* 2007;8:239-51.
29. StataCorp. Stata/MP software. College Station, TX: StataCorp; 1996-2010. 30. 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.
31. 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;63:513-23.
32. 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; 2009. Available at: http://effectivehealthcare.ahrq.gov/ehec/products/122/328/2009_0805_grading.pdf.
33. Higgins JP, Thompson SG. Quantifying heterogeneity in a meta-analysis. *Stat Med* 2002;21:1539-58.
34. Higgins JP, Thompson SG, Deeks JJ, et al. Measuring inconsistency in meta-analyses. *BMJ* 2003;327:557-60.

VI. 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 of internal anatomy from the harmonic frequencies.
4. **Magnetic resonance imaging (MRI):** A method of imaging internal anatomy by using strong magnetic fields and radiofrequency energy.
5. **Molecular breast imaging (MBI):** A variant on scintimammography that uses gamma cameras and a radioactive tracer to obtain images of metabolic patterns.
6. **Positron emission tomography:** A method of imaging the metabolic patterns of tissues by tracking the metabolism of a positron-emitting radioactive tracer.
7. **Positive likelihood ratio:** the ability of the diagnostic test to accurately predict the presence of breast cancer.
8. **Positive predictive value:** the probability of a woman actually having breast cancer after testing positive for breast cancer. $\text{Positive predictive value} = \frac{\text{true positives}}{\text{true positives} + \text{false positives}}$.
9. **Negative likelihood ratio:** the ability of the diagnostic test to accurately “rule out” the presence of breast cancer.
10. **Negative predictive value:** the probability of a woman actually not having breast cancer after testing negative for breast cancer. $\text{Negative predictive value} = \frac{\text{true negatives}}{\text{false negatives} + \text{true negatives}}$.
11. **Scintimammography:** A method of imaging metabolic patterns of tissues by tracking the metabolism of a radioactive tracer with gamma cameras.
12. **Sensitivity:** the proportion of women with breast cancer who test positive for breast cancer. $\text{Sensitivity} = \frac{\text{true positives}}{\text{true positives} + \text{false negatives}}$.
13. **Specificity:** the proportion of women with benign lesions who test negative for breast cancer. $\text{Specificity} = \frac{\text{true negatives}}{\text{false positives} + \text{true negatives}}$.
14. **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.

15. 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.

VII. Summary of Protocol Amendments

In the event of protocol amendments, the date of each amendment will be accompanied by a description of the change and the rationale.

NOTE: The following protocol elements are standard procedures for all protocols.

VIII. Review of Key Questions

For Comparative Effectiveness reviews (CERs) the key questions were posted for public comment and finalized after review of the comments. For other systematic reviews, key questions submitted by partners are reviewed and refined as needed by the EPC and the Technical Expert Panel (TEP) to assure that the questions are specific and explicit about what information is being reviewed.

IX. Technical Expert Panel (TEP)

A TEP panel is selected to provide broad expertise and perspectives specific to the topic under development. Divergent and conflicted opinions are common and perceived as health scientific discourse that results in a thoughtful, relevant systematic review. Therefore study questions, design and/or methodological approaches do not necessarily represent the views of individual technical and content experts. The TEP provides information to the EPC to identify literature search strategies, review the draft report and recommend approaches to specific issues as requested by the EPC. The TEP does not do analysis of any kind nor contribute to the writing of the report.

X. Peer Review

Approximately five experts in the field will be asked to peer review the draft report and provide comments. The peer reviewer may represent stakeholder groups such as professional or advocacy organizations with knowledge of the topic. On some specific reports such as reports requested by the Office of Medical Applications of Research, National Institutes of Health there may be other rules that apply regarding participation in the peer review process. Peer review comments on the preliminary draft of the report are considered by the EPC in preparation of the final draft of the report. The synthesis of the scientific literature presented in the final report does not necessarily represent the views of individual reviewers. The dispositions of the peer review comments are documented and will, for CERs and Technical briefs, be published three months after the publication of the Evidence report.

It is our policy not to release the names of the Peer reviewers or TEP panel members until the report is published so that they can maintain their objectivity during the review process.

Source: www.effectivehealthcare.ahrq.gov

Published Online: July 01, 2010



Source: www.effectivehealthcare.ahrq.gov
Published Online: July 01, 2010