Comparative Effectiveness Review Disposition of Comments Report

Research Review Title: *Core Needle and Open Surgical Biopsy for Diagnosis of Breast Lesions: An Update to the 2009 Report*

Draft review available for public comment from January 30, 2014 to February 26, 2014.

Research Review Citation: Dahabreh IJ, Wieland LS, Adam GP, Halladay C, Lau J, Trikalinos TA. Core Needle and Open Surgical Biopsy for Diagnosis of Breast Lesions: An Update to the 2009 Report. Comparative Effectiveness Review No. 139. (Prepared by the Brown Evidence-based Practice Center under Contract 290-2012-00012-I.) AHRQ Publication No. 14-EHC040-EF. Rockville, MD: Agency for Healthcare Research and Quality. September 2014. 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 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 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.

Source: www.effectivehealthcare.ahrq.gov/search-for-guides-reviews-and-reports/?pageaction=displayproduct&productID=1960
Published Online: September 10, 2014

Commentator & Affiliation	Section	Comment	Response
Reviewer #1	01. General	Yes. Yes. Yes.	Thank you. No further response required.
Reviewer #1	03. Introduction	No specific comments.	No response required.
Reviewer #1	04. Methods	Inclusion criteria and exclusion criteria are well justified. The search strategies are logical, involving eight electronic databases to identify the relevant prospective or retrospective cohort studies. Outcome measures and statistical methods are well justified in the document and seem valid.	Thank you. No further response required.
Reviewer #1	04. Methods	Meta-analysis of nine studies reporting the odds of requiring only one surgical procedure were demonstrated to be 13 times higher among women receiving core biopsy as compared to those receiving open surgical biopsy, a significant and important outcome.	We agree that this is an important finding. We believe we have discussed this finding adequately and have highlighted it in the Report's Abstract and Executive Summary.
Reviewer #1	04. Methods	The discussions and determinations of the risk of bias across the studies considered in this update were conservative and no significant bias should be present in this report. However, this has resulted in some clinical data being down-graded as lower in strength than it probably should be.	Thank you. We have described the items we considered in the assessment of risk of bias in the revised report. The complete dataset of extracted information (i.e. the individual risk of bias assessments for each study) are provided in SRDR. Risk of bias ("or quality") assessment is by definition subjective and cannot always distinguish between poor reporting and poor design or conduct. It is not clear to which "clinical data" the reviewer is referring to. We have provided a detailed assessment of strength of evidence in the revised report (separately for comparative and non-comparative data). We believe that our grading of the strength of evidence is reasonable and we have provided a detailed rationale for all our dispositions.
Reviewer #1	05. Results	The amount of detail presented in the results section is appropriate, well referenced, and supported by additional data in the appendices. Figures, tables, and appendices are appropriately detailed and descriptive. The key messages are explicit and applicable to current clinical practice.	Thank you. No further response required.
Reviewer #1	05. Results	Re-examining the accuracy outcomes for core biopsy of DCIS is an important and significant update to this report.	Thank you. No further response required.
Reviewer #1	05. Results	Addressing the issue of potential dissemination of cancer cells by the biopsy procedure (seeding) was a second important consideration for this update, though this has never been considered a risk in clinical practice.	Thank you. No further response required.

Commentator & Affiliation	Section	Comment	Response
Reviewer #1	06. Discussion/Conclusion	There is one important and critical aspect that underpins both image guided core biopsy and open surgical biopsy. Both rely on the quality of imaging to depict and locate a lesion in an individual woman's breast. The ability to locate the lesion can be effected by the breast density and the location of the lesion within the breast. While it is apparent that this is particularly important for image guided core biopsy, it is equally important for open surgical biopsy. Most open surgical biopsies are preceded by an image guided needle wire localization procedure. The same imaging types of studies are used to perform image guided wire localization as are used for image guided core biopsy. Both are subject to the same issues of breast density and lesion location/access with needles using imaging guidance. Therefore, the accuracy of image guided core biopsy and open surgical biopsy preceded by needle wire localization are subject to the same biases and limitations introduced by various imaging methodologies to locate lesions within the breast. This is something that was not explicitly stated in this review, nor in the earlier version of this review from years prior – and is a potential source of bias in this report. Only open surgical biopsies, guided by palpation (a minority of all open surgical biopsies performed) would not be subjected to this bias.	We agree that imaging is an important determinant of the performance of both open and core needle biopsy methods and have added this information to the Introduction and Discussion section of the revised report. However, we do not believe that not mentioning this fact was a source of bias, for the individual studies or for the body of evidence we reviewed (because not mentioning background information - which we were aware of - cannot bias the assessment of actual empirical evidence).
Reviewer #1	06. Discussion/Conclusion	The report suggests "future research needs" include studies of test performance to evaluate MRI-guided biopsy methods. One very important consideration in making this statement is that most lesions identified on MRI are then located using "second look ultrasound" and, if seen by US, the standard of care would be to biopsy these lesions using ultrasound guidance. The MRI detected lesions remaining (i.e., those not seen with second look ultrasound), are then biopsied by MRI; clearly, these are a very different and more difficult subgroup of lesions that remain after "easier" lesions are biopsied with ultrasound guidance.	Thank you. We agree that this is an important point and have mentioned it in the Future Research Needs section of the report.

Commentator & Affiliation	Section	Comment	Response
Reviewer #1	06. Discussion/Conclusion	This analysis is also similar to differences that might be detected between ultrasound guided core biopsies and stereotactic core biopsy. Ultrasound is the standard of care and first choice for any soft tissue mass that is visible with ultrasound because biopsy using this technique is easier, quicker, less expensive and more comfortable for the patient. Lesions not well visualized with ultrasound (i.e., micro calcifications not visualized by ultrasound) are then biopsied using stereotactic guidance. Again, this comprises a more difficult group of lesions than lesions typically biopsied using ultrasound guidance. To reiterate, MR guidance is only used to biopsy lesions that are not visualized by any other imaging modality.	We agree that this is an important point and have incorporated related information in the Discussion section of the revised report.
Reviewer #1	09. Clarity/ Usability	The report is well structured and organized. The target questions are well articulated and appropriately researched. Given limitations cited above, the conclusions in this report are useful to inform policy and/or practice decisions.	Thank you. Please see above for our responses to the relevant comments.
Reviewer #2	01. General	In general, this study has lost much of its practical relevance. Core biopsy has become the standard method of diagnosis for non-palpable breast lesions in the US today. That being said, the target population and audience are well defined and the key questions clearly stated.	We agree that in clinical practice the question about the relative merits of open versus core needle biopsy is considered settled. However, we believe that the evaluation of alternative core needle biopsy methods (a key aspect of this update) is a valuable contribution. Furthermore, new biopsy methods (e.g., MRI-guided biopsy) appear to represent areas for additional research.
Reviewer #2	03. Introduction	No comments.	No response required.
Reviewer #2	04. Methods	The methods are appropriate.	Thank you. No further response required.

Commentator & Affiliation	Section	Comment	Response
Reviewer #2	05. Results	Page ES-10 and 21 - It is unclear to me if for the purpose of the study a false negative core biopsy is one in which the biopsy is negative for carcinoma, but the subsequent surgery reveals a malignancy (therefore including lesions such as atypical duct hyperplasia - known to have a risk of upstage due to sampling, e.g.) or if it is meant as cases in which the lesion was missed entirely and treatment was delayed. These are 2 very different scenarios and I suspect that latter was the authors' intent. It is even less clear what is meant by false positive. If every malignancy-containing core biopsy is correctly diagnosed as malignancy, then there is a zero false positive rate. If individual cases reveal no residual tumor in the surgical specimen, this simply means the malignancy was removed entirely by the core, not that it did not exist. In this context, false positives are errors in core biopsy diagnosis. I doubt the authors were able to assess this. The issues/definitions of false negative and false positive need clarification	In the revised report we have provided additional details about the definition of each possible diagnostic test result. Please note that when the biopsy removed an entire lesion we have considered the biopsy results as “true” (TP or TN, depending on the pathology results). We have also performed additional sensitivity analyses to evaluate alternative definitions of FP findings.
Reviewer #2	05. Results	Key question 2- ES11-12, 46 The authors need to find a different word for "dissemination". Displacement of tumor cells (or benign cells) due to core biopsy and occasional mechanical transport of them to sentinel lymph nodes is NOT dissemination. Dissemination in current usage implies systemic spread throughout the body and has a very foreboding implication for an issue which is of importance to pathologists and oncologists in recognizing what it is not (namely metastatic disease). Calling it dissemination will cause undue alarm to an astute public reading this study.	When appropriate we have changed ‘dissemination’ to ‘displacement’ in both the Executive Summary and the Main Report Results section for Key Question 2 in order to be more accurate about the phenomenon. When referring to all studies deemed relevant to this Key Question we use the term “studies of dissemination or displacement”, to capture both studies of new tumor formation on the needle tract (a rare event) and the displacement of cells (e.g., to the lymphatic circulation – an outcome of unclear clinical significance).

Commentator & Affiliation	Section	Comment	Response
Reviewer #2	05. Results	ES17, 48 It is extremely surprising to me that the authors could find few studies addressing the use of MRI-guided core biopsy based on increased risk, since this in in fact one of the main uses of MRI breast assessment. The authors are incorrect in stating that "MRI is likely reserved for diagnostically challenging cases". This is simply not the case, at least in this reviewer's fairly extensive experience.	We have reviewed several dozen studies of MRI that did not meet out inclusion criteria. We also contacted all TEP members to ask for specific citations of MRI studies that we should consider. No additional studies meeting our criteria were identified via this process. Regarding the use of MRI in “diagnostically challenging cases”, we respectfully disagree with the reviewer. Several studies explicitly state that MRI was reserved for patients who could not be effectively imaged with other techniques (e.g. US). We have clarified our wording to indicate that the “diagnostic challenge” pertains to the inability to successfully use imaging methods other than MRI.
Reviewer #2	06. Discussion/Conclusion	No comment.	No further response required.
Reviewer #2	09. Clarity/ Usability	See general comments section above.	No response required.
Reviewer #3	01. General	The target population and audience are explicitly defined, and the key questions are appropriate and explicitly stated. The clinical meaningfulness of the report is limited by the quality of the evidence base.	Thank you. No further response required.
Reviewer #3	03. Introduction	The key questions are appropriate and explicitly defined. It appears that the question about MRI-guided CNB was premature, given the amount of evidence that turned out to be available.	EPC reports aim to identify, refine, and (if the literature permits) address Key Questions that are clinically important. MRI-guided biopsy was identified as an important technique to consider during the updating survey and the Key Informant and Technical Expert Panel discussions. When the available research evidence does not permit an answer to clinically important questions, the question can be prioritized as a research gap. In that sense, the MRI-guided biopsy question is helpful in identifying one area for potential research.
Reviewer #3	04. Methods	The inclusion and exclusion criteria for the studies are justifiable. The search strategies are explicitly stated and logical. The definitions used for the outcome measurements are appropriate. I am not a statistician, so will not comment on the choice of statistical methods.	Thank you. No further response required.

Commentator & Affiliation	Section	Comment	Response
Reviewer #3	05. Results	I am not aware of any studies that were missed or were included inappropriately. Figures and tables are adequate and descriptive. The characteristics and the limitations of the studies are very clearly described.	Thank you. No further response required.
Reviewer #3	05. Results	In several places the report notes that the ratings in this report are not directly comparable with those of the original report. Why did this report shift to analysis of comparative test performance?	We believe that comparisons among diagnostic tests are directly informative for clinical decisionmaking (unlike assessments of individual tests in isolation). For this reason, we emphasized assessment of the strength of evidence for comparative outcomes. However, for consistency with the original 2009 evidence report, we have added an assessment of the strength of evidence for non-comparative test performance outcomes.
Reviewer #3	06. Discussion/Conclusion	As a non-statistician, it is hard to know what to make of the series of tables F, G, H in which not ONE comparison had an overall rating better than "Low," and the overall emphasis throughout the report on the limitations of the study. Should we take comfort in the fact that differences in performance among alternative biopsy methods seem small across multiple studies of poor quality? Or did the entire effort to assemble and analyze the collection of studies yield no real value because of the poor quality of the evidence? The report is motivated by an interest in updating the 2009 report for questions about DCIS, MRI-guided CNB and freehand automated device CNB. It seems clear that the examination of MRI-guided CNB was premature, and additional studies are needed. But in what way did the conclusions of this report extend those of the prior report with respect to the other two questions?	Please note that assessment of the strength of evidence does not rely exclusively on statistical considerations. We believe that the changes we have made to the strength of evidence assessment (e.g., the inclusion of non-comparative outcomes) should address the reviewer's concerns. We believe that the evidence on each test of interest is of moderate strength. Informal comparisons across tests suggest that – when used in the populations where they are considered applicable – all image-guided methods have similar test performance. We think that this conclusion is a reasonable interpretation of our systematic review (with respect to test performance).
Reviewer #3	09. Clarity/ Usability	The report is clearly organized. It seems longer than necessary because the executive "summary" is extensive, and much of the text in the summary is repeated in the body of the report.	We have streamlined the Executive summary to the extent possible. Current guidance for our evidence reports requires that the Executive Summary be a standalone document, thus some duplication of information between the Summary and the Main Report text is unavoidable.
Reviewer #4	01. General	The report is very clear. The 3 questions addressed in the report are clearly stated, and so is the methodology used to conduct the research, as well as the results. If anything, questions, methodology and results are reiterated a number of times, making the report a bit repetitive.	Thank you for your comments. We have streamlined the Executive Summary; however, some degree of duplication between the Executive Summary and the Main Report text is unavoidable.

Commentator & Affiliation	Section	Comment	Response
Reviewer #4	03. Introduction	The introduction clearly states the problems, and specifically indicates that the current review constitutes a continuation of the prior review released in 2009.	Thank you. No further response required.
Reviewer #4	04. Methods	Inclusion and exclusion criteria used to selected the studies are justifiable, and the search strategies are clearly stated, especially with regard to use of articles that had been excluded from the 2009 report, but have been included in this review. Although the definitions of true positive, true negative and false negative results is easy to understand, the definition of false positive cases is not clearly stated in the review (at least I could not find it, even by electronically searching the pdf numerous times). I think that the definition of false positive cases should be clearly stated in this report.	Thank you. We have provided additional information on the definition of all diagnostic test result categories.
Reviewer #4	05. Results	The results are presented clearly and systematically. The supportive studies and their conclusions are clearly summarized.	Thank you. No further response required.
Reviewer #4	05. Results	Regarding false positive cases. As I commented above regarding the methods section, the definition of false positive cases should be clarified. In particular, if a core biopsy diagnosis of invasive carcinoma or DCIS was rendered, the absence of carcinoma in the open biopsy specimen does not mean that the core dx was incorrect, because the core biopsy might have completely removed the carcinoma, and the malignant dx still applies. This point is made in some of the articles (ref 1734. Rakha EA, Ho BC, Naik V, et al. Outcome of breast lesions diagnosed as lesion of uncertain malignant potential (B3) or suspicious of malignancy (B4) on needle core biopsy, including detailed review of epithelial atypia. Histopathology 2011 Mar;58(4):626-32.), but is not clearly conveyed in the review. Furthermore, the number of false positive diagnoses for different diagnostic procedures mentioned in the review and depicted in the graphs, seems to be relatively high, and should be checked again carefully.	In the revised report, we have provided additional details about the definition of each possible diagnostic test result. Please note that when the biopsy removed an entire lesion, we have considered the biopsy results as true (TP or TN, depending on the pathology results). We have also performed additional sensitivity analyses to evaluate alternative definitions of FP findings.
Reviewer #4	05. Results	Page 43: Wait time for test results. The Authors mention a study that used "a microwave processor to reduce wait times for test results reduced the average wait for results (P<0.001). ref 255" I recommend to remove mention of this study. Its results are misleading and not in compliance with the current ASCO/CAP guidelines for processing of breast tissue suitable for evaluation of ER, PR and HER2 status. The use of rapid fixation and/or tissue processing could introduce processing artifact and affect tissue immunoreactivity, and should not be mentioned in this review.	Because our selection criteria did not specify a specific breast biopsy tissue processing method, we have opted to retain this study in Key Question 3. However, to address the reviewer's concern, we have explicitly stated that the study evaluated a non-standard processing method that is not in widespread use.

Commentator & Affiliation	Section	Comment	Response
Reviewer #4	06. Discussion/Conclusion	The clinical implications and limitations of the review are clearly stated. In particular, it is somehow surprising that the data regarding MRI-guided cbx for evaluation of breast lesions in women at average and high risk was too limited for definitive conclusions. The future research section also clearly indicates possible future studies, MRI-guided core biopsy included.	Thank you for this comment. We have reviewed several dozen studies of MRI that did not meet out inclusion criteria. We also contacted all TEP members to ask for specific citations of MRI studies that we should consider. No additional studies meeting our criteria were identified via this process.
Reviewer #4	09. Clarity/ Usability	The report is clear and well structured. The conclusions are unbiased and informative.	Thank you. No further response required.
Reviewer #5	01. General	This is a very well done systematic review. See attached review.	Thank you. No further response required.
Reviewer #5	03. Introduction	See attached review.	No further response required.
Reviewer #5	03. Introduction	Same point again about the "large proportion" of women undergoing biopsy over a 10 year period. It is actually a small proportion (7% or less). Interestingly, the Hubbard, et al article did estimate risk of biopsy by age and density, which might be worth noting, ie, that some women have a higher likelihood of undergoing biopsy.	This has been corrected.
Reviewer #5	Abstract	Is diagnosis the right word? The biopsy is a procedure to gather tissue for the further evaluation of a suspicious lesions…perhaps it is part of the diagnostic process, but it is not "for diagnosis."	We have used the term "in the diagnostic assessment".
Reviewer #5	Abstract	Do you mean obviated the need for surgical biopsy procedures? Just saying surgery implies that the biopsy was therapeutic and no additional procedures was needed. Note text on page vi, line 24-25 for comparison.	We used the term "additional surgical procedures". In some cases no additional surgery was performed (because the lesion was deemed benign).
Reviewer #5	02. Executive Summary	Probably the prevalence of women with a prior diagnosis of breast cancer is not a relevant statistic	We only use this statistic to emphasize the public health importance of breast cancer. Most of these women have at one time received a breast biopsy.
Reviewer #5	02. Executive Summary	If the emphasis is on asymptomatic breast cancer, remove detection by self-exam and physical exam…detection of symptoms is not detection of asymptomatic breast cancer. It isn't really clear what point is being made here…if mammographic signs or physical symptoms are suspicious, tissue will need to be gathered to rule out breast cancer. Suggested rewrite: "Because the earliest stages of breast cancer are often asymptomatic, the process of breast cancer diagnosis is often initiated by detection of an abnormality through screening mammography, although suspicious palpable abnormalities detected by a woman or clinician also need further evaluation. If the…."	Thank you. We have adopted the suggested wording.

Commentator & Affiliation	Section	Comment	Response
Reviewer #5	02. Executive Summary	Actually, the 10 year false positive biopsy rate is 7% for annual screening and 4.8% for biennial, each considerably higher than odds of a true positive biopsy over 10 years. I don't think that amounts to a "large proportion" compared with the rate of at least one positive finding. Reference 3 is out of date—I'm also surprised that it qualified under inclusion criteria…it does not provide direct estimates, as does reference 2.	We have corrected this statement. Please note that studies cited in the Introduction are not selected on the basis of selection criteria used for the systematic review (e.g., screening results are reported for context). We think reference 3 offers relevant information.
Reviewer #5	02. Executive Summary	Not sure what is meant by "the ductal carcinoma in situ (DCIS) underestimation rate of stereotactically guided vacuum-assisted core-needle biopsy." Could this be stated differently?	We have provided clear operational definitions for underestimation in the revised report.
Reviewer #5	02. Executive Summary	Probably the report should explain central credibility intervals. Most people (apart from Bayesians) will never have heard of them.	We have added a non-technical definition of these intervals.
Reviewer #5	02. Executive Summary	At this point, I don't think there has been a clear explanation of the issue around underestimation of DCIS, or underestimation of high risk DCIS lesion underestimation rate. Perhaps this is clearer in the larger report, but so far in the executive summary there does not seem to be sufficient explanatory material on these two issues. Also, in Table C, it is not clear whether "high risk" refers to high risk women, or a high risk lesion. Several other places in the text also are not clear between the distinction of lesion level or woman level. Suggest a search on these terms throughout the document to double check for clarity.	We have clarified these terms in the ES. Specifically we added the following text: *"We defined the underestimation rate for high risk lesions (most often atypical ductal hyperplasia, ADH) as the proportion of core needle biopsy findings of high risk lesions that are found to be malignant according to the reference standard). We defined the underestimation rate for ductal carcinoma in situ (DCIS) as the proportion of core needle biopsy findings of DCIS that are found to be invasive according to the reference standard."*
Reviewer #5	02. Executive Summary	For key question 13, on the issue of choice, the implication is that women have a role in the choice of biopsy method. A question not addressed, is what factors are associated with the decision to use one vs. the others, and how often those choices are defined by risk and lesion characteristics, and whether these choices differ by professional subgroup.	We agree that this is an important issue. We have discussed how patient and lesion characteristics often guide the choice among biopsy methods. However, a systematic review of factors determining this choice was out of the scope of the current review.
Reviewer #5	02. Executive Summary	Table F. Evidence reports commonly provide strength of evidence or strength of recommendation grades without specific explanations for why the overall rating of the strength of evidence or rating was low, medium or high. Earlier examples were given, but it important, I think, to state the basis for the rating. Some rows in the table provide some insights, but others provide only conclusions from the assessment of the evidence.	The Strength of Evidence information in the executive summary is summarizing the complete assessment which is provided in the Appendix. We have referred readers to that source.

Commentator & Affiliation	Section	Comment	Response
Reviewer #5	02. Executive Summary	It may be implied, but perhaps not—either way the point could be brought out more clearly. Open biopsy is often chosen if the radiologist or surgeon would not trust benign findings from the core or needle biopsy. In other words, the lesion has an appearance that wouldn't allow acceptance of negative findings with confidence.	Thank you for this information. We have discussed in general terms the choice among alternative biopsy methods.
Reviewer #5	02. Executive Summary	General comment. I may have overlooked this, but I don't have the impression that there is a clear comparison of "what's new" since the 2009 report.	We have provided this information throughout the report.
Reviewer #5	03. Introduction	Same comment as above. If you wish to make a point about asymptomatic breast cancer, separate BSE and clinical detection from that point. Doesn't really make sense as described.	We have adopted the suggested phrasing regarding the diagnosis of breast cancer. Please see above for our response.
Reviewer #5	04. Methods	Inclusion and exclusion criteria are reasonable, as were search strategies, which were well described. Outcomes measures are also reasonable. Statistical methods are appropriate, but at a level of complexity that likely will be over the heads of nearly all individuals who will read this report.	Thank you for your comments. We have simplified the description of the statistical methods in the Executive Summary, but have retained a more detailed description in the Main Report text.
Reviewer #5	05. Results	Results: Detail is adequate, but suggest a clearer comparison of the new findings compared with the 2009 report. Findings are straightforward and well described, but I would suggest more "accessible" descriptions of the statistical methods. Explanations for grade scores typically are general and commonly not instructive with respect to individual manuscripts, i.e., "why was the risk of bias high?"	Thank you for your comments. We have simplified the description of the statistical methods in the Executive Summary, but have retained a more detailed description in the Main Report text. We have also provided more details about the changes from the 2009 version of the report. Please note that the assessment of risk of bias of individual studies is provided with the complete data extraction forms (on SRDR, to be released upon completion of the report). Given the very large number of studies, we think it is not possible to discuss each individual study in the report. However, we do discuss the various dimensions of risk of bias in more detail in the revised report.
Reviewer #5	06. Discussion/Conclusion	Unfortunately, there appear to be no data to describe the decision making process for one type of biopsy vs. another. If these data exist, some discussion would be useful.	We have provided some additional discussion of the factors that guide choice among alternative biopsy methods and their importance in designing future research studies.
Reviewer #5	09. Clarity/ Usability	The report is well constructed and organized. Although policy implications are not discussed directly, they are evident.	Thank you. The revised discussion section highlights some clinical, policy, and research implications of our findings. Of note, EPC reports do not make recommendations about clinical practice or policy.

Commentator & Affiliation	Section	Comment	Response
Reviewer #6	01. General	Is the report clinically meaningful? Within the scope of what was commissioned and intended, the report provides an excellent summary of the state of current knowledge.	Thank you. No further response required.
Reviewer #6	01. General	However, the focus on comparisons between core needle biopsy and open biopsy is not clinically relevant today. Although open biopsies are still occasionally performed for no good reason, this is a minor issue. Thus the first sentence of the abstract is not correct ("Core-needle biopsy and open surgical biopsy are the most frequently used procedures for diagnosis of suspicious breast lesions"). Core needle biopsy is now available and practiced across the country; it is the most common diagnostic procedure. Differences between mammogram, ultrasound, and MRI-guided biopsies are somewhat more meaningful, but a comparison across these is limited by the fact that the selection of the imaging modality that guides the biopsy is driven by the modality that finds the lesion. So, if calcifications are identified on mammography, ultrasound guidance is not feasible and MRI guidance is not necessary. This report will therefore inform radiologists and surgeons that if a lesion is seen on multiple modalities, the choice of X to guide the biopsy will lead to a shade higher accuracy, and possibly a shade lower harm. However, if a lesion is seen only on a single modality, there is no choice available as to which modality will guide the biopsy and differences in performance or complications are besides the point.	Please note that the report provides details about alternative core needle biopsy methods and comparisons among them, in addition to comparisons of open and core needle biopsy. Regarding the importance of lesion characteristics in determining the imaging guidance method, we have adopted some of the reviewers thinking in the Introduction and Discussion sections of the revised report.

Commentator & Affiliation	Section	Comment	Response
Reviewer #6	01. General	A notable omission from this report (again, not a reflection on the excellent work of the authors) is the diagnostic yield of each biopsy method. This is referred to only once in the entire report, with one reference from a community hospital (Hyser MJ, Am Surg 2000;66(5):438-42). There is a discussion of "harms" and concern regarding incomplete reporting of harms. But the greatest harm of a biopsy is the very fact of undergoing a benign biopsy. Benign biopsies cause tremendous stress and anxiety, influence surgical decisions, cost money, all of which contribute to women seeking surgical options that are not medically indicated, with additional downstream harms. Therefore, diagnostic yield is a crucial aspect of any examination of the performance of various techniques of biopsy. Technically, this expands the scope of the review to include the threshold at which biopsies are being recommended. But any meaningful evaluation of biopsy performance has to include the factors that prompt the biopsy in the first place. The surest route to reducing the harm of biopsies, however they are done, is to reduce the biopsy rate for American women. This aspect probably cannot be included in the current report, but is a major issue affecting the quality of life of women undergoing mammographic screening. It deserves close attention.	Diagnostic yield is defined variably by different authors. To address the reviewer's comments, we have added the following analyses to the report: (1) a descriptive statistical analysis of the proportion of patients eventually diagnosed with cancer, among those who underwent biopsy (TP + FN)/(TP + FP + FN + TN) (2) a descriptive statistical analysis of the proportion of correct diagnostic results (TP + TN)/(TP + FP + FN + TN) We note that Key Questions 2 and 3, address harms and psychological consequences of breast biopsy to the extent permitted by the information reported in published studies. Please note that screening-related issues are beyond the scope of the report.
Reviewer #6	03. Introduction	Are the target population and audience explicitly defined? The audience is not explicitly defined, but appears to be practitioners who care for women undergoing breast surveillance, mainly radiologists and surgeons. Surgeons benefit from information that open surgical biopsy is rarely needed, and is no superior to core needle biopsy. And radiologists benefit from knowledge about relative utility of the different needle-based approaches.	Various stakeholders – including patients and patient advocates, frontline clinicians, clinical researchers, funders of research, payers of health care, and manufacturers of medical technologies – use Evidence-based Practice Center reports. The audience for this particular report is implicit in the Key Questions addressed (i.e., any party interested in the answers to the pre-specified Key Questions).

Commentator & Affiliation	Section	Comment	Response
Reviewer #6	03. Introduction	However, the utility is again limited by the fact that the choice of image guidance for CNB is dictated by the imaging technique that allows the lesion to be visualized. There does not seem to be any intent to target lay organizations, advocates, lay public, policy makers, the insurance industry, the device industry, or other diverse parties, although the report will be available to them.	We have revised the report to address more explicitly the issues related to the choice of imaging techniques as they pertain to test performance comparisons. As stated in the preceding row of this table, various stakeholders – including patients and patient advocates, frontline clinicians, clinical researchers, funders of research, payers of health care, and manufacturers of medical technologies – use Evidence-based Practice Center reports. The audience for this particular report is implicit in the Key Questions addressed (i.e., any party interested in the answers to the pre-specified Key Questions).
Reviewer #6	03. Introduction	Are the key questions appropriate and explicitly stated? The relevance of the key questions vis a vis current American practice has been discussed above. The questions are precisely framed and convey the intent of the analysis.	Thank you. No further response required.
Reviewer #6	04. Methods	Are the inclusion and exclusion criteria justifiable? The following inclusion criterion is thresholded too low and needs justification: "enrolled 10 or more patients and followed at least 50 percent of them to the completion of the study". A more appropriate standard would be studies that enrolled 50 or more patients and followed for one year. If one year f/u results in the loss of many studies this could be relaxed; but including studies with an N of 10-50 likely increases the variability considerably, and adds little to the overall analysis results. Also, and allowable attrition rate of 50% is too high, 30% would be better. The authors note the biases that can be introduced with small samples and high attrition, but are too generous in what they find acceptable.	We selected these inclusion criteria to be comprehensive (i.e. to include as many informative studies as possible). The issue of "variability" raised by the reviewer does not affect our analyses, as within-study variability is modeled appropriately in the hierarchical meta-analysis model. Differences in followup duration were addressed by appropriate subgroup analyses. Finally, the selection criteria we used were chosen to be consistent with those of the original 2009 report, which facilitates the synthesis of the entire body of evidence.
Reviewer #6	04. Methods	The search strategies are explicitly stated and logical.	Thank you. No further response required.
Reviewer #6	04. Methods	Are the definitions or diagnostic criteria for the outcome measures appropriate? For Key Question 2, the outcomes included rate of inconclusive biopsy findings (e.g. inadequate sampling of lesion); it is not clear how this differs from a false negative result. If it does not, the term "false negative" should be used to describe all findings where surgical biopsy or 6-month follow up showed cancer. It is possible that by "inconclusive" the authors are referring to atypical lesions where current standard of care is to perform a follow-up surgical biopsy.	We have provided additional details regarding the definitions of non-diagnostic/inconclusive samples and diagnostic test categories (including false positives) in the revised report.

Commentator & Affiliation	Section	Comment	Response
Reviewer #6	04. Methods	Another listed outcome is “repeat biopsy rate”; presumably this too refers to the situation described above (atypical lesion requiring surgical biopsy for definitive diagnosis). If so, this is a redundant outcome and should be removed.	Repeat biopsies may be required because the results of the first biopsy were non-diagnostic. Given the potential psychological effects of additional biopsies and the delays in diagnosis and treatment that are introduced by non-diagnostic/inconclusive biopsies, we have opted to retain this outcome. The Key Informants and Technical Experts involved in the development and refinement of this report identified this as a clinically relevant outcome, and it was also included in the original 2009 evidence report.
Reviewer #6	04. Methods	A third unclear outcome is “subsequent false positive and false negative rates on mammography”; the meaning of this is opaque. It appears this refers to mammograms that are performed following the index biopsy. How would one decide that a subsequent mammogram was false negative or false positive?	This outcome was meant to capture the long-term impact of breast biopsy on future mammographic examinations. As for any imaging test for the diagnosis of breast cancer, the true disease status would be established by either biopsy examination of resected/biopsied tissue or long-term followup. This has been clarified in the text.
Reviewer #6	04. Methods	For Key Question 3, listed outcomes include “recurrence rate (for women with cancer, including local, regional, and distant recurrence), cancer-free survival and overall survival]”. These are not appropriate for an analysis of biopsy methods. The determinants of recurrence and survival go far beyond the biopsy method. There is nothing useful to be gained from including these events as outcomes for this report.	Long-term cancer outcomes represent clinical indirect effects of testing (via the impact of tests on diagnostic thinking and therapeutic choices). As such, we believe that they are relevant to the assessment of a diagnostic test, even if current research has not addressed them.
Reviewer #6	04. Methods	One outcome which is not addressed relates to the failure of wire localized surgical excision of non-palpable lesions. With current standards of care, wire-localized surgical excision should only occur if a cancer has been diagnosed (in which case it is a therapeutic procedure) or when the CNB produces ambiguous results, e.g. ADH. In the latter situation, it is possible for the surgical biopsy to fail because of poor localization or poor surgical technique, or patient-related reasons (large, mobile breast). This is usually realized during or soon after surgery because the biopsy clip is not seen on the specimen radiograph. Although rare, this is an adverse outcome which should be captured.	The number of open breast biopsy studies included in the report was small, which explains the absence of studies reporting the information requested by the reviewer. We considered failed biopsy procedures (regardless of reason) in our data extraction, but the studies often reported only the aggregate number of inconclusive/non-diagnostic procedures and did not report specific reasons (e.g., failure to localize).

Commentator & Affiliation	Section	Comment	Response
Reviewer #6	04. Methods	Are the statistical methods used appropriate? I cannot comment in any depth on the statistical methods, but I find the specificity analyses to be problematic. This may just be my ignorance, but here is the difficulty: the authors define specificity as 1-false negative. However, most specificity definitions are based on the false positive rate, which is defined in Table 1 on page 38/237 to be a high risk lesion that is benign on surgical excision. If the pathologist diagnoses ADH or ALH or radial scar on a core biopsy and a follow-up surgical biopsy shows no cancer, this does not mean that the core biopsy diagnosis was incorrect. With this consideration, a false positive result of core needle biopsy is impossible. An explanation would be helpful. This is an important issue since the great majority of current surgical biopsies are performed for these indeterminate lesions.	A typo on page 6 of the main report (definition of specificity) has been corrected. Thank you for pointing this out. For all analyses, specificity was defined as: Specificity = TN / (FP + TN) Note that 1 – FPR = = 1 – FP / (FP + TN) = = TN / (FP + TN) = Specificity When biopsy removed the entire lesion and no lesion tissue was obtained by subsequent biopsies or surgery, we considered the biopsy results to be "true" (TP or TN, depending on the index test). However, when the index biopsy suggested the presence of a high-risk lesion (or malignancy) and subsequent surgery or followup did not reveal malignancy, we considered the test results false positive. This is consistent with the operational definitions used in the original report. Further, in the revised report we have performed a sensitivity analysis by repeating all analyses after excluding high-risk lesions that were identified in the index biopsy but did not result in a malignant disease diagnosis by subsequent surgery or followup.
Reviewer #6	05. Results	The number of cores and size of needle affect complication (hematoma) rates. An analysis of these two factors relative to accuracy would be helpful in informing radiologists as to how many cores and needle size are optimal. The feeling among practitioners doing biopsies (and among manufacturers of equipment) appears to be "more is better". Putting a number on the sample required for accurate diagnosis may help dissipate this.	We agree that the number of retrieved cores and the size of the needle are important factors that may affect the test performance of breast biopsy. For this reason we have carefully considered within-study evidence that these factors modify test performance for various biopsy methods. However, because information on these factors is not well reported and because (especially for the number of cores) cross-study analyses may be susceptible to ecological bias, we have been unable to use meta-regression methods to evaluate them across studies.

Commentator & Affiliation	Section	Comment	Response
Reviewer #6	05. Results	Are the characteristics of the studies clearly described? A graphical representation of the studies included would be useful; e.g. a histogram of study size with clustered bars showing reference standard (surgical biopsy or clinical followup), to give the reader a bird's eye view of the data included in the analysis. This could be repeated for all questions since the studies included differ between them. The main message is that there are no important differences between the different core biopsy methods. This is clearly conveyed.	We have provided graphs of study estimates in the ROC space, grouping studies by index biopsy method. We have not used forest plots because the number of included studies is too large to allow effective presentation.
Reviewer #6	05. Results	Did the investigators overlook any studies that ought to have been included or conversely did they include studies that ought to have been excluded? The selection of studies was appropriate, given the caveats about selection criteria discussed above.	Thank you.
Reviewer #6	05. Results	The recurrence rate data on page 74/237 are not meaningful without information as to what treatment was used. The last study cited (252) appears to describe false negative findings rather than recurrence.	We have revised that section of the report and provided additional information on the included study.
Reviewer #6	06. Discussion/Conclusion	Discussion/conclusions the conclusion that vacuum assisted stereo core is less specific is driven by the fact that indeterminate lesions are called false positive. Most of these lesions are detected on mammogram; although not explicitly stated, these biopsies are probably also performed more recently, since digital mammography (with its greater sensitivity for detection of calcifications) became widely available, with a parallel increase in vacuum assisted procedures. There is an active debate in the literature as to the need for follow-up surgical biopsy for indeterminate lesions, with multiple studies addressing ALH, LCIS, papilloma, radial scar, mucocele-like lesion, flat epithelial atypia. I would suggest breaking out these studies separately and analyzing the "false positive" rate for these lesions specifically, rather than lumping them in with the major analysis. The sensitivity of various biopsy techniques for these lesions is presented as the "upgrade" rate for high risk lesions and DCIS in Table 3, page 48/237, and is surprisingly high and varied.	We have performed additional sensitivity analyses to the definition of false positives, as suggested.
Reviewer #6	06. Discussion/Conclusion	The implications of the major findings (no difference between various CNB methods) are not clear, given the clearly stated limitations of poor data. Again, if the intent was to guide practitioners towards more accurate and more tolerable biopsy techniques, this goal can be achieved only in the variations within a guidance method (eg automated vs vacuum assisted stereo, or automated vs vacuum assisted US core).	We have restructured the assessment of the strength of evidence regarding comparisons among biopsy methods to address this concern.

Commentator & Affiliation	Section	Comment	Response
Reviewer #6	06. Discussion/Conclusion	No important literature was omitted.	Thank you. No further response required.
Reviewer #6	06. Discussion/Conclusion	The future research section is reasonable, but ignores the issues of biopsy threshold and yield.	We agree with the reviewer that the choice of patient population (which is determined by the threshold for performing biopsy and affects yield) is a critical aspect in the design of future studies. We have noted this in the Future Research Needs section of the revised report.
Reviewer #6	07. Figures	Are figures, tables and appendices adequate and descriptive? Figure 5 on page 54/237 should have a label embedded in each plot that conveys the AUC with confidence interval. Table 7 and other similar tables should include study size. The amount of narrative detail in this and other similar tables throughout the report is too high. An attempt to distill this into 1 or 2 major points for each study would make it easier to digest. If that is not possible, I am not sure this and similar tables add anything of value. The authors point out repeatedly that the heterogeneity of the studies, the reporting the endpoints, the populations etc raise significant concerns about bias…. This table and others like it convey the same thing. I gained little from it in terms of understanding the report. Line plots: The key for lines must be in the figure, not in the legend. At present, it is difficult to relate verbal description of dots and dashes to lines in the figure.	Regarding the Strength of Evidence tables, we have opted to retain the current level of detail to ensure that our judgments are transparent to readers. We believe that the AUC for the summary (meta-analytic) ROC curve is not clinically meaningful metric and have not reported it. We have experimented with the suggestions regarding graphical presentation. However, we found that they lead to “busy” and hard to read plots and for this reason we did not retain them.

Commentator & Affiliation	Section	Comment	Response
Reviewer #6	09. Clarity/ Usability	Can the conclusions be used to inform policy and/or practice decisions? The conclusions are 1) there is no major difference between CNB procedures and 2) the existing data are generally poor. However, the policy and practice decisions that need improvement do not relate to which type of core biopsy procedure is used (see comments on page 1). The most important decisions relate to how breasts are surveyed (specific imaging methods, and the added value when more than one method is used), and what is the diagnostic yield of biopsy procedures. Although most of the future research recommendations are reasonable, they ignore the major problem American women face today, namely the performance of an excessive number of biopsies. The diagnostic yield in many centers is lower than the 0,25 assumed by the authors, and probably does differ between imaging methods. Thus the major relevant policy decision cannot be illuminated by this report, despite its very competent execution.	We agree that screening and diagnostic methods merit further research (e.g., use of MRI in these settings). However, we believe that they fall outside the scope of the current report. Whether the performance of biopsy procedures is justified (i.e., whether there is an "excess number of biopsies") is also out of the scope of the current report. We believe that this is a complex question, the answer to which would require consideration of the baseline rate of disease, the natural history of undiagnosed breast cancer, the effectiveness of treatments (at various stages of disease), and patient and physician preferences. It is likely that answering such questions requires extensive modeling efforts, given that no single study can comprehensively address all relevant issues. In summary, we agree that this is valuable area for future research, but it encompasses topics that fall outside the relatively narrow scope of the current report (which was limited to alternative diagnostic biopsy methods).

Commentator & Affiliation	Section	Comment	Response
Reviewer #7	01. General	The review would be more clinically relevant if key questions 1 and 3 considered palpable and non-palpable lesions separately. Understandably it may be difficult to discern what is palpable and what is not palpable particularly in the case of needle localized open biopsy procedures where the lesion may not preoperatively palpable but the lesion can be palpated as it is being excised intraoperatively. However some sort of rubric may be applied for the purposes of review i.e. calling all preoperative non-palpable lesions as nonpalpable regardless of intraoperative findings. While differences in palpable versus nonpalpable didn't come us as a factor affecting test performance, the findings (however limited) should still be reported. There are two reasons that this is important: 1) Patients with palpable breast lesions undergo a different work-up than patients with nonpalpable lesions. Patients with palpable lesions--in the interest of timely results by having the procedure done right in the surgeon's clinic--often undergo free-handed core bx or ultrasound guided bx. Alternatively, a patient with a non-palpable lesion would not be a candidate for a free-handed core bx or an ultrasound guided bx. This difference patient work-up based on palpable/nonpalpable nature of the lesion could cause different trends to be observed with respect to sensitivity/specificity for the methods assessed in this review.2) The standard of care OUS is likely different than the standard of care within the US where there is more screening, earlier detection and greater access to stereotactic techniques to assess nonpalpable lesions. By stratifying the results for nonpalpable lesions separate from palpable lesions, this review will better inform decisions related to the more frequent nonpalpable breast cancer clinical scenario encountered within the US.	Thank you for this suggestion. We have performed subgroup analyses by lesion palpability, when permitted by the information reported in the individual studies.
Reviewer #7	03. Introduction	No suggested changes.	No response required.
Reviewer #7	04. Methods	No suggested changes.	Thank you. No further response required.
Reviewer #7	05. Results	Breaking out results by non-palpable and palpable lesions.	We have performed subgroup analyses by lesion palpability.
Reviewer #7	06. Discussion/Conclusion	May consider suggesting more standard nomenclature for reporting study results associated with bx techniques - if an issue along these lines was encountered. More specifically may suggest reporting on palpable versus nonpalpable nature of breast lesions studied.	The importance of examining lesion-level factors is explicitly mentioned in the Future Research Needs section of the revised report. Please note that we were able to perform some exploratory meta-regression analyses by lesion palpability.
Reviewer #7	09. Clarity/ Usability	Yes - please see comments above regarding stratification of results into palpable/nonpalpable encountered within the US.	We have performed additional analyses stratified by lesion palpability.

Commentator & Affiliation	Section	Comment	Response
TEP #1	01. General	A major weakness of the review is lack of comparative performance by lesion size. Since open surgical biopsy is often for much larger lesions than those targeted by CNB, one would expect much better performance for open biopsy due to the size of lesion and not method of biopsy. This may be part of the issue with MRI performance. I think this issue warrants specific comment in discussion limitation section.	We have mentioned this limitation in the Discussion section of the revised report. I'm assuming that we didn't report on lesion size because the data was not available. If so, might as well say so.
TEP #1	02. Executive Summary	ES-1 para 1. Last sentence. Per cited ref 2, 4.8-7.0% of women screened biennially or annually per decade would receive a FP biopsy recommendation. This is not a "large proportion" as stated. This statement is also repeated in Background p1, para 2. Please correct.	We have edited the text in both places to reflect that the proportion is small.
TEP #1	02. Executive Summary	The authors should report the frequency of biopsies performed for non-screened women due to clinical concerns, not just for screened women as key question 1 relates to palpable and non-palpable abnormalities.	We could not find studies reporting a representative estimate for the proportion of breast biopsies that are indicated on the basis of clinical abnormalities. However, multiple sources point to mammographically-detected abnormalities being the most common indication. This information has been added in the Introduction.
TEP #1	04. Methods	See general comment.	No response required.
TEP #1	05. Results	Pg. 36 Dissemination of Cancerous Cells No report of studies evaluating open biopsy results to compare to CNB.	All studies meeting our criteria that reported in the displacement of cancerous cells have been included under Key Question 2.
TEP #1	06. Discussion/Conclusion	Yes, see general comments above. Ok.	Thank you. No further response required.
TEP #1	09. Clarity/ Usability	Yes. Yes. Yes.	Thank you. No further response required.
TEP #2	01. General	In general the key questions are nicely outlined and addressed throughout the text	Thank you. No further response required.
TEP #2	03. Introduction	In the abstract, it states that "open biopsy continues to be considered the gold standard diagnostic procedure." in lines 34-45. However, clinically, open biopsy is not the gold standard anymore. It may be more accurate to state, "open biopsy continues to be considered the reference standard for examining diagnostic procedures."	We have removed this statement.
TEP #2	04. Methods	Yes. [*EPC note: this is an answer to a question soliciting the reviewers opinion regarding the adequacy of the Report's methods*]	Thank you. No response required.

Commentator & Affiliation	Section	Comment	Response
TEP #2	05. Results	On page 20, in the "Contextualizing the results of test performance meta-analysis" section, I had a difficult time understanding the comparison of false positive rates (Figure A) between different biopsy modalities. Women undergo different types of core biopsies based on different presentations. If false positive refers to a pathologically positive core biopsy (with either carcinoma or a high risk lesion) and then a negative surgical excision or follow-up mammogram, it may be that the initial biopsy removed all of the abnormal tissue. It seems to me that a small cluster of calcifications sampled with a stereotactic biopsy is more likely to be completely removed than a mass that undergoes US guided biopsy. Clinically this is not treated as a false positive as the results of the core biopsy are still used for treatment and risk assessment. I think including the false positive definition here would be useful to help make this clinically relevant.	In cases where the index biopsy resulted in the removal of the entire lesion (making reference standard assessment impossible), we made the assumption that the index biopsy resulted in a "true" result (positive or negative, depending on the pathological results). This assumption has been identified as reasonable by several reviewers and was also adopted in the original 2009 evidence review.
TEP #2	06. Discussion/Conclusion	Yes.	Thank you. No further response required.
TEP #2	07. Figures	On page 53, the Figure number is not present and the figure itself is covering up the descriptive text below.	We have corrected this typesetting error. Thank you.
TEP #2	09. Clarity/ Usability	Text is easy to follow.	Thank you. No further response required.
TEP #3	01. General	The lack of published studies on performance of open surgical biopsies is worrisome because this was considered the gold standard in this comparative effectiveness review. More comment on this issue is needed to help readers understand how this weakness might influence the interpretation of literature included in the meta-analytic methods. As written, this is more of a comparative effectiveness assessment of different types of needle biopsies rather than comparison of needle to open surgical biopsy.	We have added a discussion of this issue in the revised report. Please note that this report was intended to assess both the comparative effectiveness of open versus core needle biopsy and the comparative effectiveness of alternative core needle biopsy methods.

Commentator & Affiliation	Section	Comment	Response
TEP #3	01. General	The adverse event reporting focuses on complications associated with different biopsy approaches and do not include any information on over or under diagnosis. This needs justification and inclusion as an area for future research. This topic is driving decisions about breast cancer screening, which directly affects breast cancer discovery, subsequent treatment and patient outcomes. It is too directly related to the diagnosis of breast cancer to be ignored. It seems adverse events (Question 2 could have included information on this important topic.	Breast cancer diagnosis is a complex process that includes the clinical or radiologic suspicion of the presence of disease, followed by additional diagnostic procedures, final diagnosis, and treatment. However, our report pertains only to the performance and adverse events of biopsy methods applied to women after an initial suspicious lesion has been detected. Because over-diagnosis is the end result of lead time (due to screening) and competing risks (specifically, the force of mortality due to causes of death other than breast cancer), we do not think that a discussion of over diagnosis is within the scope of the current report.
TEP #3	03. Introduction	In the abstract: I am confused by the statement that a single investigator abstracted data from each study with quantitative results and intervention descriptions verified by a second reviewer. Does this mean that a second reviewer only reviewed intervention articles with quantitative results?	Our statement was intended to indicate that a second reviewer verified both quantitative and nonquantitative data. We have clarified the statement to indicate that a second reviewer verified all extracted data.
TEP #3	03. Introduction	Also, The opening statement in the Abstract says this review was conducted because of additional studies, but the results indicate that no new studies investigating its test performance. This is confusing. Does this mean the 151 studies found were the same studies the 2009 review were based on? The first sentence indicates that as of 2009, an estimated 2.7 million women had a current or past diagnosis of breast cancer; however, the citation for this statement indicates data capture extended into 2012, which is confusing. The time period for these statistics should be checked for accuracy.	We have clarified the number of new studies identified for each key question.. Overall, we found more that 100 new studies. We have updated the prevalence data in the Background section to 2012. This information is from the American Cancer Society Report 2013-2014. Although the title of the report is 2013-2014, the prevalence statistics in the report only extend to 2012.

Commentator & Affiliation	Section	Comment	Response
TEP #3	03. Introduction	Given the extensive debate about over diagnosis of breast cancer, I was surprised to see nothing presented on this topic. It seems the difference in sensitivity and specificity might be contributing to over diagnosis, but this discussion is absent. Can this be justified?	We agree that breast cancer diagnosis is a complex process that includes the clinical or radiologic suspicion of the presence of disease, followed by additional diagnostic procedures, final diagnosis, and treatment. However, our report pertains only to the performance and adverse events of biopsy methods applied to women after an initial suspicious lesion has been detected. Because over diagnosis is the end result of lead time (due to screening) and competing risk (specifically, the force of mortality due to causes of death other than breast cancer), we do not think that a discussion of over diagnosis is relevant to the scope of the report.
TEP #3	04. Methods	I was surprised to see that the expert panel did not appear to include a breast pathologist. What was the rationale for this?	The composition of the Technical Expert panel followed the composition of the Panel convened for the original ECRI evidence report (2009). The peer review panel included experts in Clinical Pathology.
TEP #3	04. Methods	I was surprised to see that the eligibility criteria included enrollment of 10 or more patients with at least 50 percent of them followed to completion of the study. Why was this a priori follow-up rate so low?	We use minimum followup rate of 50% in order to provide a comprehensive account of the available studies. Differences in followup duration were addressed by appropriate subgroup analyses. Unfortunately, the poor reporting of followup information did not allow us to perform analyses stratified by completeness of followup.
TEP #3	04. Methods	I did not see any quality control step for checking to determine that no eligible study was inadvertently excluded from abstraction – it appears a single reviewer conducted this step without quality control for a small percent of articles. Can this be clarified?	In the third paragraph of the Literature Search and Abstract Screening section we state that each set of citations was independently screened by two reviewers. Of note, during a typical evidence review several full text articles are reviewed in team meetings (attended by all investigators). Furthermore, for this particular review a senior investigator reviewed a purposive sample of more than 30% of the included studies.
TEP #3	04. Methods	Methods described under data synthesis are quite strong – nicely done!!	Thank you. No further response required.
TEP #3	04. Methods	The expanded focus on women at higher baseline risk of breast cancer is a nice addition to this updated review. Overall, I agree with the assessment of the evidence as presented in this rather limited review.	Thank you. No further response required.

Commentator & Affiliation	Section	Comment	Response
TEP #3	05. Results	The results presented, even though they are limited are clearly described and applicable. I valued all the figures and tables presenting information.	Thank you. No further response required.
TEP #3	06. Discussion/Conclusion	The literature on the availability of a qualified pathologists is worrisome, as only two studies are reported and they conflict with each other. A great deal of literature exists on pathologists agreement on a final diagnosis with the biggest tension occurring between atypical ductal hyperplasia and ductal carcinoma in situ – There is less concern about making a diagnosis of invasive breast cancer, however both DCIS and invasive breast cancers are recommended for treatment, which subsequently affects patient outcomes. Again, the lack of attention to this important issue is a concern.	Thank you for this comment. We believe that inter-rater agreement (e.g., agreement among pathologists examining the same specimens) is not the same as "availability of qualified pathologists" (a health services outcome). As such, we considered studies reporting on the former outcome as out of the scope of the current review. The Dicussion section addresses some issues related to the reporting of pathology results.
TEP #3	09. Clarity/ Usability	The information presented is well structured and organized with the main points being clearly presented. I am not sure the evidence for all the reasons I describe are complete enough to deal with over diagnosis.	Breast cancer diagnosis is a complex process that includes the clinical or radiologic suspicion of the presence of disease, followed by additional diagnostic procedures, final diagnosis, and treatment. However, our report pertains only to the performance and adverse events of biopsy methods applied to women after an initial suspicious lesion has been detected. Because over-diagnosis is the end result of lead time (due to screening) and competing risks (specifically, the force of mortality due to causes of death other than breast cancer), we do not think that a discussion of over diagnosis is within the scope of the current report.
TEP #4	01. General	A great deal of effort seems to have been spend on this "update" with virtually no new information beyond what was presented by the report in 2009. Perhaps our resources could have been better spent.	This update added more than 128 new studies to the original evidence report. The Brown EPC does not make determinations about optimal use of resources. However we hope that integrating evidence from a total of more than 300 studies (including 54, 70, and 59 new studies for Key Questions 1,2, and 3 respectively) and performing a comprehensive assessment of test performance and comparative test performance (for the first time in this update) will be useful to readers.

Commentator & Affiliation	Section	Comment	Response
TEP #4	01. General	The report does not provide any information that is clinically useful. It is very hard to read and decipher. In some instances it is obvious that the authors do not a grasp of the clinical setting under which these biopsies are done (referring to average risk vs high risk women, comparing the guidance methods with each other, etc).	We have clarified the operational definitions of "average" and "high" risk women. Please note that we initially adopted the terminology from the original evidence report. We have also restructured the reporting and interpretation of comparative analyses.
TEP #4	01. General	Is reference 29 accurate? I don't understand how it is relevant.	We have dropped this reference.
TEP #4	03. Introduction	No comments.	No response required.
TEP #4	04. Methods	The statement that most of the studies reported on women at average risk of breast cancer is not accurate. With the exception of the studies of MRI guided biopsy, which is generally done in women at increased risk for breast cancer, most studies of needle breast biopsy include women at all risk levels, and stratification by risk level is not reported nor is it clinically relevant. The type of needle biopsy is not determined by the risk level but rather by how the lesion is best seen.	We agree with the reviewer that the level of breast cancer risk is often not clearly reported in individual studies. We have added this information in the Methods section ("Study Eligibility").
TEP #4	05. Results	Comparing accuracy of biopsy by guidance method (Tables D and E, page 19) is not clinically relevant because guidance method is determined by which modality allows best visualization of the lesion, with ultrasound preferred over stereotactic biopsy if visible by both mammography and ultrasound because of greater ease of performance.	We have limited comparative analyses to studies using the same imaging method to guide biopsy. For example, we have compared automated vs. vacuum-assisted biopsy using ultrasound guidance and automated vs. vacuum-assisted biopsy using stereotactic guidance. The tables cited by the reviewers have been modified accordingly.
TEP #4	06. Discussion/Conclusion	The studies included in this report that were not included in the 2009 version are not easily found in the reference list. It would be good to specifically reference them.	We have referenced the two groups of studies separately at the beginning of the results section for each Key Question and in the summary table at the beginning of the Results section.
TEP #4	08. References	Is reference 29 accurate? I don't understand how it is relevant.	Thank you for pointing this out. The reference has been corrected.
TEP #4	09. Clarity/ Usability	The report does not provide any information that is clinically useful. It is very hard to read and decipher. In some instances it is obvious that the authors do not have a grasp of the clinical setting under which these biopsies are done (referring to average risk vs high risk women, comparing the guidance methods with each other, etc).	We believe that the changes we have made to the report address the reviewer's concerns.
TEP #5	01. General	The report is well researched, logical and informative. Where data exists, it has been evaluated fairly and reported accurately.	Thank you. No further response required.

Commentator & Affiliation	Section	Comment	Response
TEP #5	01. General	Not sure if there is data to support two key subquestions: 1) Since most biopsies using freehand guidance are of palpable masses, it might be better to subdivide the analysis to look at results divided into palpable lesions vs non-palpable lesions (for both the free hand and image guided biopsy techniques...particularly for US guidance). It's somewhat unfair to compare freehand, which usually is used for lesions 2cm and up (palpable) vs smaller non-palpable masses which are done by imaging guidance. There is no place for freehand biopsy of non-palpable lesions.	We have performed additional subgroup analyses by lesion palpability.
TEP #5	01. General	2) Is there data on surgeon vs radiologist outcomes? Both have accreditation programs and some have more training and experience than others. What evidence comes from the radiology literature (done by radiologists) vs what comes from the surgical literature done by surgeons? Important: There is no reason to assume that the same results may be accomplished by both groups. When comparing, it would be important to include case mix considerations. Radiologists probably biopsy smaller lesions (probably more image detected lesions but also some palpable), than surgeons. It would be expected that the accuracy would be higher for the larger lesions, so that would need to be a factor in analysis. Is it possible to incorporate this into figures and Tables? Maybe there is not enough data in the existing literature. If it currently exists, this report could dig deeper and answer harder questions than just image guided bx vs surgical biopsy. If this data does not currently exist in the published literature examined, both of these are important areas for future research.	Unfortunately, this information is not fully reported in the included studies. We have identified it as a potential area for further research in the Future Research Needs section of the revised report. Information on lesion size and palpability was analyzed to the extent permitted by the reported data.
TEP #5	03. Introduction	No comments.	No response required.
TEP #5	04. Methods	Would like to see the literature broken down by palpable vs nonpalpable lesions if possible.	We extracted information on whether lesions were palpable on non-palpable, although this information was often unclearly reported. We performed a subgroup analysis by stratifying studies into three categories: “>80% lesions palpable”; “>80% lesions non-palpable”; “mixed”; and “not reported” and have added these results to the report.

Commentator & Affiliation	Section	Comment	Response
TEP #5	04. Methods	Adverse events (ES p12, p31 (p62 of 237): Line 53, p 32: "diagnosed with skin ecchymosis". Comment: It is commonplace to have ecchymosis and even a small hematoma and it is not considered a complication, and it does not mean there was excessive bleeding. The discussion doesn't handle the terms bleeding, hematoma and bleeding events that required treatment consistently. I think it needs careful definition of terms and consistent use of those terms. There is also no mention of bleeding complications related to special circumstances: anti-coagulated patients, patients on aspirin, Plavix, etc. Some practices remove patients from such drugs before a biopsy is done, to limit potential problems.	Thank you for raising this important point. We have standardized terminology related to ecchymosis, bleeding, and hematomas in the Discussion section of the revised report. During data extraction we relied on the outcome definitions provided by individual studies. We have extracted additional data regarding the antiplatelet/anticoagulant status of patients at the time of biopsy (when available). This information has been added to the revised report.
TEP #5	05. Results	p 15 (46 of 237) and also mentioned in abstract and ES, the test performance of open surgical biopsy is discussed. While the gold standard, there is likely a difference between palpable and non-palpable lesions (size matters!). For palpable lesions, the miss rate by surgeons is probably minimal. For non-palpable lesions, imaging is used to guide the excision in one of the following ways: a) pre-operative localization using imaging guidance (usually US or mammographic placement of a hook wire system by the radiologist, with subsequent excision by the surgeon; b) intra-operative ultrasound by the surgeon; c) pre-operative radioactive seed placement by the radiologist with use of detector in the operating room by the surgeon. While this detailed discussion may not belong in this report, it is brought up to mention that excision is not the gold standard that one presumes. There are failures in the system that lead to inadequate excision of non-palpable lesions, and that comment is lacking in the report.	We have incorporated some of these ideas to the Introduction and Discussion section of the report.
TEP #5	05. Results	See comments regarding palpable v non-palpable evaluated separately in the General Comments section.	We have collected additional data and performed additional analyses by lesion palpability.

Commentator & Affiliation	Section	Comment	Response
TEP #5	06. Discussion/Conclusion	The implications of major findings are clearly stated. Future research could include1) area noted above related to breakdown by palpable and non-palpable lesions 2) area noted above related to who specifically performs the biopsy. Radiologist vs surgeon, accredited by either ACR or ASBS or not accredited at all. Do these correlate with better outcomes? 3) New "dense" notification laws have passed in at least 13 states, and others are being considered, as is a national law. These inform women that other methods of screening might be warranted, most notably ultrasound screening of average risk women. (High risk women would be recommended to have screening with contrast enhanced breast MRI). Outcomes have shown that the PPV2/3 (recommended and biopsied lesions) is much lower for ultrasound screening than for mammographically detected lesions. Big area for research! Should there be a tracking system for further workups that are done as the result of such legislation? What are the oucomes of biopsies due to supplemental screening due to the enactment of these laws? What is the stress that is caused by supplemental screening? The laws suggest that women discuss risk, dense breasts and supplemental screening with their physicians. All of these things create very important public health implications!! 4) Recently coding for interventional procedures of the breast have been bundled (January 1, 2014). Bundling has economic implications. Will safety- net institutions take a bigger economic hit than private offices? Will there be reduced access to image guided core biopsies and if so, where? Important public health consideration given the conclusions of this report.	We have performed additional subgroup analyses by lesion palpability. However, we agree with the reviewer that more research is needed in this area. We agree that the research areas identified by the reviewer are interesting from a health care delivery standpoint. We have adopted some of the proposed research recommendations (those most closely related to the evidence base we reviewed) in the revised evidence report.
TEP #5	09. Clarity/ Usability	Yes, well done. Clearly supports image guided biopsy for diagnosis as the standard of care, an important advance.	Thank you. No further response required.
TEP #6	01. General	Very thorough review. There is really nothing new however that will be of interest to radiologists.	Thank you. We agree that in clinical practice the question about the relative merits of open versus core needle biopsy is considered settled. However, we believe that the evaluation of alternative core needle biopsy methods (a key aspect of this update) is a valuable contribution. Furthermore, new biopsy methods (e.g., MRI-guided biopsy) appear to represent areas for additional research.
TEP #6	03. Introduction	page 11, line 17: "large proportion need biopsies" is really not a fair statement. Of all women, even those recalled, only a small percentage need biopsies.	We have provided the range of proportions from the study cited.

Commentator & Affiliation	Section	Comment	Response
TEP #6	04. Methods	False positives: multiple locations, e.g. page 21, 57. I question why a high-risk lesion that is benign on excision is considered a "false positive" as it is not "positive". They are more benign than malignant. A true high risk false positive would mean that the patient underwent a cancer surgery for the lesion (i.e. lumpectomy and sentinel node biopsy) and this is never the case. Surgical excisions only are performed for high risk lesions.	In the revised report we have provided additional details about the definition of each possible diagnostic test result. Please note that when the biopsy removed an entire lesion, we considered the biopsy results as true (TP or TN, depending on the pathology results). We have also performed additional sensitivity analyses to evaluate alternative definitions of FP findings.
TEP #6	05. Results	Page 15, line 37. "No studies provided information on underestimation rates for open surgical biopsy." There is really no such thing as underestimation on surgical biopsy as surgical biopsies are what define under or over estimation. High-risk lesions are treated with surgical excision, so there would be no need to re-excise.	We have revised the text to clearly note that open surgical biopsy is not expected to be associated with underestimation.
TEP #6	06. Discussion/Conclusion	Yes.	Thank you. No further response required.
TEP #6	09. Clarity/ Usability	Yes.	Thank you. No further response required.
TEP #7	01. General	The key questions are explicitly stated and the targeted audience is explicitly defined. Though the report is very well done, it unfortunately falls behind current clinical management. How the studies are designed and their limitations are related more to answering questions that are more current than the ones being asked.	Thank you for these comments. We agree that in clinical practice the question about the relative merits of open versus core needle biopsy is considered settled. However, we believe that the evaluation of alternative core needle biopsy methods (a key aspect of this update) is a valuable contribution. Furthermore, new biopsy methods (e.g., MRI-guided biopsy) appear to represent areas for additional research.
TEP #7	03. Introduction	No comment.	No response required.

Commentator & Affiliation	Section	Comment	Response
TEP #7	04. Methods	The lack of studies addressing open surgical biopsy are due to the inclusion/exclusion criteria. If the start date had included the 1980s, there would have been numerous studies addressing open surgical biopsy sensitivity, specificity and adverse events. I also know there are more than 4 MRI guided biopsy studies.	Thank you for this important comment. The current report updates the original 2009 report by the ECRI EPC. As such we did not extend the literature searches to years earlier than those covered by the original report. Throughout the report, we have identified open surgical biopsy as a well-established technique with (near) perfect test performance. Regarding MRI, we have reviewed several dozen studies of MRI guided biopsy that did not meet out inclusion criteria. We also contacted all TEP members and peer reviewers to ask for specific citations that we should consider. No studies meeting our criteria were identified through this process.
TEP #7	05. Results	The amount of detail is appropriate and the key messages are explicit.	Thank you. No further response required.
TEP #7	05. Results	I have some doubt as to how applicability of the key messages. Based on consensus statements of major medical societies as well as the Standards for the National Accreditation Program of Breast Centers, Image-guided biopsy is recommended as the procedure of choice for the diagnosis of breast lesions. In fact, centers cannot get NAPBC accreditation without showing that IGBB is utilized in breast cancer diagnosis for the great majority of patients.	We believe that these recommendations are consistent with our analyses, which demonstrate that imaging-guided biopsy methods have superior test performance compared to free-hand methods. Please note that our statements about applicability refer to whether the populations, tests/interventions, comparators, and outcomes considered in research studies are “comparable” with those in clinical practice. This is required for study results to be transferable to real-world care.
TEP #7	06. Discussion/Conclusion	The findings are clearly stated. The limitations are adequately described. I have some concern about the levels of bias and strength of conclusions related to study design. Once again the studies were designed (especially the newer ones for the current review) with IGBB as an accepted modality. The patient's risk of breast cancer is not usually included because it is the lesion that is indeterminate, regardless of patient risk. With regard to the potential selective reporting of adverse events, it has become common knowledge that IGBB will result in bruising (greater for VA devices) and small localized hematomas, therefore most of these studies are looking for adverse events that cause significant pain for patients or further intervention.	Please note that the reporting of information on “baseline risk of cancer” was not used as a criterion to assess risk of bias. Regarding adverse events, we found that all adverse events (e.g., infectious complications, bleeding, etc.) were poorly reported. Similarly, the issues identified by the reviewer (clinically significant adverse events and adverse events requiring intervention) were poorly reported in the available studies.

Commentator & Affiliation	Section	Comment	Response
TEP #7	06. Discussion/Conclusion	With regard to suggestions for future studies, it is very unlikely that these studies will occur because of the current state of IGBB. Perhaps a study to try to limit the number of patients requiring an IGBB based on the patient's overall risk of breast cancer could be of interest in light of medical economics. One thing to understand, is that comparing types of image guidance is related to the type of lesion. For example stereotactic breast biopsy is mainly for microcalcifications and lesions not seen with ultrasound. Ultrasound cannot see microcalcifications. With all things equal, ultrasound guidance would be preferred if the lesion is seen with both modalities because of patient comfort, ease of accessibility. MRI biopsy is reserved only for lesions that cannot be seen with either mammogram or ultrasound because of equipment costs and patient discomfort. therefore comparison studies with MRI guided needle core bx will not be performed.	Thank you for this comment. We have adopted some of the reviewer's ideas regarding comparisons across alternative imaging methods in the revised report.
TEP #7	09. Clarity/ Usability	The report is very well structured and organized. The main points are clearly presented.	Thank you. No further response required.
TEP #7	09. Clarity/ Usability	I believe that the conclusions are unfortunately outdated but they do confirm what is already being done. I would suggest that a valuable next report would be to evaluate studies that directly compare typed of devices and stereo vs ultrasound with regard to costs. Policy decisions that can help guide payors in appropriate coverage for these procedures would be very timely.	Cost outcomes have been considered, provided they were reported by studies meeting our criteria.
TEP #8	01. General	The main point I would suggest forwarding to the authors has to do with the unfortunate lack of agreement among pathologists in the diagnosis of these biopsy specimens. Previous studies have shown less than 50% agreement among pathologists in the diagnosis of atypia and disagreement on >10% of cases of DCIS. While policy is usually in place at laboratories that require pathologists to obtain a second opinion on all cases of newly diagnosed invasive breast cancer, there are no standard policies in the U.S. about DCIS or atypia.	We agree that the reliability and validity of pathologic diagnosis is important. However, this aspect of the biopsy process was not considered to be within the scope of the current report. Nonetheless, we have provided some related suggestions in the Discussion section of the revised report.