Comparative Effectiveness Research Review Disposition of Comments Report

Research Review Title: *Noninvasive Diagnostic Tests for Breast Abnormalities: Update of a 2006 Review*

Draft review available for public comment from January 13, 2011 to February 10, 2011.

Research Review Citation: Bruening W, Uhl S, Fontanarosa J, Reston J, Treadwell J, Schoelles K. Noninvasive Diagnostic Tests for Breast Abnormalities: Update of a 2006 Review. Comparative Effectiveness Review No. 47. (Prepared by the ECRI Institute Evidence-based Practice Center under Contract No. 290- 02-0019.) AHRQ Publication No. 12-EHC014-EF. Rockville, MD: Agency for Healthcare Research and Quality; February 2012. Available at: www.effectivehealthcare.ahrq.gov/reports/final.cfm.

Comments to Research Review

The Effective Health Care (EHC) Program encourages the public to participate in the development of its research projects. Each comparative effectiveness research review is posted to the EHC Program Web site in draft form for public comment for a 4-week period. Comments can be submitted via the EHC Program Web site, mail or email. At the conclusion of the public comment period, authors use the commentators' submissions and comments to revise the draft comparative effectiveness research review.

Comments on draft reviews and the authors' responses to the comments are posted for public viewing on the EHC Program Web site approximately 3 months after the final research review is published. Comments are not edited for spelling, grammar, or other content errors. Each comment is listed with the name and affiliation of the commentator, if this information is provided. Commentators are not required to provide their names or affiliations in order to submit suggestions or comments.

The tables below include the responses by the authors of the review to each comment that was submitted for this draft review. The responses to comments in this disposition report are those of the authors, who are responsible for its contents, and do not necessarily represent the views of the Agency for Healthcare Research and Quality.

Commentator & Affiliation	Section	Comment	Response
Public Reviewer #1	Executive Summary	There is no way that radiologists interpreting breast imaging exams can accurately enough estimate the probability of malignancy for those in the 2 to 20% range. If they are wrong about the estimate and it is higher, the use of supplemental imaging does not bring the likelihood down to acceptable lesions to obviate the need for a tissue diagnosis. In addition, the prior probability of malignancy based on the patients risk of breast cancer is rarely factored into the analysis by the interpreting radiologist. They may use the imaging features to hazard a guess. But if the patient has a strong family history, gene mutation imparting very high risk or a prior benign breast biopsy showing a high risk lesion, the prior probability of malignancy would change. Furthermore, the radiation risk of PET and Scintimammography may not be "worth" imparting to gain specificity. I think that suggesting that adding layers of additional tests to improve specificity will result in use of even further additional testing of new false positive findings detected by the supplemental techniques and add layers of cost and additional biopsies. This is a report done without a true understanding of the strengths and weaknesses of breast imaging.	We have modified text throughout the document to clarify the difficulty of precisely estimating the probability of malignancy for an individual woman. We have added additional information about the radiation exposure from these tests.

Commentator & Affiliation	Section	Comment	Response
Public Reviewer #2	Executive Summary	This report, "The Effectiveness of Non-invasive Diagnostic Tests for Breast Abnormalities – An Update to the 2006 Report," is a disappointment and a disservice to breast care in the United States. The authors do not seem to understand how breast imaging is conducted. The probabilities, which are the basis of the report, do not take into consideration the actual radiographic appearance of the abnormality. All masses are not alike, nor are all calcifications, so that lumping them together into a probability range cannot reflect what should and does occur in practice. Nuclear medicine techniques and/or breast MRI should not decide the treatment plan for calcifications, regardless of calculated "probabilities". These techniques are not used this way – for good reason. For example, the authors fail to consider the radiation dose to the critical organ for nuclear medicine studies, which is not the breast for either BSGI or PEM. It is not reasonable to give a significant radiation dose to the bladder or colon in order to push a "number" down (ie, probability) to avoid a biopsy. This brings up another false underlying premise of the report: that it is more important to avoid a minimally invasive needle biopsy than to diagnose early breast cancer. For example, a very common scenario is that of a hypoechoic mass that may be a complicated cyst. This can easily be determined by aspiration with a very small 20 gauge needle. This report suggests that PEM, BSGI, or MRI would be preferred. All three of these studies are at least as invasive (all require a needle injection), PEM and BSGI involve significant radiation dose to the patient, and MRI subjects the patient to gadolinium contrast. How is this better? This report will do more harm than good. It should be replaced by a thoughtful analysis by individuals who understand the nuances of breast imaging and breast care. As it stands, this report is misguided due to lack of actual practical knowledge by the authors. Good science cannot occur in a vacuum. I strongly urge the AHRQ to withdraw this report until a clinically relevant document can be constructed.	We have modified text throughout the document to clarify the difficulty of precisely estimating the probability of malignancy for an individual woman. We have added additional information about the radiation exposure from these tests. We have modified the text to clarify that our primary conclusion is that most women won't benefit from the addition of these tests to their work-up for the purpose of distinguishing benign from malignant lesions. We have found that some women might benefit (if it were possible to identify) them based on a very low (1% - 12%) probability of malignancy prior to these additional tests.

Commentator & Affiliation	Section	Comment	Response
Public Reviewer #4	Executive Summary	PPV2 and 3 would be better metrics than sensitivity since false negatives are difficult to determine and would be closer to clinical reality.	We agree predictive values are more familiar to clinicians; however, predictive values vary as the prevalence of disease changes, and therefore providing one single estimate of PPV or NPV is misleading. An additional problem with presenting the results in terms of predictive values is that the studies included in the CER generally had much higher prevalences of disease than would be expected in a typical clinical setting, suggesting that predictive values calculated from this evidence base would not be directly applicable to many clinical settings. For these reasons, we have not added PPV data to the document. We did calculate the positive likelihood ratios and used those to illustrate post-test probability of disease across a range of pre-test probabilities. We have text in the document explaining the rationale for not using predictive values.
Peer Reviewer #2	Executive Summary	P. 7 Line 52. Define “standard” work up. In the background section it states “In current standard practice the examinations conducted after recall usually consist of diagnostic mammography and possibly ultrasound (to identify cysts).” This is a key problem with the review as ultrasound is included in the list of imaging tests reviewed, although ultrasound is often a part of “standard” work up. This continues to be a problem throughout the review as most of the advanced imaging techniques were performed after diagnostic mammography and ultrasound were performed, yet the results are presented as though all imaging tools, including US, were compared to each other after recall from screening mammogram or clinical exam alone.	Yes, this is a complex issue to address. The majority of studies were not entirely clear about how many patients received which tests before undergoing additional imaging. Some or many of the patients may have undergone ultrasound before being imaged by MRI, PET, scintimammography, or diagnostic/additional ultrasound imaging. For example, most of the studies (including those evaluating various types of US) did not enroll patients found to have simple cysts on an initial ultrasound examination, which we considered to be a reasonable criterion reflective of normal clinical practice. We have added text to clarify the issue. We have tried to use a fairly non-specific term “standard workup” throughout to encompass variations in clinical practice from center to center and variations in clinical practice triggered by specific characteristics of individual patients. We have added some text to clarify that the patient population of interest is women who have already undergone standard work-up after recall, a work-up that is likely to include diagnostic mammography and ultrasound.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #4	Executive Summary	Throughout abstract, give strength of evidence for sensitivity first, then for specificity. Since cancer detection is the primary goal, specificity becomes secondary, especially in the diagnostic setting. Doppler is not a separate examination per se, but rather may be done in addition to B-mode ultrasound. I cannot find the key questions on page 18 [comment in the table of contents] Needs reference. Add reference to BI-RADS: D'Orsi CJ, Bassett LW, Berg WA, et al. *Breast Imaging Reporting and Data System, BI-RADS: Mammography, 4th edition*. Reston: American College of Radiology, 2003. Ultrasound is used to identify abnormalities, characterize them, and guide biopsy. Many solid masses have a typically benign appearance on ultrasound: characterization goes far beyond "identifying cysts". what about positron emission mammography? need some detail about size of lesions included in study....Whole body PET is not sensitive to small cancers. What about positron emission mammography, which is designed for breast imaging? should be stated by indications--palpable mass vs. mammographic abnormality vs. nipple discharge. most studies focus on a particular clinical problem--e.g. palpable lump, nipple discharge, mammographic mass, mammographic calcifications. It would be much more pertinent to focus this analysis by problem rather than grouping all testing under one rubric. what else was considered? Was elastography considered? what is meant by "aggressive diagnostic testing"? This seems very opinionated. A needle biopsy under ultrasound is a very minor procedure.....	The strength of evidence was only rated differently for these two outcomes for MRI, and they are already in the order suggested by the reviewer. While it is true that Doppler is not a separate examination, the included studies reported data from the Doppler part of the ultrasound examination separately. The Key Questions are listed on p. 18; we believe the reviewer may have been using the page numbers provided by Adobe (which start counting from the title page) rather than the page numbers printed in the document, which start counting after the executive summary. We did not insert references into the executive summary because current AHRQ style is to provide the references in the body of the report rather than creating a separate reference list in the Executive Summary. All of the statements in the executive summary that require supporting references are derived from the body of the text and are referenced therein. We have modified the text about using ultrasound to identify cysts. No studies of positron emission mammography (PEM) met the inclusion criteria, as is stated in the document; we have added additional text to clarify this point. Very few of the included studies of PET reported the size of the lesions. The studies generally did not report data separately for palpable vs. non-palpable, and therefore the estimates of accuracy cannot, in most cases, be reported separately for different types of lesions; in the body of the text, where possible, we did report estimates of accuracy by type of lesion. Elastography was not considered, as is stated in the body of the text. We have modified the text to remove the phrase "aggressive diagnostic testing."
Public Reviewer #5	Introduction	Key question #2 is the most important in terms of deciding which test to use when but was not addressed to any useful degree in this report	We agree; unfortunately, the currently available data precludes a detailed evaluation and answer(s) to Key Question #2. Additional research needs to be conducted and published before Key Question #2 can be satisfactorily addressed. This problem is discussed in the Future Research section. Identifying gaps in knowledge is an important function of systematic reviews.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #1	Introduction	The key questions were clearly stated but were not adequately addressed by this analysis. Grouping questionable findings on screening mammography with palpable abnormalities may not be valid as the populations and the prevalence of malignancy may be very different in these different clinical settings. A very important aspect in terms of clinical applicability is key question 2, namely, when are these tests most appropriately used. This study did not adequately answer that question.	The majority of the studies enrolled both patients with palpable abnormalities and patients with screening mammography-detected abnormalities, and did not report the results separately for either group. Therefore unless we decided to exclude this type of study (the majority of the studies), we had no choice but to group these patient populations together in a single analysis. Where results were reported separately for each population of patients we did analyze results for the two populations separately. We did not add any additional text. We agree that Key Question 2 is of great importance; however, the currently available data precludes a detailed evaluation and answer(s) to Key Question #2. Additional research needs to be conducted and published before Key Question #2 can be satisfactorily addressed. This topic is discussed in the Future Research section. Identifying gaps in knowledge is an important function of systematic reviews.
Peer Reviewer #2	Introduction	P.12 line 51 “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 recommended follow up for a probably benign lesion on mammography includes one additional xray of the area in question for women in annual screening mammography program and two additional xrays of the area in question for women undergoing every two year mammography. The term “frequent repeat” mammography exams could be replaced by “additional”	We have made the suggested change.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #3	Introduction	Should there be mention of the factor of heterogeneity as regards patient preference? The link to the reference cited for survival statistics (7) is not correct - a server error message appears. There is some evidence that survival after stage IV diagnosis may be improving above the number quoted and it might be mentioned that these statistics (by nature) may not apply to a woman diagnosed today. For example, a CBCRP report cites an MDAnderson report: A recent study from M.D. Anderson Cancer Center13 that compared length of survival of metastatic breast cancer patients treated at their institution in five-year increments, found that median survival had doubled to 51 months (range 33-69 months) in 1995-2000 from a median survival of 27 months (range 21-33 months) only five years earlier, 1990-1994. Five years after their diagnosis with metastatic disease, 40 percent of these patients were still alive, as compared with 29 percent during 1990-1994. At the initiation of their study, during the period 1974-79, only 10 percent of patients were still alive at five years and the median survival was only 15 months (range 11-19 months). http://www.cbcrp.org/publications/papers/mayer/page_03.php In discussing lobular cancers, should there be any mention of the difficulties and differences known to exist as far as detecting them with imaging?	We have replaced reference 7, which refers to data from 2008, with an updated reference that refers to data from 2010, and updated the numbers about survival rates in the text. There is a considerable body of literature discussing lobular cancers and difficulties in detecting them on screening; however, because none of the included studies distinguished between types of cancers diagnosed we have not discussed the issues in detail.
Peer Reviewer #4	Introduction	most “papillary” carcinoma is DCIS. Invasive micropapillary carcinoma is aggressive and is now listed below. what about DCIS? quantify the rate of short-term follow-up after screening recall/workup. again, question the tone here. An ultrasound-guided needle biopsy is a very minor procedure--much less healing involved than from skin biopsies for moles that could be melanoma.....but much greater likelihood of being malignant. based on what? The risk of complications from needle biopsies is very low, based on AHRQ's own report. The risk from even one delayed diagnosis of breast cancer could be great.	The text in question is referring specifically to histological subtypes of invasive ductal carcinoma, not to DCIS and not to papillary carcinoma, which is discussed below the text in question. DCIS is discussed above the section marked with this comment. The rate of short-term followup after screening recall/workup varies across centers and countries. The text in question has been modified. We agree that the risk of complications from needle biopsies is low, although not zero.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #5	Introduction	The following comments; Page 25 “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” and Page 80, “The prevalence of cancers in the general population sent for breast biopsy (in the USA) has been reported to be around 23%.17”, are supported by a study done 12 years ago. More recent data from close to 200 facilities and 1.1 million women participating in the Breast Cancer Surveillance Consortium (BCSC) from1996 to 2002, funded by the National Cancer Institute, show that biopsies revealed 34% malignancies, of which 78.4% were invasive cancers and 21.6% were ductal carcinoma in situ (Performance Benchmarks for Screening Mammography. Rosenberg RD et all. Radiology October 2006 241:55-66; doi:10.1148/radiol.2411051504) Page 25 line 17: “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.” This statement is unsupported. Minimally invasive needle biopsy has an exceedingly low morbidity rate and almost no associated mortality risk. Non-invasive test such as MRI have unknown long term risk of Gd exposure and PET, BSGI and MIBI have associated risk from radiation exposure that are not minimal.	We thank the reviewer for providing this newer information and have added it to the report. After consideration, we have to agree with the reviewer that risks from radiation exposure may not be safer than core needle biopsy, and have deleted the statement.

Commentator & Affiliation	Section	Comment	Response
Public Reviewer #4	Methods	Since all the studies started with a “High” status how did all except MRI end with a low status. If “consistency unknown” was this marked as low. How is consistency weighted with regard to bias risk?	The rating system is described in detail in the Methods section, with additional details provided in the referenced Methods Guide and the Appendices. All of the studies started with a high internal validity rating (low risk of bias); however, the rating of strength of evidence incorporated other aspects of the evidence base, including consistency, precision, and directness. Many of the evidence bases were found to be inconsistent and imprecise, leading to the “Low” strength of evidence ratings. We have text in the document explaining how the strength of evidence ratings were arrived at. Consistency and risk of bias are independent concepts. Risk of bias refers to aspects of individual study design; consistency refers to how similar the data is across all of the studies. The rating of “consistency unknown” was only applied to evidence bases with a single study. Evidence bases rated as “consistency unknown” and “precise” were given an overall Low strength of evidence rating; if “consistency unknown” and imprecise, the evidence was rated as Insufficient to support an evidence-based conclusion. We have text in the document explaining these concepts.
Public Reviewer #5	Methods	It is unclear how the choice for which non-invasive tests to evaluate was made. This review does not reflect actual clinical practice. PET and PET/CT are never used in the setting of an abnormal screening mammogram or questionable palpable finding. Whole-body scintimammography is not used to any degree in this country for any indication. It does not seem reasonable to take a number of studies done with different populations and designed to evaluate different aspects of test performance, aggregate the results, then apply those results to the abnormal screening setting. Under the heading Applicability, it is stated that studies that enrolled patient populations that were not a general population of asymptomatic women participating in routine breast cancer screening programs were excluded. However, many of the studies were those conducted on women with suspicious imaging and/or clinical findings who were already scheduled to undergo biopsy. That is a very different population from women who have findings for whom the most appropriate management is unclear. Therefore, the applicability of the results of this report to the situation stated in the introduction is doubtful.	We have added additional text to the Methods section to further explain the choice of technologies. The fact that the majority of studies only enrolled women scheduled for biopsy, and had prevalences of disease that were much higher than would be expected in the intended patient population is indeed a serious limitation of the findings, as discussed in the Applicability part of the Discussion section; we have added text to the Executive Summary to clarify this important point for those readers who only have access to the Executive Summary.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #1	Methods	It is unclear why the imaging tests that were chosen were included. PET and PET/CT are never used for evaluation of suspected abnormalities on screening mammography. The inclusion criterion for MRI studies that required that an 8 channel breast coil was used, was not followed and the inclusion of this criterion serves to emphasize the fact that the authors of this study have little basic understanding of the technology they are attempting to assess. As was pointed out to them in an early conference call, 8 channel coils are not necessary to obtain high quality breast MRI examinations. Field strength is a more important determinant than the number of channels and studies that used a 0.5T magnet should not have been included as they do not reflect current clinical practice. In addition, 8 channel coils were not developed until 2004-2005, yet MRI studies from 2000 were considered. Finally, most of the cited MRI studies do not specify the number of channels in the breast coil, yet were included nevertheless. For sestamibi studies, I am not aware of any facility in the country that is using whole-body MIBI examinations for evaluation of breast abnormalities, yet these studies were included. I am unclear as to the validity of using studies that examined very different populations and then using mathematical modeling to apply results to the management of questionable findings from screening.	We have added text to clarify the choice of technologies for review. We thank the reviewer for noticing the discrepancy. We had originally proposed the use of 8 channels as an inclusion criterion, but after consultation with the Technical Expert Panel that criterion was dropped. The data abstraction forms in the Appendix contain the actual inclusion criterion that was used, namely, any dedicated breast coil with no requirement as to number of channels. However, the statement about 8 channels in the Methods section was over-looked and was not updated. We have corrected the error. The American College of Radiology guidelines on breast MRI imaging state: The selection of field strength is a major technical decision. A 1.5 Tesla magnet has traditionally been considered a minimum technical requirement because of the relationship between field strength and resolution. However, improvements in other components of the scanning process have resulted in improved scan quality at lower field strengths. Only two of the included studies used 0.5T magnets. Usually 0.5T magnets are used with open-bore systems. Patients are often interested in using open-bore systems rather than closed systems, and some manufacturers of newer open-bore 0.5T devices claim they are as accurate as older 1.5T devices. Therefore we cannot justify excluding studies of 0.5T devices at this time; information about the accuracy of 0.5T devices is important information for the clinician to have.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #2	Methods	The statistical methods are very sound and appropriate. The search strategies are logical. The definitions for outcome measures are appropriate, but the inclusion and exclusion criteria for patient populations are not clear and specifically fail to address which combinations of imaging have already been performed prior to the "added imaging test" evaluated. Throughout the document when the generic term is used for computer aided diagnosis the term CAD not CADx should be used. CADx P. 12 Line 28 the authors state "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. " With the notable exception of the ultrasound studies cited, most of the studies include a patient population that has had both diagnostic mammography and ultrasound. The presentation of the results does not make this clear and leaves one with the assumption that each of the tools can be compared to each other (eg, ultrasound compared to MRI) regarding their addition to mammography alone	The majority of included studies were not entirely clear about how many patients received which tests before undergoing additional imaging. The term CADx has been used to specifically refer to computer-aided diagnosis, as distinguished from CADe (computer-aided detection), and CADq (computer-aided quantification of tumor volume), and from the generic CAD (which, in the context of breast imaging, usually refers to the use of computer systems to assist in reading screening mammograms, a process some might refer to as CADe). Many of the studies of diagnostic ultrasound had also performed preliminary ultrasound examinations to rule out diagnoses such as simple cysts before enrolling patients. Additional text has been added clarifying this point.
Peer Reviewer #3	Methods	The paragraph about the harm of false negative vs. false positive results (pg. 39, line 26) seems a little too strongly-worded on the side of the harms of false negatives. Yes, they perhaps are of more "clinical relevance", but the sentence that includes "women may die from delayed diagnosis" seems emotional and not in the same distanced tone as the rest of the report. Perhaps wording using "morbidity and mortality", or "decreased survival" would fit in better?	We appreciate the reviewer noticing this point, and have made the suggested change.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #4	Methods	Women with symptoms/lumps are no longer considered "routine screening". This is a problem throughout this document. prevalence of disease is quite different for screening vs women with palpable lump--need to keep these separate. again, these two populations are very different in terms of disease prevalence: would keep them separate seems that diagnostic mammography was overlooked in all this. PET is not something we would recommend for further evaluation of a breast abnormality. It can be used for whole body staging in women with newly diagnosed cancer, as can PET-CT. PEM (positron emission mammography) can be used for evaluating breast abnormalities but there isn't much literature on it, and it doesn't appear you all looked at PEM. BSGI is different from whole body scintigraphic breast imaging and should be distinguished throughout For Q3 throughout, you might focus on acceptability, but would not repeat accuracy issues here. spell out also: radiation risk, tolerability of the examinations would also list breast CT, dual-energy subtraction contrast-enhanced mammography. disagree re: elastography no one would consider whole body PET or PET-CT for further evaluation of a mammographic or clinical abnormality should exclude SPECT--not current	We did not include studies of women with overt symptoms such as nipple discharge; we did include studies that included women with palpable lumps, because a significant proportion of breast cancers are only diagnosed after a woman or healthcare provider identifies a lump on routine examination (screening of asymptomatic women). The studies generally did not report data separately for palpable vs. non-palpable, and therefore the estimates of accuracy cannot, in most cases, be reported separately for different types of lesions; in the body of the text, where possible, we did report estimates of accuracy by type of lesion. Diagnostic mammography is the standard of care. The scope of the report is "What else" can be added to the standard of care to aid in the distinction between benign and malignant, and is the addition of the proposed extra test beneficial? No studies of PEM met the inclusion criteria; text has been added to clarify this point. We did distinguish between BSG1 and whole-body scintimammography. Changing the exact wording of the Key Questions at this stage in the project is not a usual procedure for systematic review development. The issues mentioned are addressed in the report, but the exact wording of the Key Questions was developed following a systematic process, including posting of the questions for public comment. Breast CT and dual-energy subtraction contrast-enhanced mammography were not at any point proposed to be included in the scope of the review. Elastography should be considered for inclusion in future systematic reviews on this topic. We identified a number of published studies that were using PET or PET/CT to further evaluate mammographic or clinical abnormalities. We identified only one study that used SPECT; this technology does appear to be fading out of current use.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #5	Methods	As the authors admit, many of the cited papers involve studies of women with considerably higher than average risk for breast cancer. Example: Page 113 Table 4 MRI papers: 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. This is listed as 960 patients and 960 lesions (which I believe is incorrect, 821 patients). The authors themselves admit "The positive predictive value of mammography was high (52.8%)... this suggest that patients with more advanced breast lesions were referred into the trial." Because most of the cited papers involve higher than average risk patients, application of this data to findings on routine screening in normal risk women is difficult.	We have corrected the error in Table 4. We have added additional text to the executive summary to draw further attention to the possible limitation in the applicability of the results.
Peer Reviewer #6	Methods	Poor Methodology- simply lumping all studies together and analyzing sensitivity and specificity does not take into account the specific clinical setting in which a modality is used- eg PET not used in a screening setting	The study inclusion/exclusion criteria ensured that all of the technologies evaluated were being used in the same specific clinical setting - to examine women who had abnormalities already detected on screening. There is text in the document explaining the choice of studies and methods used; no additional text was added.
Public Reviewer #5	Results	How was "pre-test probability of the lesion being malignant" determined? It is rare that the probability of malignancy can be so precisely defined. Therefore, these results are of no practical clinical use.	We have modified text throughout the document to clarify the difficulty of precisely estimating the probability of malignancy for an individual woman.
Peer Reviewer #1	Results	Describing the probability of malignancy given a negative or positive result on imaging based on probability of malignancy prior to the test is not clinically useful. The method of determining the pre-test probability of malignancy was not stated and in actual practice, is rarely so precise as is outlined in the results. For example, it is usually impossible to say whether a particular finding has a 5% versus a 10% chance of malignancy. Therefore, the results of this report are of no clinical utility. Also, different imaging tests have different uses depending on the appearance of the abnormality seen on the screening mammogram. This was key question #2 but was never addressed by this report.	We have modified text throughout the document to clarify the difficulty of precisely estimating the probability of malignancy for an individual woman. We discuss the lack of evidence addressing Key Question 2 in the Future Research section.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #2	Results	The study inclusion was extensive. The patient population and characteristics of the studies reported lacked the key variables of 1) specifically which imaging tests or tissue sampling had been performed PRIOR TO performance of the imaging assessed in the study and 2) were patients symptomatic or asymptomatic.	For the most part the studies were unclear as to how many patients had which tests prior to additional imaging, and were also often reporting on combinations of patients with palpable and non-palpable abnormalities. Where possible we performed analyses of the different patient populations separately.
Peer Reviewer #3	Results	Information is clearly presented, detail seems appropriate. As a relative lay person, I was somewhat confused by the lack of consistent use of terms and percentages about benefit across the different sections (executive summary and results). At the risk of redundancy, it might help to use the same phrases throughout, using both percentage and term. For example those at “real but low risk” of malignancy (results section) -- is that the same as “approximately 2 to 20% suspicion of malignancy” group noted in the executive summary? I think that it is, but am not sure, and although a clinician would catch on more quickly than me, a rushed clinician (aren't they all that?) might appreciate the clarity. This is nicely summed up in the discussion segment, but again, for those just dipping in and in a rush, consistency over each section might be helpful.	We have made changes to improve the consistency in terminology across different sections.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #4	Results	This table really has no relevance to clinical practice. We don't routinely recommend MRI for further evaluation of a mammographic or clinical abnormality except in very particular circumstances: nipple discharge and negative mammogram and ultrasound; vague new asymmetry on mammogram and not able to perform biopsy. We would never recommend MRI for further evaluation of most abnormalities--we would either biopsy them or perform short-term follow-up mammography (which is far less expensive).... usually 10 minutes per view and minimum of two views per breast for PEM this has not been the case with dedicated breast PET (positron emission mammography, or PEM) this is just not true---neither whole body PET or PET-CT have any utility in further workup of mammographic abnormalities--there is very sparse data on nonpalpable abnormalities, and what data do exist show very low sensitivity to smaller cancers. I would delete this table as we just don't use PET or PET-CT this way due to low sensitivity to small cancers. this paragraph reads almost like an advertisement? this is for the older whole body studies. Current studies nearly always use dedicated systems, with positioning analogous to mammography and gentle stabilization of the breast being imaged. It doesn't make sense to detail the older technology/positioning only. radiation dose issues need to be discussed	The nomination of this topic by CMS suggests that women are being sent for additional imaging during work-up increasingly often, and MRI is one of the most frequently recommended additional examinations suggested by some health-care providers. It is important to evaluate the evidence to see if women do or do not benefit from the addition of these imaging modalities to the standard workup for breast abnormalities. Our references indicate 5 minutes per view, but perhaps it varies from device to device. Yes, the statement was about whole-body PET, not about PEM. Yes, the data do not support the routine use of PET or PET/CT as part of the standard workup to distinguish benign from malignant lesions unless it is possible to assess prior to the test that a woman has a 1-5% probability of malignancy. We identified a number of published studies that were using PET or PET/CT to further evaluate mammographic or clinical abnormalities, so apparently some researchers have considered using this technology for this purpose. It is important to evaluate the evidence to see if women do or do not benefit from the addition of these imaging modalities to the standard workup for breast abnormalities in terms of distinguishing benign from malignant lesions. The paragraph in question is in fact describing the manufacturer's claims: "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." (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.) The majority of the studies (all published recently) are not using dedicated systems, and therefore describing how they performed the imaging is essential. Radiation dose issues were not reported by any of the included studies, and therefore were not discussed as part of the answer(s) to the key questions; however, we have added additional information about radiation doses and other possible safety hazards as part of the Background information for each technology.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #5	Results	The use of the presented data in terms Negative likelihood ratios is NOT clinically meaningful for several reasons. For possible findings found at routine screening mammography, current practice involves evaluation of these recalls using diagnostic mammographic views and or ultrasound. The majority of these questionable screening findings turn out to be either normal superimposed tissue or benign entities such as cysts requiring no further action. Lesion that are deemed suspicious (>2% chance of malignancy based on BI-RADS classification established by Sickles et al) are recommended for needle core biopsy. There are currently no standards or sufficient data for the precise estimation of pretest probability of these lesions based on diagnostic mammography and or ultrasound features. There is even less data on estimation of a lesions pre-test probability of malignancy based solely on the findings at routine screening mammography and therefore additional non-invasive tests such as PEM, MRI, BSGI and MIBI at that point would be inappropriate. The problem with the use of Negative likelihood ratio is that it requires a narrow estimation of a lesions pre-test probability of malignancy based on the initial mammographic findings. There is no data to support that this can be done accurately and current practice is based on BI-RADS categories that fall only into broad estimations (BI-RADS 4 >2% and less than 95%, BI-RADS 5 >95%)	We have modified text throughout the document to clarify the difficulty of precisely estimating the probability of malignancy for an individual woman.
Peer Reviewer #6	Results	How does one assign a cancer risk of 5%? 20%? Such numbers are impossible to determine in clinical practice and should not be used in the report.	We have modified text throughout the document to clarify the difficulty of precisely estimating the probability of malignancy for an individual woman.
Public Reviewer #3	Discussion	There must be national accreditation for all imaging modalities, and staff. Studies are performed by radiologists. In practice, the studies are executed by technologists. Without uniformity and standardization of practices and equipment, including computer technologies utilized in interpretation, the outlook for quality evaluation and treatment is dismal.	We agree, a requirement for national accreditation would most likely improve the consistency and quality of diagnostic imaging; however, because this topic is out of scope of the report, we have not made any changes to the document about the benefits of national accreditation.
Public Reviewer #4	Discussion	Surrogate endpoints for cancer [re: future research needed]	We are unsure what the reviewer is referring to here. Perhaps you are suggesting that a substitute for long-term patient oriented outcomes like survival and quality of life is needed. Surrogate outcomes are only useful if proven to be tightly linked to patient-oriented outcomes.

Commentator & Affiliation	Section	Comment	Response
Public Reviewer #5	Discussion	The statement that "the use of non-invasive imaging in addition to standard workup may be useful for women with a low (2 to 20%) suspicion of malignancy" is not adequately supported by the analysis presented in this report and is potentially harmful. The report suggests that it is more important to avoid a minimally invasive needle biopsy than to diagnose early breast cancer. For example, a very common scenario is that of a hypoechoic mass that may be a complicated cyst, which can easily be resolved by aspiration with a very small 20 gauge needle. This report postulates that PEM, BSGI, or MRI would be preferred. All three of these studies are at least as invasive as needle biopsy (all require a needle injection), PEM and BSGI involve significant radiation dose to the patient, and MRI subjects the patient to gadolinium contrast. The potential harm from the suggestions in this report are not appropriately addressed.	The clinical situation the reviewer describes actually correlates very well with the findings of the report- the patient is identified on screening; undergoes additional non-invasive imaging (ultrasound); the findings of the additional non-invasive imaging suggest the lesion is non-malignant. Our findings do not support the use of PEM, BSG1, or MRI in this scenario. We have added additional information about radiation exposure and risks from contrast agents to the document.
Peer Reviewer #1	Discussion	In the discussion, the authors themselves state that calculations based on studies with a high proportion of malignant cases may not be generalizable to other populations. Therefore, the conclusion reached by the authors, that non-invasive imaging may be acceptable to avoid biopsy in cases where there is a low probability of malignancy is highly questionable. The goal of decreasing biopsies that prove to be benign is an admirable one but it should be recognized that with minimally invasive image guided biopsy techniques, particularly ultrasound guided cores, the morbidity and cost are low and for many women is more acceptable than having yet another test that does not give as conclusive results as are offered by tissue diagnosis. In addition, it should be recognized that the cost of some of these imaging tests, particularly PET, is equal or greater than that of an image guided biopsy. Also, the radiation dose associated with PET and MIBI studies is not inconsequential and should be taken into consideration before these tests are used.	The fact that the majority of studies only enrolled women scheduled for biopsy, and had prevalences of disease that were much higher than would be expected in the intended patient population is indeed a serious limitation of the findings, as discussed in the Applicability part of the Discussion section; we have added text to the Executive Summary to clarify this important point for those readers who only have access to the Executive Summary. Clinicians and women should take all available information and personal preferences into account when making decisions. The findings of our report should be only one factor that is considered when deciding upon the best management plan for each individual woman.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #2	Discussion	The conclusion that would impact and possibly change current clinical practice is to suggest "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." The suggestion is either US or MRI could be useful for women with a low suspicion of malignancy. There is lack of clarity on "standard" work up and the studies reviewed have a heterogeneous population. Most importantly, many of the MRI studies INCLUDED patients who had ultrasound as well as mammography. This is not clarified in the report. The authors are clear regarding the limitations, stating "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." However, their conclusions and discussion do not match the key limitations to the study particularly given this is a subset of patients that were overwhelmingly a patient population recommended for biopsy.	The issue of prior ultrasounds has been addressed in the points above. The issue of applicability was addressed separately and was not incorporated into the generally low ratings of the strength of evidence. Clinicians and patients considering the use of the information in this report should consider the ratings of strength of evidence and the limitations of applicability when interpreting the results. Often one of the more important findings of systematic reviews is the gaps in the available evidence, gaps that need to be addressed by future research.
Peer Reviewer #3	Discussion	The summary is excellent. The future research section is also clear, although if the ideal RCT design is "admittedly < . . . > logistically difficult to conduct", can we add more detail about what could realistically be done to improve research?	We have added text about the use of modeling to provide information about the impact of diagnostic testing strategies on patient outcomes.
Peer Reviewer #6	Discussion	Unless non invasive testing can accurately determine a lesion to have a 2% chance or less of being malignant minimally invasive breast biopsy must remain the gold standard to rule out cancer	The evidence does not appear to support the routine use of additional non-invasive imaging during workup for screening detected abnormalities unless it is possible to identify the women with a prior suspicion of malignancy that is in the range of >0% but < 5, 10, or 12%, depending on the imaging technique being considered.

Commentator & Affiliation	Section	Comment	Response
Public Reviewer #5	Conclusion	The goal of this review is an important one. Unfortunately, this review did not produce any clinically meaningful results. It is unlikely, and rightfully so, to be used by breast imagers to determine best practice.	We believe that our finding that only women thought to be at very low risk of malignancy (1% – 12%) may benefit from additional non-invasive imaging to distinguish benign from malignant lesions is clinically relevant. The evidence does not appear to support the use of additional non-invasive imaging during workup for screening detected abnormalities for most women, which should be useful information for clinicians, patients, and payers. We have added additional text throughout to clarify our findings and to emphasize that the feasibility of estimating pre-test probability in this range is uncertain.
Peer Reviewer #2	Conclusion	The basic conclusions, based on the current literature, that available nuclear medicine tests (PET and scintimammography) and advanced ultrasound (dopper, elastography) do not have a role in clinical management of patients recommended for biopsy. However, the suggestion that US does not likely play a role in women recalled from screening is not possible from the methods of the research studies reviewed. Finally, the suggestion that there may be a role for MRI in women at low suspicion for malignancy in lieu of recommended biopsy is misleading and confusing.	We concluded that MRI is not indicated as part of a standard workup for the majority of women recalled for further investigation for the purpose of distinguishing benign from malignant lesions. We have modified the text to clarify what our major conclusions are.
Peer Reviewer #1	General	This report is clinically meaningless. It ignores the clinical context in which these tests are used and has attempted to introduce mathematical objectivity to subjective clinical practice. In doing so, the report reached clinically irrelevant conclusions that are confusing and that have the potential of misleading clinicians as to the appropriate use of these imaging tests. If the results of this report are actually adopted into clincal practice, it could result in inappropriate patient management with overutilization of imaging tests in some cases with the potential for inappropriate avoidance of biopsy. The report is unlikely to be read and used by clinicians. It is too long and too technical yet the results are too simplistic and the categorization of lesions by probability of malignancy is not as precise in actual practice as is reflected in this report. Radiologists, who largely comprise the target audience for this report, will readily recognize the lack of clinical applicability.	We have modified text throughout the document to clarify the difficulty of precisely estimating the probability of malignancy for an individual woman. Clinicians and women should take all available information and personal preferences into account when making decisions. The findings of our report should be only one factor that is considered when deciding upon the best management plan for each individual woman. Standard practice is for the findings of CERs to be translated into concise patient and clinician guides, to disseminate the findings to persons who do not have the time or desire to read the full-length report.

Commentator & Affiliation	Section	Comment	Response
Peer Reviewer #2	General	The key clinical question this review thus addresses is not the appropriate imaging work up after a recall from an abnormal screening mammogram, but rather whether or not additional imaging is useful once "standard" work up has been completed and the patient is assessed as suspicious, and thus warranting biopsy. I agree with the overall conclusion based on the literature that MRI, PET, scintimammography, Doppler US have low contribution to further evaluation prior to biopsy, but the clinical question is whether these tools have a role to replace a recommendation for biopsy of a lesion identified as suspicious using "standard" techniques of clinical exam, mammography and often ultrasound, not whether these tests are useful in further evaluation of a woman after recall from a suspicious lesion identified on routine screening. The results from studies of a patient population that has had screening and diagnostic work up with mammography and, in many cases of the studies reviewed ultrasound, cannot be expected to apply to all patients recalled after an abnormal screening exam. The target population is not explicitly defined. The key clinical question is not clear. The authors move from clinical management of women recalled from screening to women recommended for biopsy after imaging work up including mammography and ultrasound to women with symptoms to women without symptoms to women referred for short interval follow up after full diagnostic work up. Additionally, these diverse populations are mixed in the literature review as well as women who have undergone full diagnostic work up including biopsy and then have additional imaging. The patient populations in the literature reviewed are far to diverse and heterogeneous to be applied to recommendations for asymptomatic women recalled from a screening mammogram.	The patient population of interest does not change throughout the report, and it never includes women who have undergone biopsy before additional imaging. The changes to the text clarifying the patient population made in response to points above also address the point.
Peer Reviewer #5	General	My major concern regarding this study is that the key question stated at the top of page 30 "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)?" seems to contradict	Yes, we anticipate that these tests would be used after diagnostic mammography and/or ultrasound. Where these technologies (chiefly MRI and diagnostic ultrasound) are already being used in clinical practice they are almost always only used after a diagnostic mammogram has been performed. All of the included studies enrolled patients who had already been evaluated by diagnostic mammography and/or ultrasound. We have modified text throughout to clarify this point.

Commentator & Affiliation	Section	Comment	Response
		statements and figures presented elsewhere in the study. From the wording of this key question, “after identification... on routine screening” would imply that the use of the studied non-invasive diagnostics test would occur before diagnostic mammography and grayscale ultrasound. In practically all of the cited papers, inclusion of patients occurred after diagnostic mammography and ultrasound evaluation of lesions found at screening mammography, and not straight from screening. The comments on page 13 “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 noninvasive imaging studies that can be conducted after discovery of a possible abnormality on screening mammography or physical examination, studies intended to guide patient management” would imply the use of these non-invasive diagnostic tests directly after screening. However, the Figure 1. Analytical framework schema (page 27) indicates the non-invasive diagnostic test would occur after diagnostic mammographic and ultrasound work-up. Comments on page 24 “... If an available non-invasive diagnostic test could assist clinicians in evaluating women.... then many women could avoid frequent repeat mammography exams, with their attendant discomfort, inconvenience, x-ray exposure, and emotional distress.” Would again imply the purpose is to study the use of these non-invasive diagnostic tests directly after screening and instead of diagnostic mammographic views or standard ultrasound. Also, in reference to the “untoward” effects of repeat mammography, this comment does not seem logical due to the fact that the proposed alternative non-invasive diagnostic tests also involve emotional stress, discomfort (MRI positioning, Gd injection and time, BSGI and MIBI involve breast compression) and considerable radiation dose (BSGI, MIBI and PET). In general, in terms of sensitivity and specificity, I would say the results are clinically meaningful in that the presented data would support against the use of the studied non-invasive imaging tests over routine diagnostic work-up, ultrasound and minimally invasive core-needle biopsy.	In reference to the “distress” experienced by women referred for short-interval mammography evaluation, the distress seems to be mostly related to spending months not knowing if the lesion is malignant or not; if an additional test could be performed immediately after the first diagnostic mammogram, on the same day in the same center, and a diagnosis of “benign” were made, these women would not experience months of anxiety and distress. We have modified the text to clarify this point.