AHRQ Comparative Effectiveness Review Surveillance Program

CER #47:
Noninvasive Diagnostic Tests for Breast Abnormalities: Update of a 2006 Review

Original release date:
February, 2012

Surveillance Report:
January, 2013

Key Findings:
All of the key questions remain up-to-date.

Summary Decision

This CER's priority for updating is **Low**

Authors:
Jennifer Schneider Chafen, MS, MD
Sydne Newberry, PhD
Aneesa Motala, BA
Margaret Maglione, MPP
Susanne Hempel, PhD
Roberta Shanman, MS
Paul Shekelle, MD, PhD

None of the investigators has any affiliations or financial involvement that conflicts with the material presented in this report.

Acknowledgments

The authors gratefully acknowledge the following individuals for their contributions to this project:

Subject Matter Experts

Carol H. Lee, MD
Yale University School of Medicine
New Haven, Connecticut

Mark Robson, MD
Memorial Sloan Kettering Cancer Center
New York, NY

Contents

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

1. Introduction

Comparative Effectiveness Review (CER) #47, Noninvasive Diagnostic Tests for Breast Abnormalities: Update of a 2006 Review, was released in February 2012.[1] It was therefore due for a surveillance assessment in October, 2012. At that time, we contacted experts involved in the original CER and subject experts to get their opinions as to whether the conclusions had changed and need to be updated. We also conducted an update electronic literature search. Every month since the CER's original release until September 2012, we received any FDA updates on the included treatments and tests.

2. Methods

2.1 Literature Searches

Using the search strategy employed for the original report, we conducted a limited literature search. We screened PubMed for the time period January 1, 2010 to August 27, 2012. This search included five high-profile general medical interest journals (Annals of Internal Medicine, British Medical Journal, Journal of the American Medical Association, Lancet, and the New England Journal of Medicine) and three specialty journals (Breast, Radiology, and Clinical Radiology**).** The specialty journals were those most highly represented among the references for the original report. This search resulted in 161 titles / abstracts to review. Appendix A includes the search strategy.

2.2 Study selection

We used the same inclusion and exclusion criteria as the original CER. We screened the titles and abstracts and obtained full text copies of publications accordingly.

2.3 Expert Opinion

We shared the conclusions of the original report with 10 experts in the field (including the original project leader, all original technical expert panel (TEP) members and peer reviewers for their assessment of the need to update the report and their recommendations of any relevant new studies; 3 subject matter experts responded including the project lead. Appendix C shows the questionnaire matrix that was sent to the experts.

2.4 Check for qualitative and quantitative signals

After abstracting the study conditions and findings for each new included study into an evidence table, we assessed whether the new findings provided a signal according to the Ottawa Method and/or the RAND Method, suggesting the need for an update. The criteria are listed in the table below.[2, 3]

	Ottawa Method
	Ottawa Qualitative Criteria for Signals of Potentially Invalidating Changes in Evidence
A1	Opposing findings: A pivotal trial or systematic review (or guidelines) including at least one new trial that characterized the treatment in terms opposite to those used earlier.
A2	Substantial harm: A pivotal trial or systematic review (or guidelines) whose results called into question the use of the treatment based on evidence of harm or that did not proscribe use entirely but did potentially affect clinical decision making.
A3	A superior new treatment: A pivotal trial or systematic review (or guidelines) whose results identified another treatment as significantly superior to the one evaluated in the original review, based on efficacy or harm.
	Criteria for Signals of Major Changes in Evidence
A4	Important changes in effectiveness short of "opposing findings"
A5	Clinically important expansion of treatment
A6	Clinically important caveat
A7	Opposing findings from discordant meta-analysis or nonpivotal trial
	Quantitative Criteria for Signals of Potentially Invalidating Changes in Evidence
B1	A change in statistical significance (from nonsignificant to significant)
B2	A change in relative effect size of at least 50 percent
	RAND Method Indications for the Need for an Update
1	Original conclusion is still valid and this portion of the original report does not need updating
2	Original conclusion is possibly out of date and this portion of the original report may need updating
3	Original conclusion is probably out of date and this portion of the original report may need updating
4	Original conclusion is out of date

2.5 Compilation of Findings and Conclusions

For this assessment we constructed a summary table that included the key questions, the original conclusions, and the findings of the new literature search, the expert assessments, and any FDA reports that pertained to each key question. To assess the conclusions in terms of the evidence that they might need updating, we used the 4-category scheme described in the table above for the RAND Method.

In making the decision to classify a CER conclusion into one category or another, we used the following factors when making our assessments:

- If we found no new evidence or only confirmatory evidence and all responding experts assessed the CER conclusion as still valid, we classified the CER conclusion as still valid.
- If we found some new evidence that might change the CER conclusion, and /or a minority of responding experts assessed the CER conclusion as having new evidence that

might change the conclusion, then we classified the CER conclusion as possibly out of date.
- If we found substantial new evidence that might change the CER conclusion, and/or a majority of responding experts assessed the CER conclusion as having new evidence that might change the conclusion, then we classified the CER conclusion as probably out of date.
- If we found new evidence that rendered the CER conclusion out of date or no longer applicable, we classified the CER conclusion as out of date. Recognizing that our literature searches were limited, we reserved this category only for situations where a limited search would produce prima facie evidence that a conclusion was out of date, such as the withdrawal of a drug or surgical device from the market, a black box warning from FDA, etc.

2.6 Determining Priority for Updating

We used the following two criteria in making our final conclusion for this CER:

- How much of the CER is possibly, probably, or certainly out of date?
- How out of date is that portion of the CER? For example, would the potential changes to the conclusions involve refinement of original estimates or do the potential changes mean some therapies are no longer favored or may not exist? Is the portion of the CER that is probably or certainly out of date an issue of safety (a drug withdrawn from the market, a black box warning) or the availability of a new drug within class (the latter being less of a signal to update than the former)?

3. Results

3.1 Search

The literature search identified 161 titles. After title and abstract review, we further reviewed the full text of 15 journal articles. The remaining titles were rejected because they clearly did not meet inclusion criteria for any of the review questions. In addition to the electronic database searches, we followed up suggestions from the topic experts for studies not already included in the original report which resulted in two more articles

Thus, 17 articles went on to full text review. Fifteen articles were rejected because they did not meet the inclusion criteria of the original report. Two articles were included for Key Question 1 and are shown in the evidence table in Appendix B.[4, 5]

3.2 Expert Opinion

For the most part, the experts were not able to identify any new literature and believed that the vast majority of key questions were still valid.

3.3 Identifying qualitative and quantitative signals

Table 1 shows the original key questions, the conclusions of the original report, the results of the literature and drug database searches, the experts' assessments, the recommendations of the Southern California Evidence-based Practice Center (SCEPC) regarding the need for update, and qualitative signals.

Table 1: Summary Table

Conclusions From CER Executive Summary	RAND Literature Search	FDA/ Health Canada/MHRA (UK)	Expert Opinion EPC Investigator Other Experts	Conclusion from SCEPC
Key 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 identification of a possible breast abnormality on routine screening (mammography and/or clinical or self-detection of a palpable lesion)? The noninvasive tests to be evaluated are: Ultrasound (conventional B-mode, color Doppler, power Doppler, tissue harmonics, and tomography); Magnetic resonance imaging (MRI) with breast-specific coils and gadolinium-based contrast agents, with or without computer-aided diagnosis (CADx); Positron emission tomography (PET) with 18-fluorodeoxyglucose (FDG) as the tracer, with or without concurrent computed tomography (CT) scans; Scintimammography (SMM) with technetium-99m sestamibi (MIBI) as the tracer, including Breast Specific Gamma Imaging (BSGI)				
Magnetic Resonance Imaging				
We combined the data reported by all 41 studies into a bivariate binomial mixed model. The data were extremely heterogeneous (I2 = 98.4%). The summary sensitivity of MRI for all lesions was 91.7 percent (95% CI: 88.5 to 94.1%) and the summary specificity was 77.5 percent (95% CI: 71.0 to 82.9%). These summary estimates are fairly similar to our 2006 estimates of the accuracy of MRI (at the mean threshold the sensitivity was 92.5%, and the specificity was 72.4%).	A new study[4] found MRI diagnostic characteristics within similar findings of the combined data.	No new data	Three experts thought this conclusion was still valid.	Conclusion is still valid and this portion of the CER does not need updating.
We investigated the heterogeneity with meta-regression. The variables investigated were: the strength of the magnet, the type of contrast agent used, whether the study enrolled all/consecutive patients or not, whether the study was prospective in design or not, whether all diagnoses were verified by histopathology or not, whether any financial conflicts of interest from the funding source existed or not, whether the study was multi- or single-	No new data	No new data	Two experts thought this conclusion was still valid.	Conclusion is still valid and this portion of the CER does not need updating.

Conclusions From CER Executive Summary	RAND Literature Search	FDA/ Health Canada/MHRA (UK)	Expert Opinion EPC Investigator Other Experts	Conclusion from SCEPC
not, whether the study accounted for inter-reader differences or not, the geographical setting of the study, whether the study was clearly affected by spectrum bias or not, and the prevalence of disease. The prevalence of disease in the study population and whether or not readers were blinded to clinical information were both found to be statistically significantly correlated with the accuracy data reported by the studies (p = 0.02 and 0.03, respectively). However, in subgroup analyses there was a statistical correlation between blinding of readers and prevalence of disease. Graphical analysis of prevalence of disease by accuracy failed to reveal any consistent pattern; therefore it is possible that the correlation between prevalence of disease and accuracy is an artifact caused by the correlation between blinding and enrollment of a population with a higher prevalence of disease. Studies that reported they had blinded readers to clinical information had a lower sensitivity than non-blinded studies (86.8% vs. 93.9%) but approximately the same specificity (74.7% vs. 78.0%).				
Position Emission Tomography				
Seven studies reported results for 403 lesions in patients referred for further evaluation by whole-body PET scanning for	No new data	No new data	Three experts thought this conclusion was still valid.	Conclusion is still valid and this portion of the CER does not need updating.

Conclusions From CER Executive Summary	RAND Literature Search	FDA/ Health Canada/MHRA (UK)	Expert Opinion EPC Investigator Other Experts	Conclusion from SCEPC
suspicious breast lesions (abnormal mammogram and/or physical examination and/or ultrasound examination), summarized in Table 8. When combined in a mixed-effects bivariate model, the summary sensitivity of PET for all lesions was 83.0 percent (95% CI: 73.0 to 89.0%), and the summary specificity was 74.0 percent (95% CI: 58 to 86%), findings that are virtually identical to our estimates in the 2006 CER (Table 9). However, the data were found to contain significant heterogeneity (I^2 = 64.0%), indicating substantial variability across the study results. The observed heterogeneity could not be explained through meta-regression using the following covariates: position (prone versus supine), enrolled mostly patients with palpable lesions (>75% vs. <75% or not reported), and blinded to patient clinical information (versus not blinded or not reported). Because the PET data are inconsistent and imprecise, we rated the strength of evidence supporting the estimate of accuracy as "low."				
The study of PET/CT was a single-center study that enrolled a total of 44 patients with 55 suspicious breast lesions detected by physical examination, mammography, or ultrasound.[16]	No new data	No new data	Two experts thought this conclusion was still valid.	Conclusion is still valid and this portion of the CER does not need updating.

Conclusions From CER Executive Summary	RAND Literature Search	FDA/ Health Canada/MHRA (UK)	Expert Opinion EPC Investigator Other Experts	Conclusion from SCEPC
PET scanning was performed at two time points. The first acquisition (Time 1) occurred immediately after an initial whole-body PET scan, and the second one (Time 2) occurred three hours after the first. At both time points, the images of the breast were acquired in the prone position. The CT data were used for attenuation correction, and images were reconstructed using a standard iterative algorithm.				
Scintimammography				
When all 11 studies were combined in the analysis, regardless of imaging technique(s) used, the summary sensitivity of SMM for all lesions was 84.7 percent (78.0 to 89.7%) and the summary specificity was 77.0 percent (95% CI: 64.7 to 85.9%). We also meta-analyzed the data reported by the nine included studies that used standard SMM (planar and double-phase imaging) by fitting a bivariate mixed-effects model. The summary sensitivity of standard SMM for all lesions was 84 percent (95% CI: 76% to 89%) and the summary specificity was 79 percent (95% CI: 63% to 89%), approximately the same as for the full dataset. In 2006, we found that the sensitivity of scintimammography was 68.7 percent and the specificity was 84.8 percent. Improvements in technology and techniques since	A new study[5] found scintimammography characteristics within similar findings of the combined data.	No new data	Three experts thought this conclusion was still valid.	Conclusion is still valid and this portion of the CER does not need updating.

Conclusions From CER Executive Summary	RAND Literature Search	FDA/ Health Canada/MHRA (UK)	Expert Opinion EPC Investigator Other Experts	Conclusion from SCEPC
then, such as the development of double-phase imaging, may explain the improved accuracy in the more recent studies.				
There was a great deal of heterogeneity (I^2 = 93%) in the reported data. We were unable to identify with meta-regression any study- related characteristics that explained this heterogeneity, such as consecutive enrollment of patients, blinding of the diagnostic test reader to patient history/other clinical information, and use of the gold standard (biopsy) as the reference standard.	No new data	No new data	Two experts thought this conclusion was still valid.	Conclusion is still valid and this portion of the CER does not need updating.
Ultrasound (B-Mode 2D Grayscale				
Twenty-one studies of 8,199 lesions addressed the accuracy of B-mode 2D grayscale.[18,26,65-83] We combined the reported data in a bivariate binomial model. The summary sensitivity of B-mode 2D grayscale ultrasound for all lesions was 92.4 percent (95% CI: 84.6 to 96.4%) and the summary specificity was 75.8 percent (60.8 to 86.3%); there was, however, considerable heterogeneity in the data (I^2 = 99.6%). In our 2006 assessment, we found that for suspicious lesions in general, the sensitivity of B-mode ultrasound examination was 86.1 percent, considerably lower than the findings of the current update; and we also found in 2006 that the specificity was 66.4 percent,	No new data	No new data	Three experts thought this conclusion was still valid.	Conclusion is still valid and this portion of the CER does not need updating.

Conclusions From CER Executive Summary	RAND Literature Search	FDA/ Health Canada/MHRA (UK)	Expert Opinion EPC Investigator Other Experts	Conclusion from SCEPC
lower than the 75.8 percent specificity of the current update. The 2006 version of the report included only a small subset of the evidence base included in the current update.				
We conducted meta-regressions to explore the heterogeneity in the data. The variables we investigated were: whether the studies accounted for inter-reader differences; whether the studies blinded image readers to clinical information or not; whether all diagnoses were verified by histopathology or not; whether a prospective design was used; whether the study was funded by a source without a financial interest in the results or not; whether the study enrolled consecutive/ all patients; the geographical location of the study; what type(s) of breast lesions were enrolled in the study; and the prevalence of disease in the study. Two of these variables, whether the studies accounted for inter-reader differences, and whether the studies blinded image readers to clinical information or not, were statistically significantly associated with the results (p = 0.01 and 0.03, respectively). Subgroup analyses found that studies that had blinded image readers to clinical information had a higher sensitivity (96.6% vs. 87.0%) but a much lower specificity (59.5% vs. 85.1%) than unblinded studies. Studies	No new data	No new data	Two experts thought this conclusion was still valid.	Conclusion is still valid and this portion of the CER does not need updating.

Conclusions From CER Executive Summary	RAND Literature Search	FDA/ Health Canada/MHRA (UK)	Expert Opinion EPC Investigator Other Experts	Conclusion from SCEPC
that had accounted for inter-reader differences had a similar sensitivity (93.4% vs. 93.0%) but a much lower specificity (52.7% vs. 90.1%) than studies that did not account for inter-reader differences.				
Ultrasound (B-Mode 2D Grayscale, Contrast Enhanced)				
Only two studies of a total of 154 breast lesions reported on the accuracy of B-mode 2D grayscale contrast-enhanced ultrasound compared to non-contrast enhanced.[26,66] Contrast enhancement was reported to increase the sensitivity (97.5% vs. 82.7%) but to not dramatically affect the specificity (76.7% vs. 74.0%).	No new data	No new data	Three experts thought this conclusion was still valid.	Conclusion is still valid and this portion of the CER does not need updating.
Ultrasound (B-Mode 3D Grayscale)				
Only one study of 150 breast lesions, Cho et al., reported on the accuracy of B-mode 3D grayscale ultrasound.[71]	No new data	No new data	Three experts thought this conclusion was still valid.	Conclusion is still valid and this portion of the CER does not need updating.
Ultrasound (Color Doppler Ultrasound)				
Six studies of a total of 718 lesions reported on the accuracy of color Doppler ultrasound.[78,80,84-87] We combined the data reported by these studies in a bivariate binomial model. The summary sensitivity of color Doppler ultrasound for all lesions was 88.5 percent (95% CI: 74.4 to 95.4%) and the summary specificity was 76.4 percent (95% CI: 61.7 to 86.7%). There was considerable heterogeneity	No new data	No new data	Three experts thought this conclusion was still valid.	Conclusion is still valid and this portion of the CER does not need updating.

Conclusions From CER Executive Summary	RAND Literature Search	FDA/ Health Canada/MHRA (UK)	Expert Opinion EPC Investigator Other Experts	Conclusion from SCEPC
in the data (I^2 = 95.2%). Exclusion of data from two studies that enrolled only patients with palpable lesions[80,85] from the bivariate model did not affect the results. There were too few studies of color Doppler to perform full meta-regressions.				
Ultrasound (Color Doppler Ultrasound, Contrast Enhanced)				
Two studies of 146 lesions compared the accuracy of contrast-enhanced color Doppler to non-enhanced color Doppler.[84,86] Contrast-enhancement was found to slightly increase the sensitivity (97.8% vs. 95.7%) and to dramatically increase the specificity (90.7% vs. 55.6%).	No new data	No new data	Three experts thought this conclusion was still valid.	Conclusion is still valid and this portion of the CER does not need updating.
Color Doppler Ultrasound Directly Compared With B-mode Grayscale Ultrasound				
Two studies directly compared the accuracy of color Doppler ultrasound to B-mode grayscale ultrasound.[78,80] Color Doppler was found to have a higher sensitivity (74.0% vs. 53.1%) but a lower specificity than B-mode ultrasound (84.0% vs. 96.3%).	No new data	No new data	Three experts thought this conclusion was still valid.	Conclusion is still valid and this portion of the CER does not need updating.
Ultrasound (Power Doppler Ultrasound)				
Nine studies of a total of 614 lesions reported on the accuracy of power Doppler ultrasound.[65,72,75,77,86,88-91] We combined the data in a bivariate binomial model. The summary sensitivity of power Doppler ultrasound for all lesions was 70.8 percent (95% CI: 47.5 to 86.6%) and the summary	No new data	No new data	Three experts thought this conclusion was still valid.	Conclusion is still valid and this portion of the CER does not need updating.

Conclusions From CER Executive Summary	RAND Literature Search	FDA/ Health Canada/MHRA (UK)	Expert Opinion EPC Investigator Other Experts	Conclusion from SCEPC
specificity was 72.6 percent (95% CI: 59.9 to 82.5%). There was considerable heterogeneity in the data (I^2 = 97.4%).				
Ultrasound (Power Doppler Ultrasound, Contrast Enhanced)				
Seven studies of 403 lesions reported on the accuracy of contrast-enhanced power Doppler ultrasound.[72,75,77,86,88,90,91] When we combined the data in a bivariate binomial model, the summary sensitivity for all lesions was 89.3 percent (95% CI: 52.4 to 98.4%) and the summary specificity was 70.4 percent (95% CI: 55.4 to 82.0%). There was considerable heterogeneity in the data (I^2 = 87.5%).	No new data	No new data	Three experts thought this conclusion was still valid.	Conclusion is still valid and this portion of the CER does not need updating.
Ultrasound (Power Doppler Ultrasound Directly Compared With B-Mode Grayscale Ultrasound)				
Four studies of 248 lesions directly compared the accuracy of power Doppler ultrasound to B-mode grayscale ultrasound.[65,72,75,77] Power Doppler was found to have a lower sensitivity (54.7% vs. 87.7%) but a higher specificity (79.4% vs. 50.7%) than B-mode grayscale ultrasound in these four direct comparisons.	No new data	No new data	Three experts thought this conclusion was still valid.	Conclusion is still valid and this portion of the CER does not need updating.
Ultrasound (Power Doppler Ultrasound Directly Compared With Color Doppler)				
One study directly compared the accuracy of power Doppler, with and without contrast-enhancement, to color Doppler, with and without contrast-enhancement.[86] This study	No new data	No new data	Three experts thought this conclusion was still valid.	Conclusion is still valid and this portion of the CER does not need updating.

Conclusions From CER Executive Summary	RAND Literature Search	FDA/ Health Canada/MHRA (UK)	Expert Opinion EPC Investigator Other Experts	Conclusion from SCEPC
reported that all four methods had a 100 percent sensitivity, but specificity for contrast-enhanced methods was much higher than for non-contrast-enhanced methods.				
Ultrasound (Tissue Harmonics)				
Only one study of 91 lesions reported on the accuracy of tissue harmonic ultrasound methods.[68]	No new data	No new data	Two experts thought this conclusion was still valid.	Conclusion is still valid and this portion of the CER does not need updating.
Key Question 2. Are there demographic (e.g., age) and clinical risk factors (e.g., morphologic characteristics of the lesion) that affect the accuracy of the tests considered in Key Question 1?				
Magnetic Resonance Imaging				
Two studies reported the accuracy of MRI by patient age.[30,44] One of these two studies (Bluemke et al.[30]) investigated the relative accuracy by premenopausal status vs. post-menopausal status of the patients, and reported virtually no difference in either sensitivity or specificity between groups. The other study (Imbriaco et al.[44]) reported the accuracy of MRI for women 50 years of age and older vs. younger women, and found that MRI was more sensitive (100% vs. 92.9%) in younger women, but had virtually the same specificity (75.0%) in both age groups.	No new data	No new data	Three experts thought this conclusion was still valid.	Conclusion is still valid and this portion of the CER does not need updating.
Eight of the studies enrolled patients who had been referred for further investigation after identification of microcalcifications on mammography.[20,22,23,25,30,39,46,51]	No new data	No new data	Two experts thought this conclusion was still valid.	Conclusion is still valid and this portion of the CER does not need updating.

Conclusions From CER Executive Summary	RAND Literature Search	FDA/ Health Canada/MHRA (UK)	Expert Opinion EPC Investigator Other Experts	Conclusion from SCEPC
When combined in a bivariate mixed-effects model the data from these eight studies had very low heterogeneity (I^2 = 3.86%). The summary sensitivity of MRI for microcalcifications was 84.0 percent (79.5 to 88.3%) and the summary specificity was 79.4 percent (71.5 to 85.6%). The summary sensitivity of MRI for evaluation of microcalcifications is considerably lower than the sensitivity of MRI for evaluation of any/all lesions (84.0% vs. 91.7%). The specificity for microcalcifications is approximately the same (79.4% vs. 77.5%). Two studies also directly compared the sensitivity of MRI for evaluation of microcalcifications vs. other types or all types of lesions (Bluemke et al.[30] and Van Goethem et al.[51]) and reported similar results: the sensitivity of MRI for evaluation of microcalcifications is approximately 85 percent, which is considerably lower than the sensitivity of MRI for evaluation of all/other types of lesions; whereas the specificity of MRI for evaluation of microcalcifications is approximately 77 percent, which may be slightly higher than the specificity of MRI for evaluation of all/other types of lesions.				
Two studies evaluated the accuracy of MRI for dense	No new data	No new data	Two experts thought this conclusion was still valid.	Conclusion is still valid and this portion of the CER does not need

Conclusions From CER Executive Summary	RAND Literature Search	FDA/ Health Canada/MHRA (UK)	Expert Opinion EPC Investigator Other Experts	Conclusion from SCEPC
breasts vs. all or non-dense breasts (Bluemke et al.[30] and Wiberg et al.[40]), and reported virtually no difference in the accuracy of MRI for evaluation of these different categories of breast tissue.				updating.
One study enrolled only patients with lesions classified as BIRADS 3 before investigation by MRI (Gokalp and Topal[24]); however, only one enrolled patient (out of 43 total) was found to have a malignancy and therefore the patient population is too small to draw conclusions about the accuracy of MRI for probably benign lesions.	No new data	No new data	Two experts thought this conclusion was still valid.	Conclusion is still valid and this portion of the CER does not need updating.
One study each investigated the accuracy of MRI for lesions broken down by palpable vs. non-palpable (Bluemke et al.[30]) and large lesion vs. small lesion (Imbracio et al.[44]).	No new data	No new data	Two experts thought this conclusion was still valid.	Conclusion is still valid and this portion of the CER does not need updating.
Two studies reported the accuracy of MRI by patient age.[30,44] One of these two studies (Bluemke et al.[30]) investigated the relative accuracy by premenopausal status vs. post-menopausal status of the patients, and reported virtually no difference in either sensitivity or specificity between groups. The other study (Imbriaco et al.[44]) reported the accuracy of MRI for women 50 years of age	No new data	No new data	Two experts thought this conclusion was still valid.	Conclusion is still valid and this portion of the CER does not need updating.

Conclusions From CER Executive Summary	RAND Literature Search	FDA/ Health Canada/MHRA (UK)	Expert Opinion EPC Investigator Other Experts	Conclusion from SCEPC
and older vs. younger women, and found that MRI was more sensitive (100% vs. 92.9%) in younger women, but had virtually the same specificity (75.0%) in both age groups.				
Position Emission Tomography				
In three of the seven studies that addressed Key Question 1, the majority (>75.0%) of the women presented with palpable breast lesions— Kiada et al.[52]: 88.0 percent palpable, Schirrmeister et al.[54]: 76.0 percent, and Yutani et al.[55]: 93.0 percent palpable. Because there were only three studies, we could not fit the data in a bivariate model. Instead, we pooled the reported sensitivities and specificities in random-effects meta-analyses. However, the data were heterogeneous (I^2 = 68.0% and I^2 = 54.6% for sensitivity and specificity, respectively), indicating substantial variability among the study results. With only three studies, we did not attempt to explore possible reason(s) for the heterogeneity. The overall sensitivity for primarily palpable lesions is higher than that for all seven studies considered under Key 1 (86.5% vs. 83.0%), but the specificity is lower (64.2% vs. 74.0%).	No new data	No new data	Three experts thought this conclusion was still valid.	Conclusion is still valid and this portion of the CER does not need updating.
One study directly compared images acquired when patients were in prone position to images	No new data	No new data	Two experts thought this conclusion was still valid.	Conclusion is still valid and this portion of the CER does not need updating.

Conclusions From CER Executive Summary	RAND Literature Search	FDA/ Health Canada/MHRA (UK)	Expert Opinion EPC Investigator Other Experts	Conclusion from SCEPC
of the same patients in supine position.[52] In this study by Kaida et al. 2008, 118 women with 122 lesions suspected of having breast cancer underwent whole-body PET in the supine position immediately followed by prone breast PET imaging. According to the results reported in the study, the sensitivity and specificity of images in the supine position were 83.0 percent and 50.0 percent, respectively. The sensitivity and specificity of images in the prone position were 96.0 percent and 50.0 percent, respectively.				
One study, Yutani et al. 2000, reported results separately for patients with BIRADS 5, lesions 1.5 cm or larger, and who were younger than 65.[55] The authors reported that PET was more sensitive for larger lesions, but the specificity was unchanged; and for the other factors, the accuracy of PET was virtually the same as for PET for all patients.	No new data	No new data	Two experts thought this conclusion was still valid.	Conclusion is still valid and this portion of the CER does not need updating.
Scintimammography				
Two studies evaluated only patients with palpable breast masses,[57,62] one study evaluated only patients with non-palpable breast masses,[63] and one study evaluated only patients with microcalcifications detected on x-ray mammography.[61] With so few studies reporting on each	No new data	No new data	Three experts thought this conclusion was still valid.	Conclusion is still valid and this portion of the CER does not need updating.

Conclusions From CER Executive Summary	RAND Literature Search	FDA/ Health Canada/MHRA (UK)	Expert Opinion EPC Investigator Other Experts	Conclusion from SCEPC
category, evidence-based conclusions are difficult to support.				
None of the studies reported outcomes by patient demographics or any other clinical risk factors that may have affected the accuracy of SMM.	No new data	No new data	Two experts thought this conclusion was still valid.	Conclusion is still valid and this portion of the CER does not need updating.
Ultrasound				
None were identified.	No new data	No new data	Three experts thought this conclusion was still valid.	Conclusion is still valid and this portion of the CER does not need updating.
Key Question 3. Are there other factors and considerations (e.g., safety, care setting, patient preferences, ease of access to care) that may affect the accuracy or acceptability of the tests considered in Key Questions 1 and 2?				
Magnetic Resonance Imaging				
One study reported the accuracy of MRI images interpreted with and without a Computer Aided Diagnosis (CAD) software system.[12] The study reported virtually no difference in either sensitivity (77.4% vs. 78.9%) or specificity (73.2% vs. 73.2%) with or without CAD assistance.	No new data	No new data	Three experts thought this conclusion was still valid.	Conclusion is still valid and this portion of the CER does not need updating.
Position Emission Tomography				
None of the seven studies on stand-alone PET scanning or the one study on PET with CT reported information that addressed this question.	No new data	No new data	Two experts thought this was out of date and pointed to a study that was included in the original report. One expert thought this conclusion was still valid.	Conclusion is still valid and this portion of the CER does not need updating.
Scintimammography				
None were identified.	No new data	No new data	Two experts thought this was out of date and pointed to a study that was included in the original report. One expert thought this conclusion was still valid.	Conclusion is still valid and this portion of the CER does not need updating.
Ultrasound				
None were identified.	No new data	No new data	Three experts thought this	Conclusion is still valid and this

Conclusions From CER Executive Summary	RAND Literature Search	FDA/ Health Canada/MHRA (UK)	Expert Opinion EPC Investigator Other Experts	Conclusion from SCEPC
			conclusion was still valid.	portion of the CER does not need updating.

Legend: BSGI=Breast Specific Gamma Imaging (BSGI); CADx=Computer-aided diagnosis; CER=Comparative Effectiveness Review; CT=Computed Tomography; FDG= Fluorodeoxyglucose; MRI=Magnetic Resonance Imaging; PET=Positron Emission Tomography; SCEPC=Southern California Evidence-based Practice Center; SMM=Scintimammography

References

1. Bruening W, Uhl S, Fontanarosa J, et al. 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 Agency for Healthcare Research and Quality. Rockville, MD: Feb February 2012. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=22420009.

2. Shekelle PG, Newberry SJ, Maglione M, et al. Assessment of the Need to Update Comparative Effectiveness Reviews: Report of an Initial Rapid Program Assessment (2005-2009) (Prepared by the Southern California Evidence-based Practice Center). Rockville, MD: Agency for Healthcare Research and Quality; October 2009.

3. Shojania KG, Sampson M, Ansari MT, et al. How quickly do systematic reviews go out of date? A survival analysis. Ann Intern Med. 2007 Aug 21;147(4):224-33. PMID 17638714.

4. Hillman BJ, Harms SE, Stevens G, et al. Diagnostic Performance of a Dedicated 1.5-T Breast MR Imaging System. Radiology. 2012 Oct;265(1):51-8. PMID 22923716.

5. Weigert JM, Bertrand ML, Lanzkowsky L, et al. Results of a multicenter patient registry to determine the clinical impact of breast-specific gamma imaging, a molecular breast imaging technique. AJR Am J Roentgenol. 2012 Jan;198(1):W69-75. PMID 22194518.

Appendices

Appendix A: Search Methodology

Appendix B: Evidence Tables

Appendix C: Questionnaire Matrix

Appendix A. Search Methodology

DATABASE SEARCHED & TIME PERIOD COVERED:
MEDLINE ON OVID – 1/1/2010-8/27/2012

LANGUAGE:
English

SEARCH STRATEGY:
exp Breast neoplasms/ or exp breast diseases/ or exp breast cancer/ or breast carcinoma/ or ((breast or mammary) and (cancer$ or neoplasm$ or carcinoma$ or tumor$ or tumour$ or lump$ or lesion$)).mp.
AND
sensitivity and specificity"/ or early diagnosis/ or diagnostic imaging/ or diagnostic value/ or diagnostic accuracy/ or diagnostic procedure/ or tumor diagnosis/ OR diagnos$.mp. or di.xs. or "gold standard".mp. or ROC.mp. or "receiver operating characteristic".mp. or likelihood.mp. OR ((false or true) adj (positive or negative)).tw. or "predictive value".mp. or accuracy.mp. or precision.mp. or sensitivity.mp. or specificity.mp.
AND
(noninvasive or non-invasive).mp. or ultrasonography.fs. or ultrasonography, mammary/ or echogra$.mp. or echomammog$.mp. or sonogra$.mp. or sonomammogr$.mp. or ultrasound.mp. or ultrason$.mp. or echomammography/ or ultrasound/ OR exp magnetic resonance imaging/ or "magnet strength".mp. or pulse sequence.mp. or mr.mp. or mri.mp. or nuclear magnetic resonance.mp. or nmr.mp. or nuclear magnetic resonance imaging/ or magnetic resonance.mp. OR (fdg$ or f-fluorodeoxyglucose or f18 or f-18).mp. or fluorodeoxyglucose f 18/ or PET.ti. or positron emission tomography.mp. or exp tomography, emission-computed/ or (comput$ adj tomograph$).tw. or positron emission tomography/ OR (gamma camera$ or gammagraph$ or nuclear medicine or radionuclide$).mp. or radionuclide imaging.fs. or radiotracer$.mp. or radiopharmaceuticals/ or sestamibi.mp. or technetium Tc 99m Sestamibi/du or gammagraph$.mp. or pem tetrofosomin.mp. or technetium.mp. or miraluma.mp. or tetrofosmin.mp. or scintimammogr$.mp. or spectrometry, gamma/ or methoxy isobutyl isonitrile technetium tc-99/ or nuclear medicine/ or scintillation camera/ or scintimammography/ or gamma spectrometry/ or exp organotechnetium compounds/du or MIBI.mp. or BSGI.mp. or gamma cameras/ OR exp spectrometry, x-ray emission/ or SPET.mp. or SPECT.mp. or single photon emission computer tomography/ OR tomosynthesis.mp. or three dimensional imaging/ or 3-D.mp. or 3D.mp. or imaging, three dimensional/ or (three adj dimension$).mp. or (("sensitivity and specificity" or early diagnosis or diagnostic imaging or diagnostic value or diagnostic accuracy or diagnostic procedure or tumor diagnosis) adj dimension$).mp. OR tomosynthesis.mp. or three dimensional imaging/ or 3-D.mp. or 3D.mp. or imaging, three dimensional/ or (three adj dimension$).mp. OR diagnosis, computer-assisted/ or image interpretation, computer-assisted/ or radiographic image interpretation, computer-assisted/ or computer assisted diagnosis/ or digital mammography/ or (comput$ adj (aided or assisted) adj (detection or diagnos$)).tw. or digital mammogra$.mp. or CAD.mp. or exp image processing, computer-assisted/ or image analysis/
OR
ultrasonography, doppler/ or ultrasonography, doppler, duplex/ or ultrasonography, doppler, color/ or doppler echography/ or (doppler adj2 (ultrason$ or echograph$)).tw. OR ((positron-emission tomography/ or tomography, emission-computed/) and (tomography, x-ray computed/ or computer assisted tomography.mp.)) or (pet adj ct).tw. or pet/ct or (positron emission tomograph$ and comput$ tomograph$).mp.
NOT
(letter or editorial or news or comment or case reports or note or conference paper).de.) or (letter or editorial or news or comment or case reports or note or conference paper).pt.

NUMBER OF RESULTS: 3710

FILTERED IN ENDNOTE TO LIMIT TO THE FOLLOWING JOURNALS:
ANNALS OF INTERNAL MEDICINE
BMJ
JAMA
LANCET
NEW ENGLAND JOURNAL OF MEDICINE

BREAST
CLINICAL RADIOLOGY
RADIOLOGY

NUMBER OF RESULTS AFTER FILTERING FOR JOURNALS: 161

Appendix B. Evidence Table

Article ID, Author, year	Trial	n	Subjects	Primary Outcome	Duration	Study Quality/ Applicability	Findings
Key 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 identification of a possible breast abnormality on routine screening (mammography and/or clinical or self-detection of a palpable lesion)? The noninvasive tests to be evaluated are: Ultrasound (conventional B-mode, color Doppler, power Doppler, tissue harmonics, and tomography); Magnetic resonance imaging (MRI) with breast-specific coils and gadolinium-based contrast agents, with or without computer-aided diagnosis (CADx); Positron emission tomography (PET) with 18-fluorodeoxyglucose (FDG) as the tracer, with or without concurrent computed tomography (CT) scans; Scintimammography (SMM) with technetium-99m sestamibi (MIBI) as the tracer, including Breast Specific Gamma Imaging (BSGI)							
Magnetic Resonance Imaging							
Hilman, 2012[4]	Retrospective study	587	Women 25-89	Screening and diagnostic accuracy	4/06-12/07	Good/Good	Diagnostic accuracy of MRI: Sensitivity: 92 % (84.1-96.3) Specificity: 84.3% (80.8-87.5)
Ultrasound							
Weigert, 2012[5]	Retrospective study	1042	Women with 2 of the following: equivocal or negative mammogram or sonogram and an unresolved clinical concern; personal history of breast cancer or current cancer diagnosis; palpable masses negative on mammographic and sonographic examination; radiodense breast tissue; or high risk	Diagnostic accuracy	Available for 6 months os follow-up	Good/Good	Sensitivity: 91% Specificity: 77%

Article ID, Author, year	Trial	n	Subjects	Primary Outcome	Duration	Study Quality/ Applicability	Findings
			for breast cancer.				

Legend: BSGI=Breast Specific Gamma Imaging (BSGI); CADx=Computer-aided diagnosis; CER=Comparative Effectiveness Review; CT=Computed Tomography; FDG= Fluorodeoxyglucose; MRI=Magnetic Resonance Imaging; PET=Positron Emission Tomography; SCEPC=Southern California Evidence-based Practice Center; SMM=Scintimammography

Appendix C. Questionnaire Matrix

Surveillance and Identification of Triggers for Updating Systematic Reviews for the EHC Program

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

Conclusions From CER Executive Summary	Is this conclusion almost certainly still supported by the evidence?	Has there been new evidence that may change this conclusion?	Do Not Know
Key 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 identification of a possible breast abnormality on routine screening (mammography and/or clinical or self-detection of a palpable lesion)? The noninvasive tests to be evaluated are: Ultrasound (conventional B-mode, color Doppler, power Doppler, tissue harmonics, and tomography); Magnetic resonance imaging (MRI) with breast-specific coils and gadolinium-based contrast agents, with or without computer-aided diagnosis (CADx); Positron emission tomography (PET) with 18-fluorodeoxyglucose (FDG) as the tracer, with or without concurrent computed tomography (CT) scans; Scintimammography (SMM) with technetium-99m sestamibi (MIBI) as the tracer, including Breast Specific Gamma Imaging (BSGI)			
Magnetic Resonance Imaging			
We combined the data reported by all 41 studies into a bivariate binomial mixed model. The data were extremely heterogeneous (I2 = 98.4%). The summary sensitivity of MRI for all lesions was 91.7 percent (95% CI: 88.5 to 94.1%) and the summary specificity was 77.5 percent (95% CI: 71.0 to 82.9%). These summary estimates are fairly similar to our 2006 estimates of the accuracy of MRI (at the mean threshold the sensitivity was 92.5%, and the specificity was 72.4%).	☐	New Evidence:	☐
We investigated the heterogeneity with meta-regression. The variables investigated		New Evidence:	

Conclusions From CER Executive Summary	Is this conclusion almost certainly still supported by the evidence?	Has there been new evidence that may change this conclusion?	Do Not Know
were: the strength of the magnet, the type of contrast agent used, whether the study enrolled all/consecutive patients or not, whether the study was prospective in design or not, whether all diagnoses were verified by histopathology or not, whether any financial conflicts of interest from the funding source existed or not, whether the study was multi- or single-centered, whether readers were blinded to clinical information or not, whether the study accounted for inter-reader differences or not, the geographical setting of the study, whether the study was clearly affected by spectrum bias or not, and the prevalence of disease. The prevalence of disease in the study population and whether or not readers were blinded to clinical information were both found to be statistically significantly correlated with the accuracy data reported by the studies ($p = 0.02$ and 0.03, respectively). However, in subgroup analyses there was a statistical correlation between blinding of readers and prevalence of disease. Graphical analysis of prevalence of disease by accuracy failed to reveal any consistent pattern; therefore it is possible that the correlation between prevalence of disease and accuracy is an artifact caused by the correlation between blinding and enrollment of a population with a higher prevalence of disease. Studies that reported they had blinded readers to clinical information had a lower sensitivity than	☐		☐

Conclusions From CER Executive Summary	Is this conclusion almost certainly still supported by the evidence?	Has there been new evidence that may change this conclusion?	Do Not Know
non-blinded studies (86.8% vs. 93.9%) but approximately the same specificity (74.7% vs. 78.0%).			
Position Emission Tomography			
Seven studies reported results for 403 lesions in patients referred for further evaluation by whole-body PET scanning for suspicious breast lesions (abnormal mammogram and/or physical examination and/or ultrasound examination), summarized in Table 8. When combined in a mixed-effects bivariate model, the summary sensitivity of PET for all lesions was 83.0 percent (95% CI: 73.0 to 89.0%), and the summary specificity was 74.0 percent (95% CI: 58 to 86%), findings that are virtually identical to our estimates in the 2006 CER (Table 9). However, the data were found to contain significant heterogeneity (I^2 = 64.0%), indicating substantial variability across the study results. The observed heterogeneity could not be explained through meta-regression using the following covariates: position (prone versus supine), enrolled mostly patients with palpable lesions (>75% vs. <75% or not reported), and blinded to patient clinical information (versus not blinded or not reported). Because the PET data are inconsistent and imprecise, we rated the strength of evidence supporting the estimate of accuracy as "low."	☐	New Evidence:	☐

Conclusions From CER Executive Summary	Is this conclusion almost certainly still supported by the evidence?	Has there been new evidence that may change this conclusion?	Do Not Know
The study of PET/CT was a single-center study that enrolled a total of 44 patients with 55 suspicious breast lesions detected by physical examination, mammography, or ultrasound.[16] PET scanning was performed at two time points. The first acquisition (Time 1) occurred immediately after an initial whole-body PET scan, and the second one (Time 2) occurred three hours after the first. At both time points, the images of the breast were acquired in the prone position. The CT data were used for attenuation correction, and images were reconstructed using a standard iterative algorithm.	☐	New Evidence:	☐
The authors reported that dual-time point PET/CT (Time 2) demonstrated a sensitivity of 80 percent and specificity of 100 percent compared to a sensitivity of 62 percent and specificity of 100 percent for single time-point PET/CT. The authors concluded that malignant lesions showed a significant increase in FDG over time compared to benign lesions.	☐	New Evidence:	☐
Scintimammography			
When all 11 studies were combined in the analysis, regardless of imaging technique(s) used, the summary sensitivity of SMM for all lesions was 84.7 percent (78.0 to 89.7%) and the summary specificity was 77.0 percent (95% CI: 64.7 to 85.9%). We also meta-analyzed the data reported by the nine included studies that used standard SMM	☐	New Evidence:	☐

Conclusions From CER Executive Summary	Is this conclusion almost certainly still supported by the evidence?	Has there been new evidence that may change this conclusion?	Do Not Know
(planar and double-phase imaging) by fitting a bivariate mixed-effects model. The summary sensitivity of standard SMM for all lesions was 84 percent (95% CI: 76% to 89%) and the summary specificity was 79 percent (95% CI: 63% to 89%), approximately the same as for the full dataset. In 2006, we found that the sensitivity of scintimammography was 68.7 percent and the specificity was 84.8 percent. Improvements in technology and techniques since then, such as the development of double-phase imaging, may explain the improved accuracy in the more recent studies.			
There was a great deal of heterogeneity (I^2 = 93%) in the reported data. We were unable to identify with meta-regression any study- related characteristics that explained this heterogeneity, such as consecutive enrollment of patients, blinding of the diagnostic test reader to patient history/other clinical information, and use of the gold standard (biopsy) as the reference standard.	☐	New Evidence:	☐
Ultrasound (B-Mode 2D Grayscale)			
Twenty-one studies of 8,199 lesions addressed the accuracy of B-mode 2D grayscale.[18,26,65-83] We combined the reported data in a bivariate binomial model. The summary sensitivity of B-mode 2D grayscale ultrasound for all lesions was 92.4 percent (95% CI: 84.6 to 96.4%) and the	☐	New Evidence:	☐

Conclusions From CER Executive Summary	Is this conclusion almost certainly still supported by the evidence?	Has there been new evidence that may change this conclusion?	Do Not Know
summary specificity was 75.8 percent (60.8 to 86.3%); there was, however, considerable heterogeneity in the data (I^2 = 99.6%). In our 2006 assessment, we found that for suspicious lesions in general, the sensitivity of B-mode ultrasound examination was 86.1 percent, considerably lower than the findings of the current update; and we also found in 2006 that the specificity was 66.4 percent, lower than the 75.8 percent specificity of the current update. The 2006 version of the report included only a small subset of the evidence base included in the current update.			
We conducted meta-regressions to explore the heterogeneity in the data. The variables we investigated were: whether the studies accounted for inter-reader differences; whether the studies blinded image readers to clinical information or not; whether all diagnoses were verified by histopathology or not; whether a prospective design was used; whether the study was funded by a source without a financial interest in the results or not; whether the study enrolled consecutive/ all patients; the geographical location of the study; what type(s) of breast lesions were enrolled in the study; and the prevalence of disease in the study. Two of these variables, whether the studies accounted for inter-reader differences, and whether the studies blinded image readers to clinical information or not, were	☐	New Evidence:	☐

Conclusions From CER Executive Summary	Is this conclusion almost certainly still supported by the evidence?	Has there been new evidence that may change this conclusion?	Do Not Know
statistically significantly associated with the results (p = 0.01 and 0.03, respectively). Subgroup analyses found that studies that had blinded image readers to clinical information had a higher sensitivity (96.6% vs. 87.0%) but a much lower specificity (59.5% vs. 85.1%) than unblinded studies. Studies that had accounted for inter-reader differences had a similar sensitivity (93.4% vs. 93.0%) but a much lower specificity (52.7% vs. 90.1%) than studies that did not account for inter-reader differences.			
Ultrasound (B-Mode 2D Grayscale, Contrast Enhanced)			
Only two studies of a total of 154 breast lesions reported on the accuracy of B-mode 2D grayscale contrast-enhanced ultrasound[26,66] compared to non-contrast enhanced. Contrast enhancement was reported to increase the sensitivity (97.5% vs. 82.7%) but to not dramatically affect the specificity (76.7% vs. 74.0%).	☐	New Evidence:	☐
Ultrasound (B-Mode 3D Grayscale)			
Only one study of 150 breast lesions, Cho et al., reported on the accuracy of B-mode 3D grayscale ultrasound.[71]	☐	New Evidence:	☐
Ultrasound (Color Doppler Ultrasound)			

Conclusions From CER Executive Summary	Is this conclusion almost certainly still supported by the evidence?	Has there been new evidence that may change this conclusion?	Do Not Know
Six studies of a total of 718 lesions reported on the accuracy of color Doppler ultrasound.[78,80,84-87] We combined the data reported by these studies in a bivariate binomial model. The summary sensitivity of color Doppler ultrasound for all lesions was 88.5 percent (95% CI: 74.4 to 95.4%) and the summary specificity was 76.4 percent (95% CI: 61.7 to 86.7%). There was considerable heterogeneity in the data (I^2 = 95.2%). Exclusion of data from two studies that enrolled only patients with palpable lesions[80,85] from the bivariate model did not affect the results. There were too few studies of color Doppler to perform full meta-regressions.	☐	New Evidence:	☐
Ultrasound (Color Doppler Ultrasound, Contrast Enhanced)			
Two studies of 146 lesions compared the accuracy of contrast-enhanced color Doppler to non-enhanced color Doppler.[84,86] Contrast-enhancement was found to slightly increase the sensitivity (97.8% vs. 95.7%) and to dramatically increase the specificity (90.7% vs. 55.6%).	☐	New Evidence:	☐
Color Doppler Ultrasound Directly Compared With B-mode Grayscale Ultrasound			
Two studies directly compared the accuracy of color Doppler ultrasound to B-mode grayscale ultrasound.[78,80] Color Doppler was found to have a higher sensitivity (74.0% vs. 53.1%) but a lower specificity	☐	New Evidence:	☐

Conclusions From CER Executive Summary	Is this conclusion almost certainly still supported by the evidence?	Has there been new evidence that may change this conclusion?	Do Not Know
than B-mode ultrasound (84.0% vs. 96.3%).			
Ultrasound (Power Doppler Ultrasound)			
Nine studies of a total of 614 lesions reported on the accuracy of power Doppler ultrasound.[65,72,75,77,86,88-91] We combined the data in a bivariate binomial model. The summary sensitivity of power Doppler ultrasound for all lesions was 70.8 percent (95% CI: 47.5 to 86.6%) and the summary specificity was 72.6 percent (95% CI: 59.9 to 82.5%). There was considerable heterogeneity in the data (I^2 = 97.4%).	☐	New Evidence:	☐
Ultrasound (Power Doppler Ultrasound, Contrast Enhanced)			
Seven studies of 403 lesions reported on the accuracy of contrast-enhanced power Doppler ultrasound.[72,75,77,86,88,90,91] When we combined the data in a bivariate binomial model, the summary sensitivity for all lesions was 89.3 percent (95% CI: 52.4 to 98.4%) and the summary specificity was 70.4 percent (95% CI: 55.4 to 82.0%). There was considerable heterogeneity in the data (I^2 = 87.5%).	☐	New Evidence:	☐
Ultrasound (Power Doppler Ultrasound Directly Compared With B-Mode Grayscale Ultrasound)			
Four studies of 248 lesions directly compared the accuracy of power Doppler ultrasound to B-mode grayscale ultrasound.[65,72,75,77] Power Doppler was	☐	New Evidence:	☐

Conclusions From CER Executive Summary	Is this conclusion almost certainly still supported by the evidence?	Has there been new evidence that may change this conclusion?	Do Not Know
found to have a lower sensitivity (54.7% vs. 87.7%) but a higher specificity (79.4% vs. 50.7%) than B-mode grayscale ultrasound in these four direct comparisons.			
Ultrasound (Power Doppler Ultrasound Directly Compared With Color Doppler)			
One study directly compared the accuracy of power Doppler, with and without contrast-enhancement, to color Doppler, with and without contrast-enhancement.[86] This study reported that all four methods had a 100 percent sensitivity, but specificity for contrast-enhanced methods was much higher than for non-contrast-enhanced methods.	☐	New Evidence:	☐
Ultrasound (Tissue Harmonics)			
Only one study of 91 lesions reported on the accuracy of tissue harmonic ultrasound methods.[68]	☐	New Evidence:	☐
Key Question 2. Are there demographic (e.g., age) and clinical risk factors (e.g., morphologic characteristics of the lesion) that affect the accuracy of the tests considered in Key Question 1?			
Magnetic Resonance Imaging			
Two studies reported the accuracy of MRI by patient age.[30,44] One of these two studies (Bluemke et al.[30]) investigated the relative accuracy by premenopausal status vs. post-menopausal status of the patients, and reported virtually no difference in either sensitivity or specificity between groups. The other study (Imbriaco et al.[44]) reported	☐	New Evidence:	☐

Conclusions From CER Executive Summary	Is this conclusion almost certainly still supported by the evidence?	Has there been new evidence that may change this conclusion?	Do Not Know
the accuracy of MRI for women 50 years of age and older vs. younger women, and found that MRI was more sensitive (100% vs. 92.9%) in younger women, but had virtually the same specificity (75.0%) in both age groups.			
Eight of the studies enrolled patients who had been referred for further investigation after identification of microcalcifications on mammography.[20,22,23,25,30,39,46,51] When combined in a bivariate mixed-effects model the data from these eight studies had very low heterogeneity (I^2 = 3.86%). The summary sensitivity of MRI for microcalcifications was 84.0 percent (79.5 to 88.3%) and the summary specificity was 79.4 percent (71.5 to 85.6%). The summary sensitivity of MRI for evaluation of microcalcifications is considerably lower than the sensitivity of MRI for evaluation of any/all lesions (84.0% vs. 91.7%). The specificity for microcalcifications is approximately the same (79.4% vs. 77.5%). Two studies also directly compared the sensitivity of MRI for evaluation of microcalcifications vs. other types or all types of lesions (Bluemke et al.[30] and Van Goethem et al.[51]) and reported similar results: the sensitivity of MRI for evaluation of microcalcifications is approximately 85 percent, which is considerably lower than the sensitivity of	☐	New Evidence:	☐

Conclusions From CER Executive Summary	Is this conclusion almost certainly still supported by the evidence?	Has there been new evidence that may change this conclusion?	Do Not Know
MRI for evaluation of all/other types of lesions; whereas the specificity of MRI for evaluation of microcalcifications is approximately 77 percent, which may be slightly higher than the specificity of MRI for evaluation of all/other types of lesions.			
Two studies evaluated the accuracy of MRI for dense breasts vs. all or non-dense breasts (Bluemke et al.[30] and Wiberg et al.[40]), and reported virtually no difference in the accuracy of MRI for evaluation of these different categories of breast tissue.	☐	New Evidence:	☐
One study enrolled only patients with lesions classified as BIRADS 3 before investigation by MRI (Gokalp and Topal[24]); however, only one enrolled patient (out of 43 total) was found to have a malignancy and therefore the patient population is too small to draw conclusions about the accuracy of MRI for probably benign lesions.	☐	New Evidence:	☐
One study each investigated the accuracy of MRI for lesions broken down by palpable vs. non-palpable (Bluemke et al.[30]) and large lesion vs. small lesion (Imbracio et al.[44]).	☐	New Evidence:	☐
Position Emission Tomography			

Conclusions From CER Executive Summary	Is this conclusion almost certainly still supported by the evidence?	Has there been new evidence that may change this conclusion?	Do Not Know
In three of the seven studies that addressed Key Question 1, the majority (>75.0%) of the women presented with palpable breast lesions— Kiada et al.[52]: 88.0 percent palpable, Schirrmeister et al.[54]: 76.0 percent, and Yutani et al.[55]: 93.0 percent palpable. Because there were only three studies, we could not fit the data in a bivariate model. Instead, we pooled the reported sensitivities and specificities in random-effects meta-analyses. However, the data were heterogeneous (I^2 = 68.0% and I^2 = 54.6% for sensitivity and specificity, respectively), indicating substantial variability among the study results. With only three studies, we did not attempt to explore possible reason(s) for the heterogeneity. The overall sensitivity for primarily palpable lesions is higher than that for all seven studies considered under Key 1 (86.5% vs. 83.0%), but the specificity is lower (64.2% vs. 74.0%).	☐	New Evidence:	☐
One study directly compared images acquired when patients were in prone position to images of the same patients in supine position.[52] In this study by Kaida et al. 2008, 118 women with 122 lesions suspected of having breast cancer underwent whole-body PET in the supine position immediately followed by prone	☐	New Evidence:	☐

Conclusions From CER Executive Summary	Is this conclusion almost certainly still supported by the evidence?	Has there been new evidence that may change this conclusion?	Do Not Know
breast PET imaging. According to the results reported in the study, the sensitivity and specificity of images in the supine position were 83.0 percent and 50.0 percent, respectively. The sensitivity and specificity of images in the prone position were 96.0 percent and 50.0 percent, respectively.			
One study, Yutani et al. 2000, reported results separately for patients with BIRADS 5, lesions 1.5 cm or larger, and who were younger than 65.[55] The authors reported that PET was more sensitive for larger lesions, but the specificity was unchanged; and for the other factors, the accuracy of PET was virtually the same as for PET for all patients.	☐	New Evidence:	☐
Scintimammography			
Two studies evaluated only patients with palpable breast masses,[57,62] one study evaluated only patients with non-palpable breast masses,[63] and one study evaluated only patients with microcalcifications detected on x-ray mammography.[61] With so few studies reporting on each category, evidence-based conclusions are difficult to support.	☐	New Evidence:	☐
None of the studies reported outcomes by patient demographics or any other clinical risk factors that may have affected the accuracy of SMM.	☐	New Evidence:	☐

Conclusions From CER Executive Summary	Is this conclusion almost certainly still supported by the evidence?	Has there been new evidence that may change this conclusion?	Do Not Know
Ultrasound			
None were identified.			
Key Question 3. Are there other factors and considerations (e.g., safety, care setting, patient preferences, ease of access to care) that may affect the accuracy or acceptability of the tests considered in Key Questions 1 and 2?			
Magnetic Resonance Imaging			
One study reported the accuracy of MRI images interpreted with and without a Computer Aided Diagnosis (CAD) software system.[12] The study reported virtually no difference in either sensitivity (77.4% vs. 78.9%) or specificity (73.2% vs. 73.2%) with or without CAD assistance.	☐	New Evidence:	☐
Position Emission Tomography			
None of the seven studies on stand-alone PET scanning or the one study on PET with CT reported information that addressed this question.	☐	New Evidence:	☐
Scintimammography			
None were identified.	☐	New Evidence:	☐
Ultrasound			

Conclusions From CER Executive Summary	Is this conclusion almost certainly still supported by the evidence?	Has there been new evidence that may change this conclusion?	Do Not Know
None were identified.	☐	New Evidence:	☐
Are there new data that could inform the key questions that might not be addressed in the conclusions?			