



Effective Health Care Program

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

Executive Summary

Background

Approximately one in eight U.S. women will develop breast cancer during her lifetime.¹ Because the earliest stages of breast cancer are asymptomatic, the process of breast cancer diagnosis is often initiated by detecting an abnormality through self-examination, physical examination by a clinician, or screening mammography. If the initial assessment suggests that the abnormality could be breast cancer, the woman is likely to be referred for a biopsy—a sampling of cells or tissue from the suspicious lesion. Among women screened annually for 10 years, approximately 50 percent will need additional imaging, and 5–7 percent will have biopsies.^{2,3}

Three techniques for obtaining samples from suspicious breast lesions are available: fine-needle aspiration, biopsy with a hollow core needle, or open surgical retrieval of tissue. Fine-needle aspiration samples cells and does not assess tissue architecture, is generally considered less sensitive than core needle and open biopsy methods,⁴ and is used less frequently. Core-needle biopsy, which retrieves a sample of tissue, and open surgical procedures are the most frequently used biopsy methods. Lesion samples obtained by core needle or surgical biopsy are

Effective Health Care Program

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

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

evaluated by pathologists and classified into histological categories with the primary goal of determining whether the lesion is benign or malignant. Because core needle biopsy samples only part



Agency for Healthcare Research and Quality

Advancing Excellence in Health Care • www.ahrq.gov

Effective
Health Care

of the breast abnormality, a risk exists that a lesion will be classified as benign, high risk, or noninvasive when invasive cancer is in fact present in unsampled areas. Open surgical biopsy samples most or all of the lesion, and is therefore considered to have a smaller risk of misdiagnosis. However, open procedures may carry a higher risk of complications, such as bleeding or infection, compared to core needle biopsy procedures.⁵ Therefore, if core needle biopsy is also highly accurate, women and their clinicians may prefer some type of core needle biopsy to open surgical biopsy.

Alternative core needle biopsy methods differ with respect to the use of imaging (e.g., stereotactic mammography; ultrasound; or magnetic resonance imaging [MRI]), the use of vacuum to assist in tissue acquisition, the use of needles of varying diameter, and the numbers of samples taken. These and other factors may affect test performance and the rate of complications. For example, some biopsy procedures may retrieve larger amounts of tissue, improving test performance, but the retrieval of larger amounts of tissue may also result in more complications, such as bleeding. Imaging methods may also influence the performance of open surgical biopsies because the majority of such biopsies are preceded by an image-guided wire localization procedure. In general, the impact of various aspects of biopsy technique and patient or lesion characteristics on test performance and safety is not clear.

In 2009, the ECRI Evidence-based Practice Center (EPC) conducted a comparative effectiveness review for core needle versus open surgical biopsy commissioned by the Agency for Health Care Research and Quality (AHRQ).^{6,7} That evidence report assessed the diagnostic test performance and adverse events of core needle biopsy techniques compared to open surgical biopsy and evaluated differences between open biopsy and core needle biopsy with regards to patient preferences, costs, availability, and other factors. The authors concluded that core needle biopsies were almost as accurate as open surgical biopsies, had a lower risk of severe complications, and were associated with fewer subsequent surgical procedures.⁷

The publication of additional studies and changes in practice raised the concern that the conclusions of the original report may be out of date, particularly for the underestimation rate of ductal carcinoma in situ (DCIS) with stereotactically guided vacuum-assisted core needle biopsy, the performance of MRI-guided core needle biopsy, and the performance of freehand automated

device core needle technology. New studies may also provide additional information allowing the exploration of heterogeneity for test performance and safety outcomes. Therefore, an updated review of the published literature was considered necessary to synthesize all evidence on currently available methods for core needle and open surgical breast lesion biopsy.

Key Questions

On the basis of input from clinical experts during the development of our protocol, we made minor revisions to the Key Questions and study eligibility criteria to clarify the focus of the updated review. We specified the following three Key Questions to guide the conduct of the update:

Key Question 1: In women with a palpable or nonpalpable breast abnormality, what is the test performance of different types of core needle breast biopsy compared with open biopsy for diagnosis?

- What factors associated with the patient and her breast abnormality influence the test performance of different types of core needle breast biopsy compared with open biopsy for diagnosis of a breast abnormality?
- What factors associated with the procedure itself influence the test performance of different types of core needle breast biopsy compared with open biopsy for diagnosis of a breast abnormality?
- What clinician and facility factors influence the test performance of core needle breast biopsy compared with open biopsy for diagnosis of a breast abnormality?

Key Question 2: In women with a palpable or nonpalpable breast abnormality, what are the adverse events (harms) associated with different types of core needle breast biopsy compared with open biopsy for diagnosis?

- What factors associated with the patient and her breast abnormality influence the adverse events of core needle breast biopsy compared with the open biopsy technique in the diagnosis of a breast abnormality?
- What factors associated with the procedure itself influence the adverse events of core needle breast biopsy compared with the open biopsy technique in the diagnosis of a breast abnormality?
- What clinician and facility factors influence the adverse events of core needle breast biopsy compared with the open biopsy technique in the diagnosis of a breast abnormality?

Key Question 3: How do open biopsy and various core needle techniques differ in terms of patient preference, availability, costs, availability of qualified pathologist interpretations, and other factors that may influence choice of a particular technique?

Methods

We performed a systematic review of the published scientific literature using methodologies outlined in the AHRQ “Methods Guide for Comparative Effectiveness Reviews,”⁸ hereafter referred to as the Methods Guide. We followed the reporting requirements of the “Preferred Reporting Items for Systematic Reviews and Meta-analyses” (PRISMA) statement.⁹ A full description of all review steps is included in the full report and the study protocol (PROSPERO registration number CRD42013005690).

External Stakeholder Input

We convened a nine-member Technical Expert Panel (TEP), including representatives of professional societies, experts in the diagnosis and treatment of breast cancer (including radiologists and surgeons), and a patient representative. The TEP provided input to help further refine the Key Questions and protocol, identify important issues, and define the parameters for the evidence review.

Study Eligibility Criteria

We included only English-language full-text articles. Studies included for the assessment of diagnostic test performance (Key Question 1) met the following inclusion criteria: (1) enrolled women not previously diagnosed with breast cancer who received core needle or open biopsy for initial diagnosis of possible breast cancer; (2) compared diagnoses on core needle biopsy to a reference standard of open surgery or followup by clinical examination or imaging of at least 6 months; (3) reported or allowed the calculation of sensitivity, specificity, positive or negative predictive value; (4) were prospective or retrospective cohort studies (including randomized controlled trials); and (5) enrolled 10 or more patients and followed at least 50 percent of them to the completion of the study. In contrast to the original report, we did not restrict eligibility to studies including only women at average risk for breast cancer, because MRI-guided biopsy, which was identified as a topic of interest for this update, is used mainly in women at a higher-than-average risk for breast cancer. Of note, studies often do not provide information on the risk of cancer among included patients. Thus we grouped studies into two categories: (1) studies that explicitly

reported that more than 15 percent of included patients were at high risk of cancer; (2) studies that reported that fewer than 15 percent of included patients were at high risk of cancer, or did not provide information on baseline risk. Throughout this review, we refer to the latter group as “studies of women at average risk of cancer”; however, we acknowledge that this group may include studies enrolling patients at higher-than-average cancer risk.

Studies included for the assessment of possible adverse events of core needle biopsy (Key Question 2) or the assessment of patient-relevant outcomes, resource use and logistics, and availability of technology and relevant expertise (Key Question 3) were not required to compare diagnoses on core needle biopsy to a reference standard of open surgery or clinical followup, or to contain extractable information on diagnostic test performance. Furthermore, for Key Question 2 we included any primary research articles, regardless of design, that addressed the dissemination or displacement of cancer cells by the biopsy procedure (e.g., seeding).

Literature Search and Study Selection

We searched MEDLINE®, Embase®, the Cochrane Central Register of Controlled Trials, the Cochrane Database of Systematic Reviews, the Database of Abstracts of Reviews of Effects, the Health Technology Assessment Database, the U.K. National Health Service Economic Evaluation Database, the U.S. National Guideline Clearinghouse, and CINAHL. Appendix A in the full report describes our search strategy, which is based on an expansion of the search strategy used in the original report. We did not use a search filter for studies of diagnostic tests to increase search sensitivity.^{10, 11} We also searched for systematic reviews on the topic and used their lists of included studies to validate our search strategy and to make sure we identified all relevant studies.

To identify studies excluded from the original evidence report because they enrolled women at high risk for cancer, we rescreened both the set of abstracts screened for the original report and the full text of studies excluded from the original report because they included women at high risk for cancer. Titles and abstracts were manually screened in duplicate. A single reviewer screened each potentially eligible article in full text to determine eligibility and a second reviewer examined all articles deemed relevant. Disagreements regarding article eligibility were resolved by consensus involving a third reviewer.

Data Abstraction and Management

Data were extracted using electronic forms and entered into the Systematic Review Data Repository (SRDR; <http://srdr.ahrq.gov/>). We pilot-tested the forms on several studies extracted by multiple team members to ensure consistency in operational definitions. A single reviewer extracted data from each eligible study. A second reviewer verified extracted data and discrepancies were resolved by consensus including a third reviewer. We contacted

authors to clarify information reported in their papers and to verify suspected overlap between study populations in publications from the same group of investigators.

Definitions of Test Performance Outcomes and Underestimation Rates

Table A illustrates how index and reference standard results were used to construct 2×2 tables for Key Question 1 (test performance outcomes).

Table A. Definitions of diagnostic groups based on index and reference standard test results

		Reference Standard Results (open surgery or followup)	
		Malignant (invasive or in situ)	Benign
Core Needle Biopsy Results (index test)	Malignant (invasive or in situ)	considered TP	considered TP*
	High risk lesion (e.g., ADH)	considered TP	considered FP
	Benign	considered FN	considered TN

*Some study authors specifically stated that diagnoses of malignancy on core needle biopsy were assumed to be correct, whether or not a tumor was observed upon surgical excision. The original version of this review also classified all diagnoses of malignancy on core needle biopsy as true positives.
ADH = atypical ductal hyperplasia; FN = false negative; FP = false positive; TN = true negative; TP = true positive.

Two issues related to the definition of diagnostic test categories merit additional description. First, occasionally core needle biopsy removes the entire target lesion that is being biopsied, rendering subsequent surgical biopsies unable to confirm the findings of the index test procedure. In such cases of core needle diagnoses of malignancy, we considered the core needle results to be true positive. This operational definition was adopted by several of the primary studies we reviewed and the original ECRI report. Second, in our primary analysis (and consistent with the 2009 ECRI report) core needle biopsy identified high risk lesions that on subsequent surgery (or followup) are not found to be associated with malignant disease were considered false positive. To assess the impact of this operational definition on our findings we performed a sensitivity analysis where high risk lesions on index core needle biopsy found to be nonmalignant (high risk or benign) on subsequent open biopsy or surgical excision were excluded from the analyses.

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.

Assessment of Risk of Bias

We assessed the risk of bias for each individual study following the Methods Guide. We used elements from the Quality Assessment for Diagnostic Accuracy Studies instrument (QUADAS version 2), to assess risk of bias for studies of diagnostic test accuracy.¹²⁻¹⁵ The tool assesses four domains of risk of bias related to patient selection (e.g., consecutive or random selection), index test (e.g., blinding of index test assessors to reference standard results), reference standard test (e.g., blinding of reference standard assessors to the index test results), and patient flow and timing (e.g. differential and partial verification). We used items from the Newcastle-Ottawa scale,¹⁶ the Cochrane Risk of Bias tool,¹⁷ and the checklist proposed by Drummond et al.,¹⁸ to assess nonrandomized cohort studies, randomized controlled trials, and studies of resource utilization and costs, respectively.

Data Synthesis

We summarized included studies qualitatively and presented important features of the study populations, designs, tests used, outcomes, and results in summary tables. Statistical analyses were conducted using methods currently recommend for use in Comparative Effectiveness Reviews of diagnostic tests.^{19, 20}

For Key Question 1 we performed meta-analyses when studies were deemed sufficiently similar with respect to included populations, and the core needle biopsy and reference standard tests they employed.

We used a mixed effects binomial-bivariate normal regression model that accounted for different imaging guidance methods, the use of automated or vacuum-assisted devices, and the baseline of risk of cancer of included patients. This model allowed us to estimate the test performance of alternative diagnostic tests, and to perform indirect comparisons among them.²¹ Furthermore, it allowed us to derive summary receiver operating characteristic (ROC) curves.^{22,23} A univariate mixed effects logistic regression model was used for the meta-analysis of rates of DCIS and high risk lesion underestimation.²⁴ We used meta-regression methods to evaluate the impact of risk of bias items and other study-level characteristics.^{25, 26}

For Key Question 2, we found that adverse events were inconsistently reported across studies and that the methods for ascertaining their occurrence were often not presented in adequate detail. For this reason we refrained from performing meta-analyses for these outcomes. Instead, we calculated descriptive statistics (medians, 25th and 75th percentiles, minimum and maximum values) across all studies and for specific test types.

For Key Question 3, because of the heterogeneity of research designs and outcomes assessed, we were only able to perform a meta-analysis comparing core needle and open surgical biopsies with respect to the number of patients who required one versus more than one surgical procedure for treatment, after the establishment of breast cancer diagnosis. This analysis used a univariate normal random effects model with binomial within-study distribution.

All statistical analyses were performed using Bayesian methods; models were fit using Markov Chain Monte Carlo methods and non-informative prior distributions. Theory and empirical work suggest that, when the number of studies is large, this approach produces results similar to those of maximum likelihood methods (which do not require the specification of priors).²⁷ Results were summarized as medians of posterior distributions with

associated 95 percent central credible intervals (CrIs). A CrI denotes a range of values within which the parameter value is expected to fall with 95 percent probability.

Grading the Strength of Evidence

We followed the Methods Guide⁸ to evaluate the strength of the body of evidence for each Key Question with respect to the following domains: risk of bias, consistency, directness, precision, and reporting bias.^{8, 28} Generally, strength of evidence was downgraded when risk of bias was not low, in the presence of inconsistency, when evidence was indirect or imprecise, or when we suspected that results were affected by selective analysis or reporting.

We determined risk of bias (low, medium, or high) on the basis of the study design and the methodological quality. We assessed consistency on the basis of the direction and magnitude of results across studies. We considered the evidence to be indirect when we had to rely on comparisons of biopsy methods across different studies (i.e., indirect comparisons). We considered studies to be precise if the CrI was narrow enough for a clinically useful conclusion, and imprecise if the CrI was wide enough to include clinically distinct conclusions. The potential for reporting bias (“suspected” vs. “not suspected”) was evaluated with respect to publication and selective outcome and analysis reporting. We made qualitative dispositions rather than perform formal statistical tests to evaluate differences in the effect sizes between more precise (larger) and less precise (smaller) studies because such tests cannot distinguish between “true” heterogeneity between smaller and larger studies, other biases, and chance.^{29, 30} Therefore, instead of relying on statistical tests, we evaluated the reported results across studies qualitatively, on the basis of completeness of reporting, number of enrolled patients, and numbers of observed events. Judgment on the potential for selective outcome reporting bias was based on reporting patterns for each outcome of interest across studies. We acknowledge that both types of reporting bias are difficult to reliably detect on the basis of data available in published research studies. We believe that our searches (across multiple databases), combined with our plan for contacting test manufacturers (for additional data) and the authors of published studies (for data clarification) limited the impact of reporting and publication bias on our results, to the extent possible.

Finally, we rated the body of evidence using four strength of evidence levels: high, moderate, low, and insufficient.⁸ These describe our level of confidence that the evidence reflects the true effect for the major comparisons of interest.

We qualitatively evaluated similarities and differences in study populations, diagnostic methods, and outcomes among study designs. We used these comparisons to inform our judgments on applicability of study findings to clinical practice.

Results

Our literature searches identified 8,637 potentially relevant citations (including 1,127 rescreened from the original ECRI evidence report). The full-length articles of 2,480 of these studies were obtained and examined in full text. Finally, 128 new studies were considered eligible for inclusion in the updated review (54, 70, and 59 new studies for Key Questions 1, 2, and 3 respectively), for a total of 316 included studies.

Key Question 1: In women with a palpable or nonpalpable breast abnormality, what is the diagnostic test performance of different types of core needle breast biopsy compared with open biopsy or with each other?

One hundred and sixty studies, published between 1990 and 2013, provided information on test performance outcomes of core needle biopsy (54 new studies and 106 studies included in the original evidence report; another study included in the original report overlapped with one of the newer studies and was excluded). Fifty studies were prospectively designed, and 58 were conducted in the United States. Ten studies provided outcome information on more than one group of patients (typically undergoing biopsy with a different biopsy device). In statistical analyses, these groups were treated separately, leading to a total of 171 independent patient groups with information on 69,804 breast lesions.

Test Performance of Open Surgical Biopsy

Published information on the test performance of open surgical biopsy was limited. However, research studies of needle biopsy methods and technical experts generally suggested that open surgical biopsy could be considered a “gold” standard test (i.e., a test without measurement error). One study included in the ECRI report stated that open surgical biopsy may miss one to two percent of breast cancers (i.e., sensitivity of 98% or greater). The original evidence report did not identify any information on underestimation rates for open surgical biopsy. We found a single study that reported underestimation in 16.7 percent of ADH lesions (1 of 6) and 7.1 percent of DCIS lesions (1 of 14) diagnosed thorough open biopsy. The small

number of lesions in this study precludes reliable conclusions. Because open surgical biopsy samples the entire target lesion or a large part of it, in theory underestimation rates can be reduced to zero.

Test Performance of Core Needle Biopsy Methods

A total of 160 studies contributed information to analyses of test performance of core needle biopsy methods; 154 enrolled women at average risk and only 6 enrolled women specified to be at high risk of cancer. Studies varied by type of imaging guidance (stereotactic guidance, ultrasound guidance, MRI guidance, other guidance, or freehand), how the biopsy sample was extracted (automated or vacuum), and other factors (e.g., needle size). If studies included multiple cohorts of patients undergoing biopsy by different methods (e.g., some patients were biopsied with vacuum-assistance and others were not) but the study did not report the test performance of each method, these groups were treated together as ‘multiple methods’ in statistical analyses for that factor. One hundred and thirty-one study groups reported the use of a single form of imaging guidance (83 stereotactic; 41 ultrasound; 6 MRI; 1 grid), whereas 10 used freehand methods, 29 used multiple methods, and one did not report adequate details. Sixty study arms used vacuum-assisted methods to obtain the biopsy sample; 80 used automated methods; 30 used multiple methods; and 1 did not report adequate details. Needle gauge also varied across studies: 57 used 14G needles, 9 used smaller needles, 46 used larger bores, and 48 studies did not report relevant information, or used a range of needle sizes. Reference standard tests also differed across studies: 26 used open biopsy on all included patients; 94 used mean or median followup of between 6 and 24 months for test negative patients, and 40 used mean or median followup of 24 months or more for test negative cases. Additional study details are available in the SRDR. Consistent with the findings of the original report, the risk of bias for this body of evidence was considered moderate to high, mainly due to concerns about spectrum bias, retrospective data collection, differential verification, and lack of information regarding the blinding of reference standard test assessors to the index test results.

The frequency of malignant disease (invasive cancer or DCIS, at the lesion level) ranged from 1 percent to 94 percent, with a median of 34 percent. The proportion of correct diagnoses ranged from 68 percent to 100 percent, with a median of 96 percent. Table B summarizes meta-

analysis results for alternative diagnostic biopsy methods, together with information on the number of lesions evaluated with each method, for women at average risk of cancer. Sensitivity estimates were higher than 0.90 and specificity estimates were higher than 0.91 for all methods. CrIs, particularly for ultrasound- and stereotactically-guided biopsy methods, were fairly precise, reflecting the large number of studies reporting information on the test performance of these methods. In contrast, results for MRI-guided methods were based on only three studies and were imprecise, particularly for sensitivity. Table C summarizes the same information for women deemed to be at high risk for cancer (e.g., due to genetic factors or strong family history). Information for this subgroup

was limited (6 studies) and we did not find evidence to suggest that the test performance of breast biopsy methods was different between women at average and high risk of cancer. However, there was substantial uncertainty around the relative test performance estimates of the two groups. Table D summarizes the results of analyses of underestimation rates for women at average risk of breast cancer. Results were rather imprecise (e.g., CrI widths were often wider than 0.1) for all estimates except the underestimation rate for stereotactically guided, vacuum-assisted biopsy methods. Analyses of underestimation rates were not possible for women at high risk of cancer because of lack of data.

Table B. Summary estimates of test performance for alternative core needle biopsy methods— women at average risk of cancer

Biopsy Method or Device	N Studies [N biopsies] for Sensitivity & Specificity	Sensitivity	Specificity
Freehand, automated	10 [786]	0.91 (0.80 to 0.96)	0.98 (0.95 to 1.00)
US-guided, automated	27 [16287]	0.99 (0.98 to 0.99)	0.97 (0.95 to 0.98)
US-guided, vacuum-assisted	12 [1543]	0.97 (0.92 to 0.99)	0.98 (0.96 to 0.99)
Stereotactically guided, automated	37 [9535]	0.97 (0.95 to 0.98)	0.97 (0.96 to 0.98)
Stereotactically guided, vacuum-assisted	43 [14667]	0.99 (0.98 to 0.99)	0.92 (0.89 to 0.94)
MRI-guided, automated	2 [89]	0.90 (0.57 to 0.99)	0.99 (0.91 to 1.00)
MRI-guided, vacuum-assisted	1 [10]	1.00 (0.98 to 1.00)	0.91 (0.54 to 0.99)
Multiple methods/other	33 [26028]	0.99 (0.98 to 0.99)	0.96 (0.93 to 0.97)

All numbers are medians with 95% CrIs, unless otherwise stated. “Other” denotes one study using grid guidance and one study that did not report information on the use of vacuum assistance.

CrI = credible interval; DCIS = ductal carcinoma in situ; MRI = magnetic resonance imaging; N = number; NA = not applicable; US = ultrasound.

Table C. Summary estimates of test performance for alternative core needle biopsy methods— women at high risk of cancer

Biopsy Method or Device	N Studies [N biopsies] for Sensitivity and Specificity	Sensitivity (95% CrI)	Specificity (95% CrI)
Stereotactically guided, automated	1 [416]	0.97 (0.82 to 1.00)	0.97 (0.91 to 0.99)
Stereotactically guided, vacuum-assisted	2 [311]	0.99 (0.93 to 1.00)	0.93 (0.79 to 0.98)
MRI-guided, automated	2 [56]	0.90 (0.58 to 0.98)	0.99 (0.92 to 1.00)
MRI-guided, vacuum-assisted	1 [76]	1.00 (0.98 to 1.00)	0.92 (0.61 to 0.99)

No studies provided information on the test performance of freehand or US-guided biopsy methods, or the use of multiple methods in populations of women at high risk of cancer. Results are based on the model with risk group as a covariate.

CrI = credible interval; DCIS = ductal carcinoma in situ; MRI = magnetic resonance imaging; N = number; US = ultrasound.

Table D. Summary estimates of underestimation rates for alternative core needle biopsy methods—women at average risk of cancer

Biopsy Method or Device	N Studies [N biopsies] for DCIS Underestimation	DCIS Underestimation Probability	N Studies [N biopsies] for High-Risk Lesion Underestimation	High-Risk Lesion Underestimation Probability
Freehand, automated	0 [0]	NA	1 [6]	0.88 (0.32 to 1.00)
US-guided, automated	14 [307]	0.38 (0.26 to 0.51)	21 [601]	0.25 (0.16 to 0.36)
US-guided, vacuum-assisted	5 [48]	0.09 (0.02 to 0.26)	9 [20]	0.11 (0.02 to 0.33)
Stereotactically guided, automated	18 [664]	0.26 (0.19 to 0.36)	29 [357]	0.47 (0.37 to 0.58)
Stereotactically guided, vacuum-assisted	34 [1899]	0.11 (0.08 to 0.14)	40 [1002]	0.18 (0.13 to 0.24)
MRI-guided, automated	0 [0]	NA	1 [1]	0.49 (0.02 to 0.97)
MRI-guided, vacuum-assisted	1 [1]	0.00 (0.00 to 0.38)	0 [0]	NA
Multiple methods/other	18 [628]	0.22 (0.15 to 0.30)	25 [866]	0.32 (0.23 to 0.41)

Analyses for underestimation were not possible for high risk women due to sparse data.

CrI = credible interval; DCIS = ductal carcinoma in situ; MRI = magnetic resonance imaging; N = number; NA = not applicable; US = ultrasound.

Comparative Test Performance

To compare test performance across different biopsy methods we used indirect (meta-regression-based) comparisons. Table E presents comparisons between pairs of biopsy methods using the same imaging guidance for sensitivity and specificity. We only examined comparisons between biopsy methods using the same imaging modality because lesion characteristics (e.g., palpability, ability to visualize a lesions) strongly influence the choice of

imaging modality. In general, differences among tests were relatively small: for example, differences in sensitivity or specificity never exceeded 0.1 (i.e., 10% absolute difference). Stereotactically guided automated biopsy had a specificity that was higher by 0.05 compared to vacuum-assisted biopsy methods, and a sensitivity that was 0.02 lower. Comparisons among MRI-guided biopsy methods were imprecise, reflecting the small number of available studies.

Table E. Differences in sensitivity between pairs of biopsy methods (meta-regression based indirect comparisons)

Biopsy Methods Compared	Difference in Sensitivity (95% CrI)	Difference in Specificity (95% CrI)
US-guided, automated vs. vacuum-assisted	0.01 (-0.01, 0.06)	-0.01 (-0.03, 0.01)
Stereotactically guided, automated vs. vacuum-assisted	-0.02 (-0.04, -0.01)	0.05 (0.02, 0.08)
MRI-guided, automated vs. vacuum assisted	-0.10 (-0.43, -0.01)	0.07 (-0.03, 0.43)

All results are shown as medians of differences (95% CrI). Positive values denote that the first-listed biopsy method has higher performance than the comparator (second listed method). CrIs that do not include the null value (0) are highlighted in bold. CrI = credible interval; MRI = magnetic resonance imaging.
CrI = credible interval; MRI = magnetic resonance imaging; US = ultrasound.

Factors That Affect Test Performance

We considered evidence on the impact of patient or study level-factors on test performance from two complementary sources: (1) within-study evidence (i.e., comparisons of test performance over levels of a factor within the patient population enrolled in a study) and (2) evidence from meta-regression analyses (that combine information across studies). Ideally, all studies would consistently report comparisons of test performance across well-defined subgroups (e.g., by patient, or lesion characteristics). Such within-study comparisons are more informative than comparisons across studies: factors related to study setting are common for all patients within the same study and other patient differences can be addressed (at least to some extent) by appropriate analytic methods (e.g., regression adjustment). In the absence of such information, one has to rely on indirect (across-study) comparisons that are generally less convincing because they cannot account for all differences across included studies.

Twenty studies provided information that allowed an evaluation of the impact of any factor on test performance. Specifically, 16 studies provided information on patient and lesion-related factors, 10 on procedural factors, and 3 on clinician and facility factors (some studies provided information on multiple factors). Of note, the majority of studies (140 of 160) did not allow investigation of the impact of any factors on test performance, raising concerns about selective analysis or reporting of results on modifiers of test performance. Among the 20 studies reporting relevant results, factors were coded inconsistently and details that would allow formal statistical testing were not available. Because of these reasons, within-study comparisons could not support conclusions regarding possible modifiers of test performance.
Meta-regression analyses were possible for the following factors: needle gauge, choice of reference standard, proportion of lesions that were palpable, country where the study was performed, whether multiple centers

contributed patients to a study, study design, and risk of bias. In general, test performance was not affected by the factors examined (i.e., CrIs included the null value), with the exception of higher sensitivity in studies conducted in the United States (vs. any other country), and higher specificity in studies using followup of 6 to 24 months (as compared to studies using surgical pathology results for all patients) and studies with a prospective design (as compared to studies with a retrospective design). These results must be interpreted with caution given that they reflect indirect comparisons across studies, which cannot be adjusted for other factors that vary across studies.

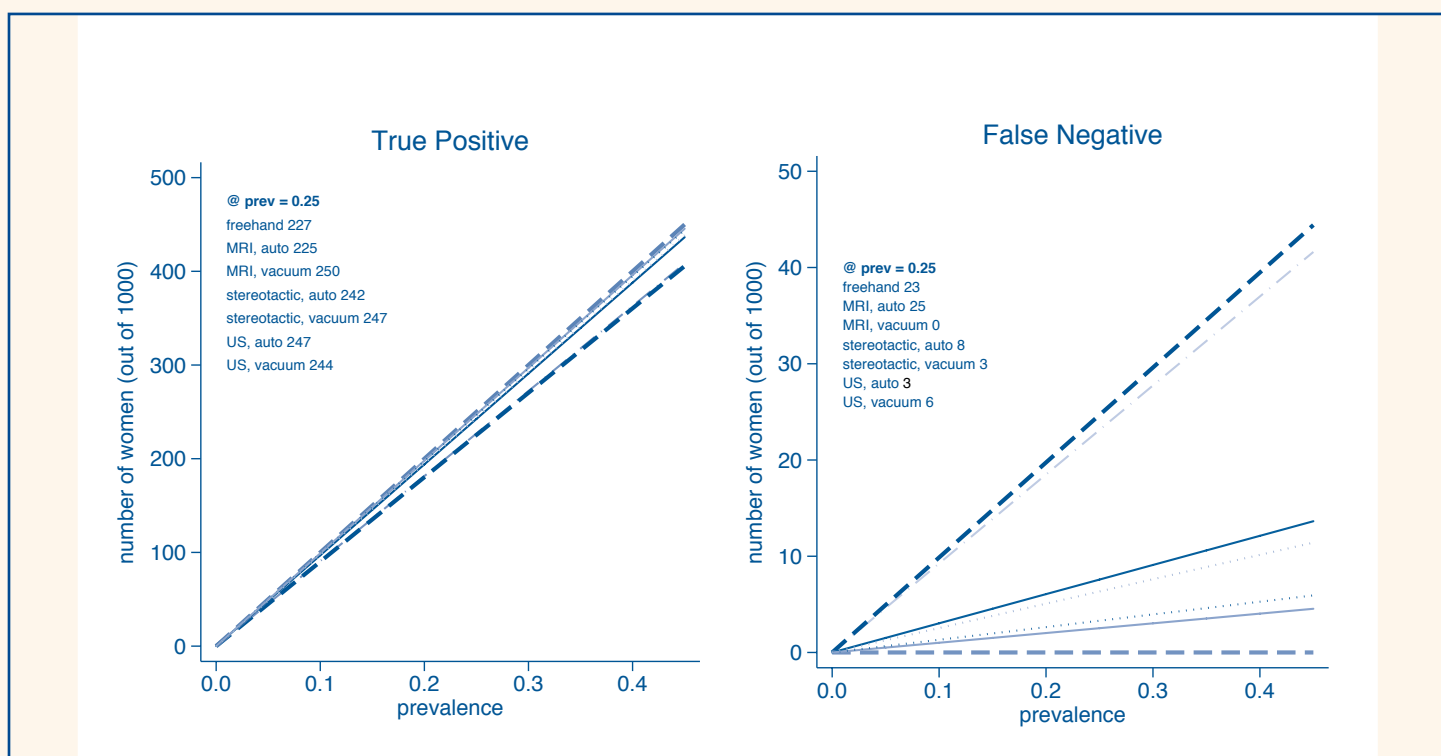
Overall, within-study analyses and meta-regression analyses were insufficient to confirm (or exclude) any single factor as a modifier of test performance.

Contextualizing the Results of Test Performance Meta-Analyses

To contextualize the results of the test performance meta-analyses presented in the preceding sections we evaluated the impact of testing in a hypothetical cohort of 1,000 women, under alternative scenarios for disease prevalence.

Because delayed diagnosis on the basis of biopsy results is the most important (adverse) outcome related to testing we highlight here results based on false negative biopsies (and their complement, true positive biopsies) in Figure A. In populations with low cancer prevalence, the number of cases where treatment may be delayed on the basis of biopsy results (i.e., false negative biopsies) is expected to be small (e.g., for all ultrasound or stereotactically guided biopsy methods less than 5 out of 1,000 women, if prevalence is 10 percent or less). As prevalence increases the number of false negative results increases for all biopsy methods, but more rapidly for MRI-guided automated and freehand methods, which had the lowest sensitivity. However, results for MRI-guided automated methods were based on only six studies. Figure A also presents numerical results for a prevalence of 25 percent, which is approximately the prevalence of breast cancer among women referred for breast biopsy in the United States. All stereotactically and ultrasound-guided methods, and MRI-guided vacuum assisted methods are expected to have fewer than 10 false negative results (for every 1,000 women undergoing biopsy), even when prevalence is as high as 0.30.

Figure A. Outcomes of testing in a hypothetical cohort of 1,000 women



Lines correspond to different test modalities: light blue dashed-dotted = freehand; blue solid = stereotactically guided, automated; light blue solid = stereotactically guided, vacuum-assisted; dark blue dotted = US-guided, automated; light blue dotted = US-guided, vacuum-assisted; dark blue dashed = MRI-guided, automated; light blue dashed = MRI-guided, vacuum assisted.

Key Question 2: In women with a palpable or nonpalpable breast abnormality, what are the adverse events (harms) associated with different types of core needle breast biopsy compared with open biopsy for diagnosis?

We synthesized information on adverse events from a total of 144 studies (70 new studies and 74 from the original evidence report) reporting on at least one of the outcomes relevant to Key Question 2. Overall, studies were considered to be of moderate to high risk of bias. Selective outcome reporting was considered likely for all adverse events examined, because of the large proportion of studies with unclear or missing data.

Adverse Events of Open Biopsy

Very few studies reported information about complications occurring in association with open surgical biopsy procedures. One study reported that 10.2 percent of wire-localized open biopsy procedures were complicated by vasovagal reactions. A narrative review reported that 2 to 10 percent of breast surgeries are complicated by hematoma formation, and that 3.8 are complicated by infection. Another study reported that 6.3 percent of open surgical biopsies were complicated by infections. One study reported low levels of pain with open biopsy when local lidocaine was used. A fifth study reported that 2.1 percent of open biopsy procedures were complicated by the development of an abscess, but zero abscesses complicated 234 ultrasound-guided vacuum-assisted core needle procedures. A sixth study reported that 4 of 100 surgical biopsies required repeat biopsy, compared to 2 of 100 vacuum-assisted core needle biopsies.

Adverse Events of Core Needle Biopsy

We identified 141 studies reporting information on at least one of the adverse events of interest following core needle biopsy (26 reported information related to the displacement of cancerous cells during biopsy). Overall, core needle biopsy appeared to have a lower risk of complications than open surgical biopsy; however, direct comparative information was sparse. The incidence of severe complications with core needle biopsy was less than 1 percent. The incidence of all adverse events was low: in more than 50 percent of studies reporting information on hematomas, bleeding, vasovagal reactions, and infections, the percentage of patients experiencing each of the aforementioned outcomes was less than 1.5 percent; in 75 percent of studies the event rate was less than 1 percent for infections, less than 5 percent for bleeding and vasovagal reactions, and less than 9 percent for hematoma formation.

Overall, 47 studies provided information on bleeding events that required additional treatment; more than half of the studies reported that no bleeding events requiring treatment were observed and the rate was lower than 0.14 percent in 75 percent of the studies. Use of vacuum assistance was associated with a greater rate of bleeding and hematoma formation.

Of 14 studies that used histopathology to demonstrate displacement of cells by core needle biopsy procedures (9 cohort and 5 case series or case reports), the percentage of needle tracks reported to contain displaced cancerous cells ranged from 0 to 69 percent. The clinical significance of these findings is unclear; tumor development on the biopsy needle track is extremely rare.

Factors That Affect the Development of Adverse Events

Five studies provided information on patient and lesion-related factors, eight studies provided information on procedural factors, and one study provided information on clinician and facility factors. The vast majority of studies reporting on adverse events from core needle biopsy did not allow investigation of the impact of factors on adverse events and no individual factor was evaluated by more than five of the total included studies, raising concerns regarding selective outcome and analysis reporting. No studies reported information on factors that affect the development of adverse events from open biopsy. We did not perform meta-regression analyses because studies reported information on adverse events inconsistently and because data were missing from more than half of the studies for all adverse events. Studies suggested that vacuum-assisted biopsy methods led to increased bleeding and performing biopsies with patients seated upright was associated with increased incidence of vasovagal reactions; however, results were reported in a way that precluded quantitation of the relative risk.

Key Question 3: How do open biopsy and various core needle techniques differ in terms of patient preference, availability, costs, availability of qualified pathologist interpretations, and other factors that may influence choice of a particular technique?

We reviewed a total of 143 studies for Key Question 3 (59 new studies and 84 studies from the original report). Generally, the evidence supported the conclusions of the original report that core needle biopsy costs less than open surgical biopsy, consumes fewer resources, and is preferred by patients. In addition, utilization of core needle biopsy has grown consistently since the mid-1990s.

Studies reported that women were generally satisfied with the cosmetic results of core needle procedures. Transient intense anxiety just before and during the procedure may be common, and may be partially ameliorated with the use of medication, relaxation and empathy techniques, or hypnosis. Based on 42 studies providing relevant information, core needle biopsy obviated the need for surgical procedures in about 75 percent of women. Ten studies reported comparisons against open surgical biopsy with respect to the number of patients requiring only one surgical procedure (vs. more than one) after cancer diagnosis. Meta-analysis of these studies suggested that the odds of requiring only one surgical procedure were almost 15 times higher among women receiving core needle biopsy; odds ratio = 14.8 (95% CrI, 7.2 to 50.2). This result should be interpreted with caution because confounding by indication is likely.

Discussion

Key Findings and Assessment of the Strength of Evidence

In this update of the 2009 Comparative Effectiveness Review on breast biopsy methods we synthesized evidence from a total of 316 studies (128 new studies and 188 from the original report). We found few studies providing information on the test performance of open surgical biopsy. In contrast, the evidence base on core needle biopsy methods now includes a large number of studies reporting on almost 70,000 breast lesions. This allowed us to assess the comparative performance of tests (when using the same type of imaging guidance), in addition to updating the 2009 report's evaluation of the performance of individual biopsy methods. Tables F-H summarize our assessment of the strength of evidence. Following the original evidence report, and in view of the paucity of evidence on open surgical biopsy, we refrained from rating the strength of evidence for this technique for all Key Questions. For Key Questions 1 and 2, we assessed the strength of evidence by integrating our (subjective) judgments on the risk of bias of included studies, the

consistency of their findings, the directness of the available data, and the precision of quantitative results. For Key Question 3 we only rated the strength of evidence for the outcome of additional surgical procedures required after biopsy. We did not rate the strength of evidence for other Key Question 3 outcomes because of the diversity of designs employed and outcomes addressed (see the Methods section for our approach to rating the strength of evidence).

Test Performance and Comparative Test Performance

Among women at average risk of cancer, core needle biopsy using ultrasound or stereotactic guidance had average sensitivities ranging from 0.97 to 0.99 and average specificities ranging from 0.92 to 0.98. Freehand biopsy methods appeared to have lower average sensitivity (0.91) compared to other methods, but similar specificity. Stereotactically guided automated techniques were associated with lower sensitivity and higher specificity compared to stereotactically guided vacuum-assisted methods. Although these results were fairly precise, they were derived from indirect comparisons across studies of moderate to high risk of bias. MRI-guided biopsies were evaluated in only six studies with small sample sizes, leading to substantial uncertainty around estimates of test performance. Table F summarizes our assessment of the strength of evidence for alternative biopsy methods in women at average risk of cancer and for comparisons among biopsy methods using the same imaging guidance modality.

We did not find a difference in test performance between women at low and high risk of breast cancer. Because the number of studies of women at high risk of cancer was small, comparisons of test performance between low and high risk women had substantial uncertainty and results were not sufficient to support definitive conclusions. Evidence on modifiers of test performance was also sparse for all biopsy methods, raising concerns about selective outcome and analysis reporting.

Table F. Strength of evidence about comparative test performance in women at average risk of breast cancer

Outcome	Comparison or Biopsy Method	Overall Rating	Key Findings and Comments
Test performance of individual biopsy methods	Freehand	Low	– Sensitivity: 0.91 (0.80 to 0.96) – Specificity: 0.98 (0.95 to 1.00)
	Ultrasound, automated	Moderate	– Sensitivity: 0.99 (0.98 to 0.99) – Specificity: 0.97 (0.95 to 0.98)
	Ultrasound, vacuum-assisted	Moderate	– Sensitivity: 0.97 (0.92 to 0.99) – Specificity: 0.98 (0.96 to 0.99)
	Stereotactically guided, automated	Moderate	– Sensitivity: 0.97 (0.95 to 0.98) – Specificity: 0.97 (0.96 to 0.98)
	Stereotactically guided, vacuum-assisted	Moderate	– Sensitivity: 0.99 (0.98 to 0.99) – Specificity: 0.92 (0.89 to 0.94)
	MRI-guided, automated	Insufficient	– Sensitivity: 0.90 (0.57 to 0.99) – Specificity: 0.99 (0.91 to 1.00)
	MRI-guided, vacuum-assisted	Insufficient	– Sensitivity: 1.00 (0.98 to 1.00) – Specificity: 0.91 (0.54 to 0.99)
Comparison of test performance among alternative biopsy methods	Ultrasound-guided, automated vs. vacuum-assisted	Low	– Difference in sensitivity: 0.01 (-0.01 to 0.06) [no difference] – Difference in specificity: -0.01 (-0.03 to 0.01) [no difference]
	Stereotactically guided, automated vs. vacuum-assisted	Low	– Difference in sensitivity: -0.02 (-0.04 to -0.01) [vacuum-assisted is better] – Difference in specificity: 0.05 (0.02 to 0.08) [automated is better]
	MRI-guided, automated vs. vacuum-assisted	Insufficient	– Difference in sensitivity: -0.10 (-0.43 to -0.01) [vacuum-assisted is better] – Difference in specificity: 0.07 (-0.03 to 0.43) [no difference]

Table F. Strength of evidence about comparative test performance in women at average risk of breast cancer (continued)

Outcome	Comparison or Biopsy Method	Overall Rating	Key Findings and Comments
Modifiers of test performance for women at average and high risk of breast cancer	All biopsy methods	Insufficient	<ul style="list-style-type: none"> – Few studies provided within-sample information for each modifier of interest; meta-regression results rely on cross-study comparisons so consistency of effects cannot be assessed – Within-study (direct) evidence was sparse; between study evidence relied on indirect comparisons across studies – In meta-regression analyses CrIs were wide; extreme odds ratio values were often observed because sensitivity and specificity for all tests were very close to 1 (see Results)

CrI = credible interval; MRI = magnetic resonance imaging.

Underestimation Rates

Underestimation rates varied among alternative biopsy methods and were often imprecisely estimated because of the relatively small number of lesions contributing data for these analyses. In general, underestimation was less

common with stereotactically guided vacuum-assisted biopsy methods, as compared to stereotactically or ultrasound-guided automated methods. Our assessment of the strength of evidence for this outcome is summarized in Table G.

Table G. Strength of evidence for underestimation rates in women at average risk of cancer

Outcome	Comparison or Biopsy Method	Overall Rating	Key Findings and Comments
DCIS underestimation	Ultrasound-guided, automated	Low	– Average underestimation probability: 0.38 (0.26 to 0.51) [14 studies]
	Ultrasound-guided, vacuum-assisted	Low	– Average underestimation probability: 0.09 (0.02 to 0.26) [5 studies]
	Stereotactically guided, automated	Low	– Average underestimation probability: 0.26 (0.19 to 0.36) [18 studies]
	Stereotactically guided, vacuum-assisted	Low	– Average underestimation probability: 0.11 (0.08 to 0.14) [34 studies]
	Other biopsy methods	Insufficient	No available studies or few studies with small numbers of lesions

Table G. Strength of evidence for underestimation rates in women at average risk of cancer (continued)

Outcome	Comparison or Biopsy Method	Overall Rating	Key Findings and Comments
High risk lesion underestimation rate	Ultrasound-guided, automated	Low	– Average underestimation probability: 0.25 (0.16 to 0.36) [21 studies]
	Ultrasound-guided, vacuum-assisted	Low	– Average underestimation probability: 0.11 (0.02 to 0.33) [9 studies]
	Stereotactically guided, automated	Low	– Average underestimation probability: 0.47 (0.37 to 0.58) [29 studies]
	Stereotactically guided, vacuum-assisted	Low	– Average underestimation probability: 0.18 (0.13 to 0.24) [40 studies]
	Other biopsy methods	Insufficient	No available studies or few studies with small numbers of lesions

DCIS = ductal carcinoma in situ.

Adverse Events and Additional Surgeries After Biopsy

In general, adverse events were reported inconsistently, raising concerns about selective outcome and analysis reporting. Few studies provided information on the harms of open surgical biopsy. Core needle biopsy was only infrequently associated with serious adverse events. Comparisons between open and core needle biopsy are based on indirect comparisons and expert opinion, with limited empirical evidence. Open biopsy appeared to be associated with an increased incidence of adverse events (including serious adverse events) compared to core needle biopsy. Our assessment of the strength of evidence for adverse events is summarized in Table H.

Among core needle biopsy methods, vacuum-assisted methods appeared to be associated with increased bleeding. Sitting upright during the biopsy procedure was associated with more vasovagal reactions. Information about the dissemination or displacement of cancer cells during the biopsy procedure was provided by a small

number of studies with various designs. Studies reported that women were generally satisfied with the cosmetic results of core needle procedures.

Women diagnosed with breast cancer by core needle biopsy were able to have their cancer treated with a single surgical procedure more often than women diagnosed by open surgical biopsy. Although the magnitude of this association was large (the ratio of the odds was approximately 15), women and their physicians are likely to choose biopsy methods on the basis of factors (e.g., lesion location, or characteristics of the lesion on imaging) that may also be associated with the need for additional surgeries. Thus, confounding by indication is likely, and we rated the strength of evidence for this association as moderate. A difference in the rate of additional surgeries among women diagnosed with alternative biopsy methods is likely, but we have less confidence that it is an effect of the biopsy methods per se or that the magnitude of the difference is known.

Table H. Strength of evidence assessment for adverse events of breast biopsy

Outcomes	Test or Comparison	Overall Rating	Key Findings
Bleeding (any severity)	Alternative core needle biopsy methods	Low	<ul style="list-style-type: none"> – Median %: 1.21 (25th perc. = 0.33; 75th perc = 3.97) – Selective outcome and analysis reporting likely – Few studies reported bleeding requiring treatment; the event rate was low (<0.40 perc.) in those studies
Bleeding events that require treatment	Comparisons among alternative core needle biopsy methods	Low	<ul style="list-style-type: none"> – Median %: 0 (25th perc. = 0; 75th perc = 0.14) – Selective outcome and analysis reporting likely – Few studies reported bleeding requiring treatment; the event rate was low
Hematoma formation	Alternative core needle biopsy methods	Low	<ul style="list-style-type: none"> – Median %: 1.44 (25th perc. = 0.25; 75th perc = 8.57) – Selective outcome and analysis reporting likely
Infectious complications	Alternative core needle biopsy methods	Low	<ul style="list-style-type: none"> – Median %: 0 (25th perc. = 0; 75th perc = 0.33) – Selective outcome and analysis reporting likely
Vasovagal reactions:	Alternative core needle biopsy methods	Low	<ul style="list-style-type: none"> – Median %: 1.27 (25th perc. = 0.37; 75th perc = 3.88) – Potential for selective outcome and analysis reporting
Pain and severe pain	Alternative core needle biopsy methods	Low	25 studies of a wide variety of biopsy methods reported information about patient pain during the procedure (pain was assessed heterogeneously across studies)
Other adverse events	Alternative core needle biopsy methods	Insufficient	<ul style="list-style-type: none"> – Most events were reported by a single study precluding assessment of consistency – Individual studies did not provide adequate information for precise estimation of the event rate) – Only informal indirect comparisons among biopsy methods were possible – Selective outcome and analysis reporting likely
Modifiers of adverse events – vasovagal reactions	Sitting upright during the biopsy procedure	Low	<ul style="list-style-type: none"> – Vasovagal reactions were more common among patients sitting during the biopsy procedure – Results were reported in few studies (11 studies; 8 from the original evidence report and 3 from this update) – Selective outcome and analysis reporting likely
Modifiers of adverse events – bleeding	Vacuum-assisted versus non-vacuum assisted biopsy methods	Low	<ul style="list-style-type: none"> – Vacuum-assisted procedures were generally associated with increased rates of bleeding and hematoma formation – Bleeding events were generally uncommon – Comparisons among biopsy methods were based on informal indirect comparisons (across studies) – Selective outcome and analysis reporting likely
All other modifiers of adverse events	Comparisons among alternative core needle biopsy methods	Insufficient	<ul style="list-style-type: none"> – Most factors assessed by a single study limiting our ability to assess consistency – Selective outcome and analysis reporting likely – Within-study comparisons provided direct evidence

perc. = percentile.

Applicability of Review Findings

The existing evidence base on core needle biopsy of breast lesions in women at average risk of cancer appears to be applicable to clinical practice in the United States. The average age was similar to that of women undergoing breast biopsy in the United States, and the indications were similar to the prevalent indications in clinical practice (i.e. mammographic findings of suspicious lesions). Almost all studies were carried out in either the United States or in industrialized European or Asian countries where core-biopsy methods are likely sufficiently similar to those used in the United States. The applicability of our findings to women at high risk of breast cancer is uncertain because we found few studies explicitly reporting on groups of patients at high baseline risk of breast cancer and comparisons of test performance between subgroups of women produced imprecise results.

Limitations of the Evidence Base

We believe that the evidence regarding the performance of core needle biopsy for diagnosis of breast lesions is limited in the following ways: (1) published evidence on the test performance and adverse events of open surgical biopsy was sparse; (2) available studies were at moderate to high risk of bias and information on patient selection criteria, patient or lesion characteristics, adverse events, or patient-relevant outcomes was often missing or inconsistently reported, and pathology results were not reported with adequate granularity; (3) studies typically used lesions (or biopsy procedures) as the unit of analysis, instead of patients, reporting results in a way that did not allow for the correlation to be accounted for in our statistical analyses; (4) studies provided limited information to assess the impact of various patient-, lesion-, procedure-, or system- related factors on the outcomes of breast biopsy; (5) the number of studies on MRI-guided biopsy for women at average or high risk of cancer was small; (6) limited information existed on the comparative effectiveness of alternative biopsy methods on patient-relevant outcomes, resource use and logistics, and availability of technology and expertise for different core needle biopsy techniques.

Limitations of This Review

Our work has several limitations, which—to a large extent—reflect the limitations of the underlying evidence base. Because of selective, incomplete, or no reporting of necessary information, our ability to explore between-study heterogeneity was limited. Further, because we relied on published information, we were unable to evaluate the

impact of patient- or lesion-level factors on outcomes of interest. We did not include studies published in languages other than English; however, given the very large number of studies from diverse geographic locations included in the review, we believe that the addition of non-English language studies would not affect our conclusions.

The reference standard in the reviewed studies was a combination of clinical followup and pathologic confirmation. We assumed that these diagnostic methods have negligible measurement error (i.e., that they represent a “gold” standard). It is unlikely that this assumption is exactly true. However, we believe that the error rate of the reference standard is low enough that its influence on our estimates is unlikely to be substantial.

Future Research Needs

There is now a large body of evidence indicating that stereotactic and ultrasound guided core needle techniques have comparable sensitivity to each other and to open biopsy. The next focus of research should be biopsy under MRI guidance, which is a new technique that is likely to come into wider use. The data is not yet adequate to define its advantages or disadvantages of MRI guided biopsy compared with alternative techniques. Studies should be powered to achieve adequate precision (i.e., produce confidence intervals or CrIs that are narrow enough to allow clinically meaningful conclusions), have a prospective design, enroll patients across multiple centers, and use standardized histological classification systems for pathological classification.^{31, 32} For all biopsy methods, additional well-designed and fully reported prospective cohort studies are needed, primarily for addressing questions about the impact of patient-, lesion-, procedure-, or system-level factors on test performance, adverse events, and patient-relevant outcomes. This would help resolve uncertainties regarding effect modification (e.g., over patient and lesion factors) that cannot be resolved with the currently available data. Such studies could be conducted at relatively low cost, and large-scale databases of prospectively-collected observational data on breast biopsy procedures and outcomes could be used to evaluate the comparative effectiveness of alternative biopsy methods with respect to short and long term outcomes, and potential modifying factors. In all future studies, baseline risk of cancer development should be characterized using consistent and widely accepted criteria to allow appropriate subgroup analyses. We believe that a randomized comparison of alternative biopsy methods would not be fruitful because existing studies indicate that biopsy procedures have sensitivities

and specificities that are fairly similar and also close to 1. Additional information is also needed to identify factors that may influence the rate of adverse events of specific biopsy methods. Future research needs to be reported in accordance with recent reporting guidelines (e.g., STANDards for the Reporting of Diagnostic accuracy studies; www.stard-statement.org/), for progress to be made on these questions.³³

Conclusions

A large body of evidence indicates that ultrasound- and stereotactically-guided core needle biopsy procedures have sensitivity and specificity close to that of open biopsy procedures, and are associated with fewer adverse events. The strength of the evidence on the test performance of these methods is deemed moderate because studies are at medium to high risk of bias, but provide precise results and exhibit low heterogeneity. Freehand procedures have lower sensitivity than imaging-guided methods. The strength of conclusions about the comparative test performance of automated and vacuum-assisted devices (when using the same imaging guidance) is deemed low, because of concerns about the risk of bias of included studies and the reliance on indirect comparisons. There were insufficient data to draw conclusions for MRI-guided biopsy or women at high baseline risk of cancer. Harms were reported inconsistently, raising concerns about selective outcome and analysis reporting. There is low strength of evidence that vacuum-assisted procedures appear to have a higher risk of bleeding than automated methods. There is moderate strength of evidence that women diagnosed with breast cancer by core needle biopsy were more likely to have their cancer treated with a single surgical procedure, compared with women diagnosed by open surgical biopsy.

References

1. American Cancer Society. Breast Cancer Facts & Figures 2013-2014. Atlanta, GA: American Cancer Society; 2013. www.cancer.org/acs/groups/content/@research/documents/document/acspc-042725.pdf. Accessed March 26, 2014.
2. Hubbard RA, Kerlikowske K, Flowers CI, et al. Cumulative probability of false-positive recall or biopsy recommendation after 10 years of screening mammography: a cohort study. *Ann Intern Med*. 2011 Oct 18;155(8):481-92. PMID: 22007042.
3. Elmore JG, Barton MB, Moceri VM, et al. Ten-year risk of false positive screening mammograms and clinical breast examinations. *N Engl J Med*. 1998 Apr 16;338(16):1089-96. PMID: 9545356.
4. Yu YH, Wei W, Liu JL. Diagnostic value of fine-needle aspiration biopsy for breast mass: a systematic review and meta-analysis. *BMC Cancer*. 2012;12:41. PMID: 22277164.
5. Bernardi D, Borsato G, Pellegrini M, et al. On the diagnostic accuracy of stereotactic vacuum-assisted biopsy of nonpalpable breast abnormalities. Results in a consecutive series of 769 procedures performed at the Trento Department of Breast Diagnosis. *Tumori*. 2012 Jan-Feb;98(1):113-8. PMID: 22495711.
6. Bruening W, Schoelles K, Treadwell J, et al. Comparative Effectiveness of Core-Needle and Open Surgical Biopsy for the Diagnosis of Breast Lesions. Comparative Effectiveness Review No. 19 (Prepared by the ECRI Institute Evidence-based Practice Center under Contract No. 290-02-0019). AHRQ Publication No. 10-EHC007-EF. Rockville, MD: Agency for Healthcare Research and Quality; December 2009. www.ncbi.nlm.nih.gov/books/NBK45220/pdf/TOC/pdf. Accessed April 25, 2013.
7. Bruening W, Fontanarosa J, Tipton K, et al. Systematic review: comparative effectiveness of core-needle and open surgical biopsy to diagnose breast lesions. *Ann Intern Med*. 2010 Feb 16;152(4):238-46. PMID: 20008742.
8. Agency for Healthcare Research and Quality. Methods Guide for Effectiveness and Comparative Effectiveness Reviews. Rockville, MD: Agency for Healthcare Research and Quality; April 2012. www.effectivehealthcare.ahrq.gov/ehc/products/60/318/MethodsGuide_Prepublication-Draft_20120523.pdf. Accessed April 25, 2013.
9. Moher D, Liberati A, Tetzlaff J, et al. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Int J Surg*. 2010;8(5):336-41. PMID: 20171303.
10. Leeflang MM, Scholten RJ, Rutjes AW, et al. Use of methodological search filters to identify diagnostic accuracy studies can lead to the omission of relevant studies. *J Clin Epidemiol*. 2006 Mar;59(3):234-40. PMID: 16488353.
11. Whiting P, Westwood M, Burke M, et al. Systematic reviews of test accuracy should search a range of databases to identify primary studies. *J Clin Epidemiol*. 2008 Apr;61(4):357-64. PMID: 18313560.
12. Whiting PF, Rutjes AW, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med*. 2011 Oct 18;155(8):529-36. PMID: 22007046.
13. Whiting P, Rutjes AW, Reitsma JB, et al. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. *BMC Med Res Methodol*. 2003 Nov 10;3:25. PMID: 14606960.
14. Whiting P, Rutjes AW, Dinnes J, et al. Development and validation of methods for assessing the quality of diagnostic accuracy studies. *Health Technol Assess*. 2004 Jun;8(25):iii, 1-234. PMID: 15193208.
15. Whiting PF, Weswood ME, Rutjes AW, et al. Evaluation of QUADAS, a tool for the quality assessment of diagnostic accuracy studies. *BMC Med Res Methodol*. 2006 Mar 6;6:9. PMID: 16519814.
16. Wells G, Shea B, O'Connell D, et al. The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses. www.ohri.ca/programs/clinical_epidemiology/oxford.asp. Accessed 29 April 2013.

17. Higgins JP, Altman DG, Gotzsche PC, et al; Cochrane Bias Methods Group/Cochrane Statistical Methods Group. The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ*. 2011 Oct 18;343:d5928. PMID: 22008217.
18. Drummond MF, Jefferson TO. Guidelines for authors and peer reviewers of economic submissions to the BMJ. The BMJ Economic Evaluation Working Party. *BMJ*. 1996 Aug 3;313(7052):275-83. PMID: 8704542.
19. Trikalinos TA, Balion CM, Coleman CI, et al. Chapter 8: meta-analysis of test performance when there is a "gold standard". *J Gen Intern Med*. 2012 Jun;27 Suppl 1:S56-66. PMID: 22648676.
20. Trikalinos TA, Balion CM. Chapter 9: options for summarizing medical test performance in the absence of a "gold standard." *J Gen Intern Med*. 2012 Jun;27 Suppl 1:S67-75. PMID: 22648677.
21. Hamza TH, van Houwelingen HC, Stijnen T. The binomial distribution of meta-analysis was preferred to model within-study variability. *J Clin Epidemiol*. 2008 Jan;61(1):41-51. PMID: 18083461.
22. Rutter CM, Gatsonis CA. A hierarchical regression approach to meta-analysis of diagnostic test accuracy evaluations. *Stat Med*. 2001 Oct 15;20(19):2865-84. PMID: 11568945.
23. Arends LR, Hamza TH, van Houwelingen JC, et al. Bivariate random effects meta-analysis of ROC curves. *Med Decis Making*. 2008 Sep-Oct;28(5):621-38. PMID: 18591542.
24. Stijnen T, Hamza TH, Ozdemir P. Random effects meta-analysis of event outcome in the framework of the generalized linear mixed model with applications in sparse data. *Stat Med*. 2010 Dec 20;29(29):3046-67. PMID: 20827667.
25. Hamza TH, van Houwelingen HC, Heijenbrok-Kal MH, et al. Associating explanatory variables with summary receiver operating characteristic curves in diagnostic meta-analysis. *J Clin Epidemiol*. 2009 Dec;62(12):1284-91. PMID: 19398297.
26. van Houwelingen HC, Arends LR, Stijnen T. Advanced methods in meta-analysis: multivariate approach and meta-regression. *Stat Med*. 2002 Feb 28;21(4):589-624. PMID: 11836738.
27. Dahabreh IJ, Trikalinos TA, Lau J, et al. An Empirical Assessment of Bivariate Methods for Meta-Analysis of Test Accuracy. Rockville (MD); 2012.
28. Singh S, Chang SM, Matchar DB, et al. Grading a body of evidence on diagnostic tests. In: Chang SMM, D. B. Smetana, G. W. Umscheid, C. A., ed *Methods Guide for Medical Test Reviews*. AHRQ Publication No. 12-EHC017. Rockville, MD: Agency for Healthcare Research and Quality; June 2012:chapter 7. www.ncbi.nlm.nih.gov/books/NBK98241/ Accessed April 25, 2013.
29. Lau J, Ioannidis JP, Terrin N, et al. The case of the misleading funnel plot. *BMJ*. 2006 Sep 16;333(7568):597-600. PMID: 16974018.
30. Sterne JA, Sutton AJ, Ioannidis JP, et al. Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. *BMJ*. 2011;343:d4002. PMID: 21784880.
31. Ellis IO, Humphreys S, Michell M, et al. Best Practice No 179. Guidelines for breast needle core biopsy handling and reporting in breast screening assessment. *J Clin Pathol*. 2004 Sep;57(9):897-902. PMID: 15333647.
32. Compton CC. Reporting on cancer specimens: case summaries and background documentation: A publication of the College of American Pathologists Cancer Committee. 2003.
33. Bossuyt PM, Reitsma JB, Bruns DE, et al. Towards complete and accurate reporting of studies of diagnostic accuracy: The STARD Initiative. *Ann Intern Med*. 2003 Jan 7;138(1):40-4. PMID: 12513043.

Full Report

This executive summary is part of the following document: 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.



