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

Prepared for:
Agency for Healthcare Research and Quality
U.S. Department of Health and Human Services
540 Gaither Road
Rockville, MD 20850
www.ahrq.gov

Contract No. 290-2012-00012-I

Prepared by:
Brown Evidence-Based Practice Center
Providence, RI

Investigators:
Issa J. Dahabreh, M.D., M.S.
Lisa Susan Wieland, Ph.D.
Gaelen P. Adam, M.L.I.S.
Christopher Halladay, B.A., M.S.
Joseph Lau, M.D.
Thomas A. Trikalinos, M.D.

AHRQ Publication No. 14-EHC040-EF
September 2014
This report is based on research conducted by Brown Evidence-based Practice Center (EPC) under contract to the Agency for Healthcare Research and Quality (AHRQ), Rockville, MD (Contract No. 290-2012-00012-I). The findings and conclusions in this document are those of the authors, who are responsible for its contents; the findings and conclusions do not necessarily represent the views of AHRQ. Therefore, no statement in this report should be construed as an official position of AHRQ or of the U.S. Department of Health and Human Services.

The information in this report is intended to help health care decisionmakers—patients and clinicians, health system leaders, and policymakers, among others—make well-informed decisions and thereby improve the quality of health care services. This report is not intended to be a substitute for the application of clinical judgment. Anyone who makes decisions concerning the provision of clinical care should consider this report in the same way as any medical reference and in conjunction with all other pertinent information, i.e., in the context of available resources and circumstances presented by individual patients.

This report may be used, in whole or in part, as the basis for development of clinical practice guidelines and other quality enhancement tools, or as a basis for reimbursement and coverage policies. AHRQ or U.S. Department of Health and Human Services endorsement of such derivative products may not be stated or implied.

This report may periodically be assessed for the urgency to update. If an assessment is done, the resulting surveillance report describing the methodology and findings will be found on the Effective Health Care Program Web site at: www.effectivehealthcare.ahrq.gov. Search on the title of the report.

This document is in the public domain and may be used and reprinted without permission except those copyrighted materials that are clearly noted in the document. Further reproduction of those copyrighted materials is prohibited without the specific permission of copyright holders.

Persons using assistive technology may not be able to fully access information in this report. For assistance contact EffectiveHealthCare@ahrq.hhs.gov.

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

Preface

The Agency for Healthcare Research and Quality (AHRQ), through its Evidence-based Practice Centers (EPCs), sponsors the development of systematic reviews to assist public- and private-sector organizations in their efforts to improve the quality of health care in the United States. These reviews provide comprehensive, science-based information on common, costly medical conditions, and new health care technologies and strategies.

Systematic reviews are the building blocks underlying evidence-based practice; they focus attention on the strength and limits of evidence from research studies about the effectiveness and safety of a clinical intervention. In the context of developing recommendations for practice, systematic reviews can help clarify whether assertions about the value of the intervention are based on strong evidence from clinical studies. For more information about AHRQ EPC systematic reviews, see www.effectivehealthcare.ahrq.gov/reference/purpose.cfm.

AHRQ expects that these systematic reviews will be helpful to health plans, providers, purchasers, government programs, and the health care system as a whole. Transparency and stakeholder input are essential to the Effective Health Care Program. Please visit the Web site (www.effectivehealthcare.ahrq.gov) to see draft research questions and reports or to join an e-mail list to learn about new program products and opportunities for input.

We welcome comments on this systematic review. They may be sent by mail to the Task Order Officer named below at: Agency for Healthcare Research and Quality, 540 Gaither Road, Rockville, MD 20850, or by email to epc@ahrq.hhs.gov.

Richard Kronick, Ph.D.  Yen-Pin Chiang, Ph.D.
Director  Acting Deputy Director
Agency for Healthcare Research and Quality  Center for Evidence and Practice

Stephanie Chang, M.D., M.P.H.  Elisabeth Kato, M.D., M.R.P.
Director  Task Order Officer
Evidence-based Practice Program  Center for Evidence and Practice
Center for Evidence and Practice  Improvement
Agency for Healthcare Research and Quality  Agency for Healthcare Research and Quality
Technical Expert Panel

In designing the study questions and methodology at the outset of this report, the EPC consulted several technical and content experts. Broad expertise and perspectives were sought. Divergent and conflicting opinions are common and perceived as healthy scientific discourse that results in a thoughtful, relevant systematic review. Therefore, in the end, study questions, design, methodologic approaches, and/or conclusions do not necessarily represent the views of individual technical and content experts.

Technical Experts must disclose any financial conflicts of interest greater than $10,000 and any other relevant business or professional conflicts of interest. Because of their unique clinical or content expertise, individuals with potential conflicts may be retained. The TOO and the EPC work to balance, manage, or mitigate any potential conflicts of interest identified.

One Technical Expert was a patient and remains anonymous. The list of the remaining Technical Experts who participated in developing this report follows:

Patricia Carney, Ph.D.  
Oregon Health and Science University  
Portland, OR

Joanne Elmore, M.D., M.P.H.  
University of Washington  
School of Public Health  
Seattle, WA

Richard E. Fine, M.D.  
Department of Surgery  
Department of Surgical Oncology  
University of Tennessee College of Medicine  
The West Clinic Comprehensive Breast Center  
Memphis, TN

Mark A. Helvie, M.D.  
University of Michigan Health System  
Ann Arbor, MI

Carol Lee, M.D.  
Memorial Sloan Kettering Cancer Center  
New York City, NY

Barabara Monsees, M.D.  
Evens Professor of Women's Health  
Department of Radiology  
Washington University School of Medicine  
St. Louis, MO

Liane Philpotts, M.D.  
Yale University School of Medicine  
New Haven, CT

Melissa Pilewskie, M.D.  
Memorial Sloan-Kettering Cancer Center  
Evelyn H. Lauder Breast Center  
New York City, NY
Peer Reviewers

Prior to publication of the final evidence report, EPCs sought input from independent Peer Reviewers without financial conflicts of interest. However, the conclusions and synthesis of the scientific literature presented in this report does not necessarily represent the views of individual reviewers.

Peer Reviewers must disclose any financial conflicts of interest greater than $10,000 and any other relevant business or professional conflicts of interest. Because of their unique clinical or content expertise, individuals with potential non-financial conflicts may be retained. The TOO and the EPC work to balance, manage, or mitigate any potential non-financial conflicts of interest identified.

The list of Peer Reviewers follows:

Binita Ashar, M.D.
Center for Devices and Radiological Health
U.S. Food and Drug Administration
Baltimore, MD

Ira J. Bleiweiss, M.D.
Icahn School of Medicine at Mount Sinai
New York City, NY

Edi Brogi, M.D.
Memorial Sloan-Kettering Cancer Center
New York City, NY

Laurie L. Fajardo, M.D.
University of Iowa Health Care
Iowa City, IA

Seema Khan, M.D.
Feinberg School of Medicine
Northwestern University
Chicago, IL

Tracy Lively, Ph.D.
Cancer Diagnosis Program
National Cancer Institute
Bethesda, MD

Robert A. Smith, Ph.D.
American Cancer Society
Atlanta, GA
Structured Abstract

**Objective.** Core needle biopsy and open surgical biopsy are the most frequently used procedures for diagnosis of suspicious breast lesions. An AHRQ evidence report on the comparative effectiveness and adverse events of breast biopsy methods was completed in 2009. The availability of additional studies and the uncertainties surrounding newer biopsy techniques prompted an update of that report.

**Study eligibility criteria.** We searched nine electronic databases (last search on December 16, 2013) for English-language full-text reports of prospective or retrospective cohort studies of women not previously diagnosed with breast cancer who were undergoing biopsy for diagnosis of a breast lesion.

**Study appraisal and synthesis methods.** A single investigator extracted quantitative and qualitative data from each study; a second reviewer verified extracted data. We assessed the strength and applicability of the evidence. We performed Bayesian meta-analyses to estimate summary test performance and performed indirect comparisons to assess the relative effectiveness of alternative core needle biopsy methods. Statistical models accounted for between-study heterogeneity.

**Results.** One hundred and sixty studies of moderate to high risk of bias provided information on the test performance of alternative core needle biopsy techniques. We found one new study investigating the test performance of open biopsy. For women at average risk of cancer, both ultrasound- and stereotactically guided biopsies had average sensitivities higher than 0.97 and average specificities ranging from 0.92 to 0.99; freehand biopsy methods had average sensitivity of 0.91 and specificity of 0.98. However, evidence on the test performance of magnetic resonance imaging (MRI)-guided biopsy (6 studies) was insufficient to draw conclusions. Test performance did not differ between women at average and high baseline risk of cancer, but results were imprecise. Test performance of automated and vacuum-assisted devices (when using the same imaging guidance) was fairly similar (absolute differences in sensitivity and specificity ≤ 0.1). One hundred and forty-one studies contributed information on potential harms of different core needle biopsy techniques. Overall, core needle biopsy had a lower risk of complications than open surgical biopsy; however information on the latter was sparse. The absolute incidence of adverse events was low and the incidence of severe complications was less than 1 percent for all techniques. Vacuum-assisted procedures appeared to be associated with increased bleeding and hematoma formation; biopsies performed with patients seated upright appeared to be associated with increased risk of vasovagal reactions. Harms were reported inconsistently, raising concerns about selective outcome reporting. We found 10 case reports of patients developing tumors at the site of prior core needle biopsies. We found information on only a few patient-relevant and resource-related outcomes. Based on 42 studies, core needle biopsy obviated the need for surgical procedures in about 75 percent of women. Meta-analysis of
10 studies reporting the number of surgical procedures required after biopsy suggested that the odds of requiring only one procedure were almost 15 times as high among women receiving core needle biopsy, as compared to those receiving open surgical biopsy. However, this result may be confounded by indication.

**Limitations.** Patient-level data were unavailable and information about study- or population-level characteristics was too limited to allow the identification of modifiers of test performance, adverse events, or clinical outcomes. Studies reported adverse events incompletely, and did not provide details of their outcome ascertainment methods.

**Conclusions.** A large body of evidence suggests 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 and fairly consistent results. Freehand procedures have lower sensitivity than imaging-guided methods. The strength of evidence on 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. There is low strength of evidence that vacuum-assisted procedures 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 are more likely to have their cancer treated with a single surgical procedure, compared with women diagnosed by open surgical biopsy.
# Contents

## Executive Summary

Breast Cancer Epidemiology and Clinical Diagnosis ........................................ 1
Original Evidence Report and Rationale for the Update .............................. 2
Key Questions ......................................................................................... 2

## Background

- Breast Cancer Epidemiology and Clinical Diagnosis ........................................ 1
- Original Evidence Report and Rationale for the Update .............................. 2
- Key Questions ......................................................................................... 2

## Methods

- AHRQ Task Order Officer ........................................................................ 4
- External Stakeholder Input ....................................................................... 4
- Key Questions ......................................................................................... 4
- Analytic Framework ................................................................................ 4
- Scope of the Review .............................................................................. 5
  - Populations and Conditions of Interest .................................................. 5
  - Interventions ......................................................................................... 5
  - Comparators (Index and Reference Standard Tests) .............................. 5
  - Outcomes ........................................................................................... 6
  - Timing .................................................................................................. 6
  - Setting ................................................................................................ 7
  - Study Design and Additional Criteria ................................................... 7
- Literature Search and Abstract Screening ............................................... 8
- Study Selection and Eligibility Criteria ................................................... 9
- Data Abstraction and Management ........................................................ 9
- Assessment of the Risk of Bias of Individual Studies ............................. 10
- Data Synthesis ....................................................................................... 10
- Grading the Strength of Evidence .......................................................... 12
- Assessing Applicability ......................................................................... 12
- Peer Review .......................................................................................... 13

## Results

- Included Studies .................................................................................... 16
- Test Performance for Breast Cancer Diagnosis ........................................ 16
  - Test Performance of Open Surgical Biopsy ........................................... 16
  - Test Performance of Core Needle Biopsy Methods ................................ 16
- Contextualizing the Results of Test Performance Meta-Analyses .......................... 22
- Comparative Test Performance ............................................................... 24
- Factors That Affect Test Performance .................................................... 25
  - Within-Study Evidence ...................................................................... 25
  - Meta-Regression Analyses ................................................................. 29
- Risk Of Bias Assessment for Studies Addressing Key Question 1 ............ 30
- Adverse Events of Open Biopsy ............................................................. 30
- Adverse Events of Core Needle Biopsy .................................................. 31
- Hematomas and Bleeding ..................................................................... 31
Infections ................................................................................................................. 33
Pain and Use of Pain Medications .......................................................................... 34
Vasovagal Reactions ................................................................................................ 34
Impact of Biopsy Procedure on Usual Activities and Time to Recovery .................. 35
Impact of Biopsy Procedure on Subsequent Mammographic Procedures ............... 35
Miscellaneous Reported Adverse Events ............................................................... 36
Dissemination and Displacement of Cancerous Cells During the Biopsy Procedure .... 36
Factors That Modify the Association of Biopsy Procedures With Adverse Events .... 37
Anxiety and Distress .............................................................................................. 40
Procedure Preference ......................................................................................... 41
Surgical Procedures Avoided ............................................................................. 41
Cosmetic Results ................................................................................................. 41
Resource Utilization and Costs .......................................................................... 42
Physician Experience .......................................................................................... 42
Availability of a Qualified Pathologist ................................................................. 43
Availability of Equipment/Utilization ................................................................. 43
Procedure Duration Time ................................................................................. 43
Time to Complete Tumor Removal ................................................................... 44
Wait Time for Test Results ............................................................................... 44
Recurrence Rates ............................................................................................... 45

Discussion ............................................................................................................. 46
Key Findings and Assessment of the Strength of Evidence ..................................... 46
Test Performance and Comparative Test Performance ......................................... 46
Underestimation Rates ....................................................................................... 47
Adverse Events and Additional Surgeries After Biopsy ........................................ 48
Limitations of the Evidence Base ........................................................................ 49
Strengths and Limitations of This Review .......................................................... 50
Applicability of Review Findings ....................................................................... 51
Evidence Gaps and Ongoing Research ............................................................... 51
Future Research Needs ...................................................................................... 52
Conclusions ......................................................................................................... 53
Abbreviations ................................................................................................ ...... 54
References .......................................................................................................... 55

Tables

Table A. Definitions of diagnostic groups based on index and reference standard test results
Table B. Summary estimates of test performance for alternative core needle biopsy methods—
women at average risk of cancer ........................................................................ 46
Table C. Summary estimates of test performance for alternative core needle biopsy methods—
women at high risk of cancer ............................................................................ 47
Table D. Summary estimates of underestimation rates for alternative core needle biopsy
methods—women at average risk of cancer ......................................................... 48
Table E. Differences in sensitivity between pairs of biopsy methods (meta-regression based
indirect comparisons) ..................................................................................... 49
Table F. Strength of evidence about comparative test performance in women at average risk of breast cancer............................................................................................................. ES-16
Table G. Strength of evidence for underestimation rates in women at average risk of cancer........................................................................................................................ ES-17
Table H. Strength of evidence assessment for adverse events of breast biopsy.................... ES-18
Table 1. Definitions of diagnostic groups based on index and reference standard test results...... 7
Table 2. Summary of new evidence evaluated in this update....................................................... 15
Table 3. Summary estimates of test performance for alternative core needle biopsy methods – women at average risk of cancer........................................................................ 18
Table 4. Summary estimates of test performance for alternative core needle biopsy methods – women at high risk of cancer .................................................................................... 18
Table 5. Differences in sensitivity between pairs of biopsy methods (meta-regression based indirect comparisons)........................................................................................................... 25
Table 6. Studies evaluating factors that may affect test performance (20 studies, reporting on multiple factors each) .................................................................................................... 26
Table 7. Meta-regression analysis for test performance outcomes.............................................. 29
Table 8. Adverse events associated with core needle biopsy for breast cancer diagnosis........ 31
Table 9. Core needle biopsy procedures and rates of hematoma formation and bleeding ........ 33
Table 10. Core needle biopsy procedures and rates of infectious complications ....................... 34
Table 11. Core needle biopsy procedures and rates of vasovagal reactions ............................... 35
Table 12. Studies evaluating factors that may affect the incidence of adverse events .................. 38
Table 13. Strength of evidence about comparative test performance in women at average risk of breast cancer ............................................................................................................. 47
Table 14. Strength of evidence for underestimation rates in women at average risk of cancer ... 48
Table 15. Strength of evidence assessment for adverse events of biopsy ................................. 49
Table 16. Evidence gaps for biopsy methods for the diagnosis of breast cancer ....................... 51

Figures
Figure A. Outcomes of testing in a hypothetical cohort of 1,000 women.............................. ES-13
Figure 1. Analytic framework ......................................................................................................... 5
Figure 2. Flow chart of included studies ....................................................................................... 14
Figure 3. Scatterplot of results in the receiver operating characteristic space and summary receiver operating characteristic curves of alternative core needle biopsy methods for the diagnosis of breast cancer ................................................................................. 19
Figure 4. Outcomes of testing in a hypothetical cohort of 1,000 women ..................................... 23
Figure 5. Positive predictive value of alternative biopsy methods ............................................. 24

Appendix
Appendix A. Search Strategy
Appendix B. Excluded Studies
Appendix C. Sensitivity Analysis to the Exclusion of High Risk Lesions on Core Needle Biopsy That Were Confirmed to be High Risk Lesions in Subsequent Open Biopsy or Surgical Excision
Appendix D. Assessment of the Strength of Evidence
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 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 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,” 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. 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 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. 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>
<thead>
<tr>
<th>Core Needle Biopsy Results (index test)</th>
<th>Reference Standard Results (open surgery or followup)</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Malignant (invasive or in situ)</td>
<td>Malignant (invasive or in situ)</td>
<td>considered TP</td>
<td>considered TP*</td>
</tr>
<tr>
<td>High risk lesion (e.g., ADH)</td>
<td>High risk lesion (e.g., ADH)</td>
<td>considered TP</td>
<td>considered FP</td>
</tr>
<tr>
<td>Benign</td>
<td>Benign</td>
<td>considered FN</td>
<td>considered TN</td>
</tr>
</tbody>
</table>

*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.\textsuperscript{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,\textsuperscript{16} the Cochrane Risk of Bias tool,\textsuperscript{17} and the checklist proposed by Drummond et al.,\textsuperscript{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.\textsuperscript{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.\textsuperscript{21} Furthermore, it allowed us to derive summary receiver operating characteristic (ROC) curves.\textsuperscript{22,23} A univariate mixed effects logistic regression model was used for the meta-analysis of rates of DCIS and high risk lesion underestimation.\textsuperscript{24} We used meta-regression methods to evaluate the impact of risk of bias items and other study-level characteristics.\textsuperscript{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. 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

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

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

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

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

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Freehand, automated</td>
<td>0 [0]</td>
<td>NA</td>
<td>1 [6]</td>
<td>0.88 (0.32 to 1.00)</td>
</tr>
<tr>
<td>US-guided, automated</td>
<td>14 [307]</td>
<td>0.38 (0.26 to 0.51)</td>
<td>21 [601]</td>
<td>0.25 (0.16 to 0.36)</td>
</tr>
<tr>
<td>US-guided, vacuum-assisted</td>
<td>5 [48]</td>
<td>0.09 (0.02 to 0.26)</td>
<td>9 [20]</td>
<td>0.11 (0.02 to 0.33)</td>
</tr>
<tr>
<td>Stereotactically guided, automated</td>
<td>18 [664]</td>
<td>0.26 (0.19 to 0.36)</td>
<td>29 [357]</td>
<td>0.47 (0.37 to 0.58)</td>
</tr>
<tr>
<td>Stereotactically guided, vacuum-assisted</td>
<td>34 [1899]</td>
<td>0.11 (0.08 to 0.14)</td>
<td>40 [1002]</td>
<td>0.18 (0.13 to 0.24)</td>
</tr>
<tr>
<td>MRI-guided, automated</td>
<td>0 [0]</td>
<td>NA</td>
<td>1 [1]</td>
<td>0.49 (0.02 to 0.97)</td>
</tr>
<tr>
<td>MRI-guided, vacuum-assisted</td>
<td>1 [1]</td>
<td>0.00 (0.00 to 0.38)</td>
<td>0 [0]</td>
<td>NA</td>
</tr>
<tr>
<td>Multiple methods/other</td>
<td>18 [628]</td>
<td>0.22 (0.15 to 0.30)</td>
<td>25 [866]</td>
<td>0.32 (0.23 to 0.41)</td>
</tr>
</tbody>
</table>

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 lesion) 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)

<table>
<thead>
<tr>
<th>Biopsy Methods Compared</th>
<th>Difference in Sensitivity (95% CrI)</th>
<th>Difference in Specificity (95% CrI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>US-guided, automated vs. vacuum-assisted</td>
<td>0.01 (-0.01, 0.06)</td>
<td>-0.01 (-0.03, 0.01)</td>
</tr>
<tr>
<td>Stereotactically guided, automated vs. vacuum-assisted</td>
<td>-0.02 (-0.04, -0.01)</td>
<td>0.05 (0.02, 0.08)</td>
</tr>
<tr>
<td>MRI-guided, automated vs. vacuum assisted</td>
<td>-0.10 (-0.43, -0.01)</td>
<td>0.07 (-0.03, 0.43)</td>
</tr>
</tbody>
</table>

CrI = credible interval; MRI = magnetic resonance imaging; US = ultrasound.

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.

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 conducted.
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: grey dashed-dotted = freehand; black solid = stereotactically guided, automated; grey solid = stereotactically guided, vacuum-assisted; black dotted = US-guided, automated; grey dotted = US-guided, vacuum-assisted; black dashed = MRI-guided, automated; grey 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 than 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

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Comparison or Biopsy Method</th>
<th>Overall Rating</th>
<th>Key Findings and Comments</th>
</tr>
</thead>
</table>
| 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] |
| Modifiers of test performance for women at average and high risk of breast cancer | All biopsy methods | Insufficient | – 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 Crls were wide; extreme odds ratio values were often observed because sensitivity and specificity for all tests were very close to 1 (see Results) |

Crl = 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

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Comparison or Biopsy Method</th>
<th>Overall Rating</th>
<th>Key Findings and Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>DCIS underestimation</td>
<td>Ultrasound-guided, automated</td>
<td>Low</td>
<td>– Average underestimation probability: 0.38 (0.26 to 0.51) [14 studies]</td>
</tr>
<tr>
<td></td>
<td>Ultrasound-guided, vacuum-assisted</td>
<td>Low</td>
<td>– Average underestimation probability: 0.09 (0.02 to 0.26) [5 studies]</td>
</tr>
<tr>
<td></td>
<td>Stereotactically guided, automated</td>
<td>Low</td>
<td>– Average underestimation probability: 0.26 (0.19 to 0.36) [18 studies]</td>
</tr>
<tr>
<td></td>
<td>Stereotactically guided, vacuum-assisted</td>
<td>Low</td>
<td>– Average underestimation probability: 0.11 (0.08 to 0.14) [34 studies]</td>
</tr>
<tr>
<td></td>
<td>Other biopsy methods</td>
<td>Insufficient</td>
<td>No available studies or few studies with small numbers of lesions</td>
</tr>
<tr>
<td>High risk lesion underestimation rate</td>
<td>Ultrasound-guided, automated</td>
<td>Low</td>
<td>– Average underestimation probability: 0.25 (0.16 to 0.36) [21 studies]</td>
</tr>
<tr>
<td></td>
<td>Ultrasound-guided, vacuum-assisted</td>
<td>Low</td>
<td>– Average underestimation probability: 0.11 (0.02 to 0.33) [9 studies]</td>
</tr>
<tr>
<td></td>
<td>Stereotactically guided, automated</td>
<td>Low</td>
<td>– Average underestimation probability: 0.47 (0.37 to 0.58) [29 studies]</td>
</tr>
<tr>
<td></td>
<td>Stereotactically guided, vacuum-assisted</td>
<td>Low</td>
<td>– Average underestimation probability: 0.18 (0.13 to 0.24) [40 studies]</td>
</tr>
<tr>
<td></td>
<td>Other biopsy methods</td>
<td>Insufficient</td>
<td>No available studies or few studies with small numbers of lesions</td>
</tr>
</tbody>
</table>

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>
<thead>
<tr>
<th>Outcomes</th>
<th>Test or Comparison</th>
<th>Overall Rating</th>
<th>Key Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bleeding (any severity)</td>
<td>Alternative core needle biopsy methods</td>
<td>Low</td>
<td>– Median %: 1.21 (25th perc. = 0.33; 75th perc = 3.97)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>– Selective outcome and analysis reporting likely</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>– Few studies reported bleeding requiring treatment; the event rate was low</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(&lt;0.40 perc.) in those studies</td>
</tr>
<tr>
<td>Bleeding events that require</td>
<td>Comparisons among alternative core needle biopsy</td>
<td>Low</td>
<td>– Median %: 0 (25th perc. = 0; 75th perc = 0.14)</td>
</tr>
<tr>
<td>treatment</td>
<td>methods</td>
<td></td>
<td>– Selective outcome and analysis reporting likely</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>– Few studies reported bleeding requiring treatment; the event rate was low</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(&lt;0.40 perc.) in those studies</td>
</tr>
<tr>
<td>Hematoma formation</td>
<td>Alternative core needle biopsy methods</td>
<td>Low</td>
<td>– Median %: 1.44 (25th perc. = 0.25; 75th perc = 8.57)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>– Selective outcome and analysis reporting likely</td>
</tr>
<tr>
<td>Infectious complications</td>
<td>Alternative core needle biopsy methods</td>
<td>Low</td>
<td>– Median %: 0 (25th perc. = 0; 75th perc = 0.33)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>– Selective outcome and analysis reporting likely</td>
</tr>
<tr>
<td>Vasovagal reactions:</td>
<td>Alternative core needle biopsy methods</td>
<td>Low</td>
<td>– Median %: 1.27 (25th perc. = 0.37; 75th perc = 3.88)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>– Potential for selective outcome and analysis reporting</td>
</tr>
<tr>
<td>Pain and severe pain</td>
<td>Alternative core needle biopsy methods</td>
<td>Low</td>
<td>25 studies of a wide variety of biopsy methods reported</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>information about patient pain during the procedure (pain was assessed</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>heterogeneously across studies)</td>
</tr>
<tr>
<td>Other adverse events</td>
<td>Alternative core needle biopsy methods</td>
<td>Insufficient</td>
<td>– Most events were reported by a single study precluding</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>assessment of consistency</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>– Individual studies did not provide adequate information for</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>precise estimation of the event rate</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>– Only informal indirect comparisons among biopsy methods were possible</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>– Selective outcome and analysis reporting likely</td>
</tr>
<tr>
<td></td>
<td>Sitting upright during the biopsy procedure</td>
<td>Low</td>
<td>Vasovagal reactions were more common among patients sitting during the</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>biopsy procedure</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Results were reported in few studies (11 studies; 8 from the</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>original evidence report and 3 from this update)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>– Selective outcome and analysis reporting likely</td>
</tr>
<tr>
<td></td>
<td>Vacuum-assisted versus non-vacuum assisted biopsy</td>
<td>Low</td>
<td>Vacuum-assisted procedures were generally associated with</td>
</tr>
<tr>
<td></td>
<td>methods</td>
<td></td>
<td>increased rates of bleeding and hematoma formation</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>– Bleeding events were generally uncommon</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>– Comparisons among biopsy methods were based on</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>informal indirect comparisons (across studies)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>– Selective outcome and analysis reporting likely</td>
</tr>
<tr>
<td></td>
<td>Comparisons among alternative core needle biopsy</td>
<td>Insufficient</td>
<td>– Most factors assessed by a single study limiting our ability to</td>
</tr>
<tr>
<td></td>
<td>methods</td>
<td></td>
<td>assess consistency</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>– Selective outcome and analysis reporting likely</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>– Within-study comparisons provided direct evidence</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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


Background

Breast Cancer Epidemiology and Clinical Diagnosis

Among women in the United States, breast cancer is the second most common malignancy (after skin cancer), and the second most common cause of cancer death (after lung cancer). Approximately one in eight women in the United States will develop breast cancer during their lifetime. The American Cancer Society estimates that 232,340 new cases of invasive breast cancer and 64,640 new cases of non-invasive breast cancer will be diagnosed in 2013, and 39,620 women will die of breast cancer.

During the earliest stages of breast cancer, there are usually no symptoms. The process of breast cancer diagnosis is initiated by detection of an abnormality through self-examination, physical examination by a clinician, or screening mammography. Data from the Behavioral Risk Factor Surveillance System show that, in 2010, 75.4 percent of U.S. women aged \( \geq 40 \) years and 79.7 percent of women aged 50 to 74 years reported having a mammogram within the past 2 years. If initial assessment suggests that the abnormality may be breast cancer, the woman may be referred for a biopsy, which is a sampling of cells or tissue from the suspicious lesion. In the United States, the most common indication for breast biopsy is the detection of suspicious abnormalities by screening mammography. Among women screened annually for 10 years, approximately 50 percent will need additional imaging, and 5-7 percent will have biopsies. Over a million women have breast biopsies each year in the United States. There are currently three techniques for obtaining samples from suspicious breast lesions: 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, and is used less frequently. For these reasons it will not be discussed further in this report. Core-needle biopsy, which retrieves a sample of tissue, and open surgical procedures are therefore the most frequently used biopsy methods.

Samples obtained by any of these methods are 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 often samples only part of the breast abnormality, there is the risk that a lesion will be classified as benign or as high risk (e.g., atypical ductal hyperplasia, [ADH]) or non-invasive (e.g., ductal carcinoma in situ, DCIS) when invasive cancer is in fact present in unsampled areas. In contrast, open surgical biopsy often samples most or all of the lesion, and it is thought that there is a smaller risk of misdiagnosis. However, while open surgical biopsy methods are considered to be the most accurate, they also appear to carry a higher risk of complications, such as bleeding or infection, compared to core needle biopsy. 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.

Core needle biopsy may be carried out using a range of techniques. If the breast lesion to be biopsied is not palpable, an imaging method (i.e., stereotactic mammography, ultrasound, or magnetic resonance imaging (MRI)) may be used to locate the lesion. The biopsy may be carried out with needles of varying diameter, and one or more samples of tissue may be taken. Sometimes a vacuum device is used to assist in removing the tissue sample through the needle. It is thought that these and other variations in how core needle biopsy is carried out may affect the accuracy and rate of complications of the biopsy. 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 aspects of biopsy technique on test performance and safety are not clear.

**Original Evidence Report and Rationale for the Update**

In 2009, the ECRI Evidence-based Practice Center (EPC) conducted a comparative effectiveness review for core needle versus open surgical biopsy.\(^5\)\(^6\) The original report provided a detailed description of the technical aspects of alternative biopsy methods and we have not repeated this information here. The original report assessed the diagnostic test performance and adverse events of multiple core needle biopsy techniques and tools, compared to open surgical biopsy, and also evaluated differences between open biopsy and core needle biopsy with regards to patient preference, costs, availability, and other factors. The key conclusions were 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. The need for update of the 2009 report was assessed in 2010 by the RAND EPC. Several high-impact general medical and specialty journals were searched, a panel of experts in the field was consulted, and an overall assessment of the need to update the report was produced. The conclusion of the update Surveillance Report was that additional studies and changes in practice render some conclusions of the original report possibly out of date. Specifically, the Surveillance Report noted the following:

- New studies are available regarding—
  - the DCIS underestimation rate of stereotactic vacuum-assisted core needle biopsy
  - test performance of MRI-guided core needle biopsy
  - test performance of freehand automated device core needle technology
- New studies on the test performance of core needle biopsy may allow the exploration of heterogeneity for test performance or harm outcomes

On the basis of the Surveillance Report findings, 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 biopsy.

**Key Questions**

To determine the Key Questions and study selection criteria (population, intervention, comparator, outcome, timing and setting; PICOTS) for this update, we began by considering the criteria used in the original Evidence Report. 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

This report updates a previously completed Comparative Effectiveness Review on core needle and open surgical biopsy methods for the diagnosis of breast cancer. To update the report we performed a systematic review of the published scientific literature using established methodologies as outlined in the Agency for Healthcare Research and Quality’s (AHRQ) “Methods Guide for Comparative Effectiveness Reviews,” which is available at: http://effectivehealthcare.ahrq.gov. The main sections in this chapter reflect the elements of the protocol that guided this review. We have followed the reporting requirements of the “Preferred Reporting Items for Systematic Reviews and Meta-analyses” (PRISMA) checklist. All key methodological decisions were made a priori. The protocol was developed with input from external clinical and methodological experts, in consultation with the AHRQ task order officer (TOO), and was posted online to solicit additional public comments. Its PROSPERO registration number is CRD42013005690.

AHRQ Task Order Officer

The AHRQ Task Order Officer (TOO) was responsible for overseeing all aspects of this project. The TOO facilitated a common understanding among all parties involved in the project, resolved ambiguities, and fielded all Evidence-based Practice Center (EPC) queries regarding the scope and processes of the project. The TOO and other staff at AHRQ helped to establish the Key Questions and protocol and reviewed the report for consistency, clarity, and to ensure that it conforms to AHRQ standards.

External Stakeholder Input

A new panel of experts was convened to form the Technical Expert Panel (TEP). The TEP included 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 review of evidence. Discussions among the EPC, TOO, and the TEP occurred during a series of teleconferences and via email.

Key Questions

The final Key Questions are listed at the end of the Background section. The refinement of the Key Questions took into account the patient populations, interventions, comparators, outcomes, and study designs that are clinically relevant for core needle biopsies.

Analytic Framework

We used an analytic framework (Figure 1) that maps the Key Questions within the context of populations, interventions, comparators, and outcomes of interest. The framework was adapted from that used in the original 2009 CER. It depicts the chain of logic that links the test performance of core needle biopsy for the diagnosis of breast abnormalities (Key Question 1) with patient-relevant outcomes (Key Question 3) and adverse events of testing (Key Question 2).
Scope of the Review

Populations and Conditions of Interest

The population of interest for all Key Questions was women who have been referred for biopsy for the diagnosis of primary breast cancer (including multifocal and bilateral disease) following self-examination, physical examination, or screening mammography. Studies carried out in women who had been previously diagnosed with breast cancer and were being examined for recurrence or to assess the extent of disease (staging) were excluded. The original report excluded studies carried out in women at high risk of breast cancer; however, MRI-guided biopsy is used mainly in this subset of patients. For this reason, following extensive discussions with the TEP, we decided to broaden the scope of the review to include studies carried out in women at high baseline risk of breast cancer (e.g., on the basis of BRCA genetic testing or family history of 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 either enrolled less than 15 percent of patients at high risk of cancer or did not provide information on baseline risk. Throughout this report, we refer to the latter group as “studies of women at average risk of cancer”; however, we acknowledge that it may include studies enrolling patients at higher-than-average cancer risk but failing to report the relevant information.

Interventions

For all Key Questions, the interventions of interest were core needle and open biopsy done to evaluate whether a breast lesion is malignant. Other uses of biopsy techniques (e.g., use of biopsy to examine the sentinel lymph nodes in women with an established diagnosis of breast cancer) were not considered. Studies were required to have used biopsy instrumentation that is currently commercially available, as studies of discontinued devices are not applicable to current practice.

Comparators (Index and Reference Standard Tests)

For test performance outcomes (Key Question 1) the reference standard was either open surgical biopsy, or followup by clinical examination and/or mammography for at least six
months. The diagnostic performance of each core biopsy technique (each index test) was quantified versus the reference standard. Most assessments of diagnostic performance quantify the sensitivity and the specificity of each index test – here each needle core biopsy technique. Sensitivity and specificity are probabilities conditional on true disease status, and are noncomparative in nature. The reference standard is used in their definition, and is not a “comparator test”. The comparative diagnostic performance of alternative needle core biopsy techniques was also evaluated. For adverse events and patient-relevant outcomes (outcomes other than diagnostic performance; Key Questions 2 and 3) the comparators of interest were: open surgical biopsy, followup by clinical examination and/or mammography for at least six months, or alternative core needle biopsy methods (e.g., stereotactic mammography versus ultrasound to locate the breast lesion; use versus non-use of vacuum-assistance to extract tissue samples).

Outcomes

For Key Question 1, the outcome of interest was test performance, as assessed by sensitivity (proportion of cancers detected by the reference standard that are also detected by core needle biopsy); specificity (proportion of negative findings according to core needle biopsy that were classified as negative by the reference standard; equal to one minus the false positive rate); underestimation rate for high risk lesions (most often atypical ductal hyperplasia, ADH), defined as the proportion of core needle biopsy findings of high risk lesions that are found to be malignant according to the reference standard); and underestimation rate for ductal carcinoma in situ (DCIS), defined as the proportion of core needle biopsy findings of DCIS that are found to be invasive according to the reference standard.

For Key Question 2 we looked for the following outcomes: rate of inconclusive biopsy findings (e.g. inadequate sampling of lesion); comparisons of repeat biopsy rates between core needle and open surgical biopsy; subsequent false positive and false negative rates on mammography (impact of breast biopsy on future mammographic examinations); dissemination or displacement of cancerous cells along the needle track; and patient-centered outcomes (including bruising, bleeding or hematomas, pain, use of pain medication, infections, fainting or near fainting, time to recover). Because adverse events were not consistently defined across studies, we accepted the definitions used in the individual studies (when available).

For Key Question 3, we considered patient-relevant outcomes (patient preferences for specific procedures, cosmetic results, quality of life, anxiety and other psychological outcomes, time to complete tumor removal [for women with cancer], recurrence rate [for women with cancer, including local, regional, and distant recurrence], cancer-free survival and overall survival); resource use and logistics (costs, resource utilization other than cost [number of additional surgical procedures, procedural time], subsequent surgical procedures, wait time for test results); and availability of technology and relevant expertise (physician experience, availability of equipment, availability of [qualified] pathologists to evaluate biopsy samples).

Timing

We required that the duration of clinical and/or mammography followup was at least six months in studies where open surgical biopsy was not performed.
Setting

Studies in all geographic locations and care settings were evaluated, including general hospitals, academic medical centers, and ambulatory surgical centers, among others.

Study Design and Additional Criteria

We required that studies had been published in peer-reviewed journals as full articles. For all Key Questions, studies were required to have been published in English. Restricting included studies to those published in English, which was also an inclusion criterion in the original review, was deemed unlikely to bias the results of the review and avoids the resource-intensive translation of research articles published in languages other than English.

For Key Question 1 eligible studies were prospective or retrospective cohort studies or randomized controlled trials. Retrospective case studies (“case series”) and other studies sampling patients on the basis of outcomes (e.g. diagnostic case-control studies, or studies selecting cases on the basis of specific histological findings) were excluded. Empirical evidence from meta-epidemiological studies suggests that diagnostic case-control studies may overestimate test performance. Studies were required to report information on the sensitivity, specificity, positive or negative predictive value of tests, or to include data that allow the calculation of one or more of these outcomes. Specifically, studies needed to provide adequate information to reconstruct 2×2 tables of test performance of the index against the reference standard. Table 1 illustrates how index and reference standard results were used to construct such 2×2 tables.

<table>
<thead>
<tr>
<th>Core Needle Biopsy Results (index test)</th>
<th>Reference Standard Results (open surgery or followup)</th>
<th>Benign</th>
</tr>
</thead>
<tbody>
<tr>
<td>Malignant (invasive or in situ)</td>
<td>considered TP</td>
<td>considered TP*</td>
</tr>
<tr>
<td>High risk lesion (e.g., ADH)</td>
<td>considered TP</td>
<td>considered FN</td>
</tr>
<tr>
<td>Benign</td>
<td>considered FN</td>
<td>considered TN</td>
</tr>
</tbody>
</table>

*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 non-malignant (high
risk or benign) on subsequent open biopsy or surgery were excluded from the analyses. Noncomparative studies of test performance (i.e. studies of a single index test) were required to have enrolled at least 10 participants per arm or per comparison group. This inclusion criterion was intended to reduce the risk of bias from non-representative participants in small studies. Further, smaller studies do not produce precise estimates of test performance and as such are unlikely to substantially affect results. Studies were also required to have followed at least fifty percent of participants to completion. This criterion was intended to reduce the risk of bias from high rates of attrition.

Key Question 2 was addressed by extracting harm-related information for core needle biopsy and open surgical biopsy from studies meeting the criteria for Key Question 1. In addition, we included studies that met all other selection criteria for Key Question 1 except for the use of a reference standard and the reporting of information on test performance outcomes. This allowed us to consider additional sources of evidence that assess adverse events. Finally, for this Key Question, we also reviewed primary research articles, regardless of design (i.e., case reports and case series, case-control studies, cohort studies, randomized trials), that address the dissemination or displacement of cancer cells by the biopsy procedure, a relatively rare harm that is specific to core biopsy.

The original report did not use formal criteria for study selection for Key Question 3. Based on the findings of the original report, we used the same PICOTS criteria described above and considered the following study designs:

- Randomized controlled trials, cohort studies, and cross-sectional studies on patient preferences, cosmetic results of biopsy procedures, physician experience (including studies of the “learning curve” for different biopsy methods and tools).
- Cost studies, including cost-minimization and cost-consequence analyses, were used to obtain information on resource utilization and unit costs. Given the large variability of cost information among different jurisdictions, we only considered studies conducted in the U.S. setting and published after 2004.10
- Cost-effectiveness/cost-utility analyses based on primary trials of breast biopsy interventions were used to obtain information on unit costs and resource utilization.11 Specifically, we considered the components of cost and resource use but did not use cost-effectiveness ratios or other summary measures of cost-effectiveness/utility. As for cost studies, we only considered primary cost-effectiveness/utility studies conducted in the US setting and published after 2004.10 We did not use model-based cost-effectiveness results.
- Studies of pathologist qualifications for interpreting core needle biopsy results; including interlaboratory initiatives to standardize diagnostic criteria (e.g., proficiency testing) or minimal competency requirements.
- Surveys of the availability of equipment for obtaining core needle biopsies and of qualified pathologists to examine biopsy samples.

**Literature Search and Abstract Screening**

We searched MEDLINE®, Embase®, the Cochrane Central Register of Controlled Trials (CENTRAL), the Cochrane Database of Systematic Reviews, the Database of Abstracts of Reviews of Effects (DARE), the Health Technology Assessment Database (HTA), the U.K. National Health Service Economic Evaluation Database (NHS EED), the U.S. National Guideline Clearinghouse (NGC), and the Cumulative Index to Nursing and Allied Health
Literature (CINAHL®); last search on December 16, 2013. Appendix A describes the search strategy we employed which is a revision and expansion of the search strategy used in the original report. Of note, the original report used a search filter for studies of diagnostic tests to increase search specificity; this is a reasonable approach given the large volume of literature on studies on diagnostic biopsy methods for breast cancer. Because this update covered a short time period (from 2009 to 2013) we opted to not use this filter, in order to increase search sensitivity.12, 13 Our searches covered the time period from six months before the most recent search date in the original report, to ensure adequate overlap.

To identify studies excluded from the original report because they enrolled women at high risk for cancer, the set of abstracts screened for the original report was obtained and rescreened for potentially eligible studies of high risk women. In addition, the list of studies excluded from the original report following full text review was checked to identify studies excluded because they included women at high risk for cancer. We also performed a search for systematic reviews on the topic and used their reference lists of included studies to validate our search strategy and to make sure we identified all relevant studies.

All reviewers screened a common set of 200 abstracts (in 2 pilot rounds, each with 100 abstracts), and discussed discrepancies, in order to standardize screening practices and ensure understanding of screening criteria. The remaining citations were split into nonoverlapping sets, each screened by two reviewers independently. Discrepancies were resolved by consensus involving a third investigator.

We asked the TEP to provide citations of potentially relevant articles. Additional studies were identified through the perusal of reference lists of eligible studies, published clinical practice guidelines, relevant narrative and systematic reviews, Scientific Information Packages from manufacturers, and a search of U.S. Food and Drug Administration databases. All articles identified through these sources were screened for eligibility against the same criteria as for articles identified through literature searches. We sent the final list of included studies to the TEP to ensure that no key publications had been missed.

**Study Selection and Eligibility Criteria**

Potentially eligible citations were obtained in full text and reviewed for eligibility on the basis of the predefined inclusion criteria. A single reviewer screened each potentially eligible article in full-text to determine eligibility; reviewers were instructed to be inclusive. A second reviewer verified all relevant articles. Disagreements regarding article eligibility were resolved by consensus involving a third reviewer. Appendix B lists all the studies excluded after full-text screening and the reason for exclusion.

**Data Abstraction and Management**

Data was extracted using electronic forms and entered into the Systematic Review Data Repository (SRDR; http://srdr.ahrq.gov/). The basic elements and design of these forms is similar to those we have used for other reviews of diagnostic tests and includes elements that address population characteristics, sample size, study design, descriptions of the index and reference standard tests of interest, analytic details, and outcome data. Prior to data extraction, forms were customized to capture all elements relevant to the Key Questions. We used separate sections in the extraction forms for Key Questions related to short-term outcomes, including classification of breast abnormalities, intermediate outcomes (such as clear surgical margins),
patient-relevant outcomes (such as quality of life), and factors affecting (modifying) test performance. 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. At least one other team member reviewed and confirmed all data (data verification). Disagreements were resolved by consensus including a third reviewer. We contacted authors (1) to clarify information reported in the papers that is hard to interpret (e.g., inconsistencies between tables and text); and (2) to verify suspected overlap between study populations in publications from the same group of investigators.

Assessment of the Risk of Bias of Individual Studies

We assessed the risk of bias for each individual study using the assessment methods detailed in the AHRQ Methods Guide for Effectiveness and Comparative Effectiveness Review hereafter referred to as the Methods Guide. We used elements from the Quality Assessment for Diagnostic Accuracy Studies instrument (QUADAS version 2), to assess the risk of bias (methodological quality or internal validity) of the diagnostic test studies included in the review (these studies comprise the majority of the available studies).14-17 The tool assesses four domains of risk of bias related to patient selection, index test, reference standard test, and patient flow and timing. For studies of other designs we used appropriate sets of items to assess risk of bias or methodological “quality”: for nonrandomized cohort studies we used items from the Newcastle-Ottawa scale,18 for randomized controlled trials we used items from the Cochrane Risk of Bias tool,19 and for studies of resource utilization and costs we used items from the checklist proposed by Drummond et al.20

We assessed and reported methodological quality items (as “Yes”, “No”, or “Unclear/Not Reported”) for each eligible study. We then rated each study as being of low, intermediate, or high risk of bias on the basis of adherence to accepted methodological principles. Generally, studies with low risk of bias have the following features: lowest likelihood of confounding due to comparison to a randomized controlled group; a clear description of the population, setting, interventions, and comparison groups; appropriate measurement of outcomes; appropriate statistical and analytic methods and reporting; no reporting inconsistencies; clear reporting of dropouts and a low dropout rate; and no other apparent sources of bias. Studies with moderate risk of bias are susceptible to some bias but not sufficiently to invalidate results. They do not meet all the criteria for low risk of bias owing to some deficiencies, but none are likely to introduce major bias. Studies with moderate risk of bias may not be randomized or may be missing information, making it difficult to assess limitations and potential problems. Studies with high risk of bias are those with indications of bias that may invalidate the reported findings (e.g., observational studies not adjusting for any confounders, studies using historical controls, or studies with very high dropout rates). These studies have serious errors in design, analysis, or reporting and contain discrepancies in reporting or have large amounts of missing information. We discuss the handling of high risk of bias studies in evidence synthesis in the following sections. Studies of different designs were graded within the context of their study design.

Data Synthesis

We summarized included studies qualitatively and presented important features of the study populations, designs, tests used, outcomes, and results in summary tables. Population characteristics of interest included age, race/ethnicity, and palpability of lesion. Design
characteristics included methods of population selection and sampling, and followup duration. Test characteristics included imaging-guided versus not imaging-guided, and vacuum-assisted versus not vacuum-assisted methods. We looked for information on test performance, adverse events, patient preferences, and resource utilization including costs.

Statistical analyses were conducted using methods currently recommend for use in Comparative Effectiveness Reviews of diagnostic tests.\textsuperscript{21, 22} For all outcomes we assessed heterogeneity graphically (e.g. by inspecting a scatterplot of studies in the receiver operating characteristic, ROC, space) and by examining the posterior distribution of between-study variance parameters.

For Key Question 1 we performed meta-analysis on studies that were deemed sufficiently similar. Based on the technical characteristics of the different tests, and the findings of the original Evidence Report, we developed a mixed effects binomial-bivariate normal regression model that accounted for different imaging methods (e.g. US, stereotactic mammography, MRI), the use of vacuum (yes vs. not), the baseline of risk of cancer of included patients (high versus average risk), and residual (unexplained) heterogeneity.\textsuperscript{23-25} This model allowed us to estimate the test performance of alternative diagnostic tests, and perform indirect comparisons among them.\textsuperscript{23} Furthermore, it allowed us to model the correlation between sensitivity and specificity and to derive meta-analytic ROC curves.\textsuperscript{24, 25} A univariate mixed effects logistic regression (binomial-normal) model was used for the meta-analysis of DCIS and high risk lesion underestimation rates.\textsuperscript{26}

We performed meta-regression analyses (e.g. to evaluate the impact of study risk of bias items, or the effect of other study-level characteristics) by extending the model to include additional appropriately coded terms in the regression equations.\textsuperscript{27, 28} Such analyses were planned for patient and breast lesion factors (e.g., age, density of breast tissue, microcalcifications, and palpability of the lesions), biopsy procedure factors (e.g., needle size, imaging guidance, vacuum extraction, and number of samples), clinician and facility-related factors (e.g., training of the operator, country were the study was conducted), and risk of bias items. We performed additional sensitivity analyses (e.g., leave-one-out meta-analysis and comparisons of studies added in the update versus studies included in the original report).\textsuperscript{29}

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, 25\textsuperscript{th} and 75\textsuperscript{th} 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, for all outcomes except the number of surgical procedures, we did not perform meta-analysis but instead chose to summarize the data qualitatively. We performed 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 procedures for treatment, after the establishment of breast cancer diagnosis. This analysis used a standard univariate normal random effects model with a binomial distribution for the within-study likelihood of each biopsy group (core needle vs. open).

All statistical analyses were performed using Bayesian methods; models were fit using Markov Chain Monte Carlo methods and noninformative prior distributions. Theory and empirical comparisons 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).\textsuperscript{30} 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% probability.

Grading the Strength of Evidence

We followed the Methods Guide\textsuperscript{7} 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.\textsuperscript{7,31} 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, selective outcome reporting, and selective 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.\textsuperscript{32,33} 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.\textsuperscript{7} These describe our level of confidence that the evidence reflects the true effect for the major comparisons of interest.

Assessing Applicability

We followed the Methods Guide\textsuperscript{7} in evaluating the applicability of included studies to patient populations of interest. Applicability to the population of interest was also judged separately on the basis of patient characteristics (e.g., age may affect test performance because the consistency of the breast tissue changes over time), method by which suspicion is established (e.g., mammography vs. other methods may affect test performance through spectrum effects), baseline risk of cancer (“average risk” vs. “high risk” women may affect estimated test performance because of differences in diagnostic algorithms), outcomes (e.g., prevalence of breast cancers diagnosed upon biopsy may also be a marker of spectrum effects), and setting of care (because differences in patient populations, diagnostic algorithms, and available technologies may affect test results).
Peer Review

The initial draft report was pre-reviewed by the TOO and an AHRQ Associate Editor (a senior member of another EPC). Following revisions, the draft report was sent to invited peer reviewers and was simultaneously uploaded to the AHRQ Web site where it was available for public comment for 30 days. All reviewer comments (both invited and from the public) were collated and individually addressed. The revised report and the EPC’s responses to invited and public reviewers’ comments were again reviewed by the TOO and Associate Editor prior to completion of the report. The authors of the report had final discretion as to how the report was revised based on the reviewer comments, with oversight by the TOO and Associate Editor.
Results

Our literature searches identified 8,637 potentially relevant citations (including 1,127 rescreened from the original ECRI evidence report). After review of the abstracts, 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, for a total of 316 included studies. Figure 2 presents the literature flow and Table 2 summarizes the additions to the original report, separately by Key Question.

Figure 2. Flow chart of included studies

Citations retrieved from MEDLINE, Embase, Cochrane Central Register of Controlled Clinical Trials, CINAHL, Guidelines Clearinghouse (7,510 publications)

- From review of SIP (0 publications)
- From hand search of reference lists (0 publications)
- From ECRI files of searches (1,127 publications)

- From ECRI report (208 publications)

Full text articles retrieved (2,272 publications)

- Excluded (2,164 publications):
  - Duplicate references
  - Full text not in English
  - N<10, no seeding
  - Less than 50% follow-up
  - Reference standard incomplete
  - More than 15% of women have previous or current cancer
  - Used a core-needle instrument that is no longer commercially available
  - Case control or retrospective case study
  - No primary data (commentary, narrative review, proposed procedure, letter)
  - No CNB or CNB not for diagnosis of primary breast cancer in women
  - Cases selected on the basis of CNB or final outcomes
  - No outcomes of interest (no KQ1, KQ2, KQ3)

- Full-text articles included (316 studies)

KQ1 161 studies
KQ2 144 studies
KQ3 143 studies

CNB = core needle biopsy; KQ = Key Question; N = number of patients; SIP = Scientific Information Packet.
**Table 2. Summary of new evidence evaluated in this update**

<table>
<thead>
<tr>
<th>Key Question</th>
<th>Studies Included in the Original Report</th>
<th>Studies Identified by the Updating Process</th>
<th>Total Number of Studies Synthesized in This Report</th>
</tr>
</thead>
<tbody>
<tr>
<td>Key Question 1: What is the test performance of different types of core needle breast biopsy compared with open biopsy in the diagnosis of breast cancer?</td>
<td>107**</td>
<td>54</td>
<td>161</td>
</tr>
<tr>
<td>Key Question 2: What are the adverse events (harms) associated with core needle breast biopsy compared to the open biopsy in the diagnosis of breast cancer?</td>
<td>74</td>
<td>70</td>
<td>144</td>
</tr>
<tr>
<td>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?</td>
<td>84</td>
<td>59</td>
<td>143</td>
</tr>
</tbody>
</table>

*Some studies addressed multiple Key Questions

** The original ECRI report included a total of 108 studies; one core needle biopsy study overlapped with one of the studies identified in our update which enrolled a larger patient population; thus the smaller study was excluded from Key Question 1.
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?

Included Studies

Fifty-four new studies identified by this update met the inclusion criteria for Key Question 1. We synthesized these studies with the 106 studies identified by the original evidence report, for a total of 160 studies providing information on test performance outcomes. Studies had been published between 1990 and 2013. Fifty studies were prospectively designed, and 58 were conducted in the United States. Ten studies provided information on more than one group of patients (typically undergoing biopsy with a different biopsy device). In statistical analyses these groups were treated as separate strata, leading to a total of 171 complete $2 \times 2$ tables of diagnostic test results, with information on 69,804 breast lesions.

Test Performance for Breast Cancer Diagnosis

Test Performance of Open Surgical Biopsy

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 identified by the original evidence report, provided information on the test performance of open surgical biopsy, using published literature and primary patient data (patient charts) from patients evaluated at a single medical center. Based on a re-review of archived open biopsy material by a second pathologist, patient chart review, study of cases with benign results on biopsy after suspicious mammography results, and expert opinion, the authors concluded that open surgical biopsy may miss one to two percent of breast cancers (i.e. sensitivity of 98% or greater). We found a single clinical study of patients undergoing surgical biopsies who were followed by imaging for 12 months after biopsy. This study reported underestimation in 16.7 percent of ADH lesions (1 of 6 lesions) and 7.1 percent of DCIS lesions (1/14 lesions) 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. Six studies enrolled women at high risk of cancer development and 154 enrolled women at average risk. The studies reported on a variety of biopsy techniques: 131 study arms reported on the use of a single form of imaging guidance (83 stereotactic; 41 ultrasound; 6 MRI; 1 grid) whereas ten used freehand methods, 29 used multiple methods, including freehand techniques in some cases (and did not report test performance results separately by each method), 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 size also varied across studies: 57 used 14G needles, nine used smaller and 46 used larger bores; 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 follow up 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. Details about study design, selection criteria, enrolled populations, biopsy methods and results, are publically available in the SRDR. Consistent with the findings of the original report, the overall risk of bias 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. Detailed results from our risk of bias assessment are provided at the end of this section.

The prevalence of malignant disease (invasive 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 3 summarizes the results for alternative diagnostic biopsy methods, together with information on the number of lesions evaluated by each test and summary test performance information, for women at average risk of cancer. Table 4 summarizes the same information for women deemed to be at high risk for cancer (e.g. due to genetic factors or strong family history). Figure 3 presents individual study estimates and meta-analytic results in the ROC space for both groups of women. These plots indicate that results were fairly homogeneous across studies for each test and that test sensitivity and specificity were close to 1 (studies cluster at the top left corner of the space). In analyses excluding high risk lesions on core needle biopsy that were also classified as high risk lesions on the reference standard test, the summary sensitivity of the various tests was unaffected; and summary specificity was somewhat increased (Appendix C).

Key findings with respect to test performance (sensitivity, specificity, and positive and negative likelihood ratios\(^*\)) and underestimation rates are summarized narratively below. As mentioned, only six studies reported results on the test performance of various biopsy methods for breast cancer diagnosis in high risk women. Of these studies, only three reported information on underestimation rates (all three for high risk lesions; none for DCIS).

\[^{*}\text{To aid in the interpretation of likelihood ratios we remind readers that these statistics can be used to convert pre-test probabilities to post-test probabilities. For example, before testing, assume that a patient has probability of disease}
\]

\[
\text{pre-test } p = 0.1 \text{ and pre-test odds } = \frac{\text{pre-test } p}{1 - \text{pre-test } p} = \frac{0.1}{0.9} = 0.11. \text{ If the diagnostic test has a positive likelihood ratio (LR)}^+ \text{ of 15 then the post-test odds are}
\]

\[
\text{post-test odds} = \frac{\text{pre-test odds} \times LR^+}{1} = 0.11 \times 15 = 1.67. \text{ This corresponds to a}
\]

\[
\text{post-test probability of } = \frac{\text{post-test odds}}{1 + \text{post-test odds}} = \frac{1.67}{1 + 1.67} = 0.625 \text{(i.e. the post-test probability is approximately 6}
\]

times greater than the pre-test value). If the test results had been negative and the test had a negative likelihood ratio (LR\(^-\)) of 0.1, the post-tests odds would be

\[
\text{post-test odds} = \frac{\text{pre-test odds} \times LR^-}{1} = 0.11 \times 0.1 = 0.011, \text{ which corresponds to post-test } p = \frac{0.011}{0.011 + 1} = 0.011
\]

(i.e. approximately the post-test probability is approximately 10 times lower than the pre-test value). As a rule of thumb, \(LR^+ > 10\) and \(LR^- < 0.1\) are generally considered clinically meaningful.
Table 3. Summary estimates of test performance for alternative core needle biopsy methods – women at average risk of cancer

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Freehand, automated</td>
<td>10 [786]</td>
<td>0.91 (0.80 to 0.96)</td>
<td>0.98 (0.95 to 1.00)</td>
<td>0 [0]</td>
<td>NA</td>
<td>1 [6]</td>
<td>0.88 (0.32 to 1.00)</td>
</tr>
<tr>
<td>US-guided, automated</td>
<td>27 [16287]</td>
<td>0.99 (0.98 to 0.99)</td>
<td>0.97 (0.95 to 0.98)</td>
<td>14 [307]</td>
<td>0.38 (0.26 to 0.51)</td>
<td>21 [601]</td>
<td>0.25 (0.16 to 0.36)</td>
</tr>
<tr>
<td>US-guided, vacuum-assisted</td>
<td>12 [1543]</td>
<td>0.97 (0.92 to 0.99)</td>
<td>0.98 (0.96 to 0.99)</td>
<td>5 [48]</td>
<td>0.09 (0.02 to 0.26)</td>
<td>9 [20]</td>
<td>0.11 (0.02 to 0.33)</td>
</tr>
<tr>
<td>Stereotactically guided, automated</td>
<td>37 [9535]</td>
<td>0.97 (0.95 to 0.98)</td>
<td>0.97 (0.96 to 0.98)</td>
<td>18 [664]</td>
<td>0.26 (0.19 to 0.36)</td>
<td>29 [357]</td>
<td>0.47 (0.37 to 0.58)</td>
</tr>
<tr>
<td>Stereotactically guided, vacuum-assisted</td>
<td>43 [14667]</td>
<td>0.99 (0.98 to 0.99)</td>
<td>0.92 (0.89 to 0.94)</td>
<td>34 [1899]</td>
<td>0.11 (0.08 to 0.14)</td>
<td>40 [1002]</td>
<td>0.18 (0.13 to 0.24)</td>
</tr>
<tr>
<td>MRI-guided, automated</td>
<td>2 [89]</td>
<td>0.90 (0.57 to 1.00)</td>
<td>0.99 (0.91 to 1.00)</td>
<td>0 [0]</td>
<td>NA</td>
<td>1 [1]</td>
<td>0.49 (0.02 to 0.97)</td>
</tr>
<tr>
<td>MRI-guided, vacuum-assisted</td>
<td>1 [10]</td>
<td>1.00 (0.98 to 1.00)</td>
<td>0.91 (0.54 to 0.99)</td>
<td>1 [1]</td>
<td>0.00 (0.00 to 0.38)</td>
<td>0 [0]</td>
<td>NA</td>
</tr>
<tr>
<td>Multiple methods/other</td>
<td>33 [26028]</td>
<td>0.99 (0.98 to 0.99)</td>
<td>0.96 (0.93 to 0.97)</td>
<td>18 [628]</td>
<td>0.22 (0.15 to 0.30)</td>
<td>25 [866]</td>
<td>0.32 (0.23 to 0.41)</td>
</tr>
</tbody>
</table>

All numbers are medians with 95% CrIs, unless otherwise stated.
CrI = credible interval; DCIS = ductal carcinoma in situ; MRI = magnetic resonance imaging; N = number; NA = not applicable; US = ultrasound.

Table 4. Summary estimates of test performance for alternative core needle biopsy methods – women at high risk of cancer

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

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 bivariate model with risk group as a covariate.
CrI = credible interval; DCIS = ductal carcinoma in situ; MRI = magnetic resonance imaging; N = number; US = ultrasound.
Figure 3. Scatterplot of results in the receiver operating characteristic space and summary receiver operating characteristic curves of alternative core needle biopsy methods for the diagnosis of breast cancer

Solid black lines represent results for average risk women; dashed gray lines represent results for high risk women (when studies were available). The numbers of observations and studies include cohorts of women both at average and high risk of cancer. The number of observations reflects the total number of data points in the analysis; some studies contributed patient-level and others lesion-level results. MRI = magnetic resonance imaging; US = ultrasound.
Freehand Core Needle Biopsies

Women at average risk of cancer: Ten cohorts reported data on the accuracy of nonguided (i.e., freehand) core needle biopsies performed with automated biopsy devices. The summary sensitivity was 0.91 (95% CrI, 0.80 to 0.96) and the summary specificity was 0.98 (95% CrI, 0.95 to 1.00), corresponding to a positive likelihood ratio of 58.4 (95% CrI, 19.0 to 226.9) and a negative likelihood ratio of 0.09 (95% CrI, 0.04 to 0.20). Only one study provided information on the high risk lesion underestimation rate (five cancers misclassified as high risk lesions among a total of six such lesions on core needle biopsy). No studies provided information on the DCIS underestimation rate.

Women at high risk of cancer: No studies provided information on the test performance (sensitivity, specificity, or underestimation rates) of freehand core needle biopsy techniques in women at high risk of breast cancer.

Ultrasound Guided Automated Device Core Needle Biopsies

Women at average risk of cancer: Twenty-seven cohorts of 16,287 biopsies used ultrasound guidance and an automated biopsy device. The summary sensitivity was 0.99 (95% CrI, 0.98 to 0.99) and the summary specificity was 0.97 (95% CrI, 0.95 to 0.98), corresponding to a positive likelihood ratio of 33.5 (95% CrI, 20.7 to 56.9) and a negative likelihood ratio of 0.01 (95% CrI, 0.01 to 0.02). Fourteen studies provided information on the DCIS underestimation; the summary rate was 0.38 (95% CrI, 0.26 to 0.51). Twenty-one studies provided information on high risk lesion underestimation; the summary rate was 0.25 (95% CrI, 0.16 to 0.36).

Women at high risk of cancer: No studies provided information on the test performance (sensitivity, specificity, or underestimation rates) of ultrasound-guided automated core needle biopsy techniques in women at high risk of breast cancer.

Ultrasound Guided Vacuum-Assisted Core Needle Biopsies

Women at average risk of cancer: Twelve cohorts of 1,543 biopsies used ultrasound guidance and a vacuum-assisted device to perform breast biopsies. The summary sensitivity was 0.97 (95% CrI, 0.92 to 0.99) and the summary specificity was 0.98 (95% CrI, 0.96 to 0.99), corresponding to a positive likelihood ratio of 57.7 (95% CrI, 25.8 to 138.7) and a negative likelihood ratio of 0.03 (95% CrI, 0.01 to 0.08). Five studies provided information on DCIS underestimation; the summary rate was 0.09 (95% CrI, 0.02 to 0.26). Nine studies provided information on high risk lesion underestimation: the summary rate was 0.11 (95% CrI, 0.02 to 0.33).

Women at high risk of cancer: No studies provided information on the test performance (sensitivity, specificity, or underestimation rates) of ultrasound-guided automated core needle biopsy techniques in women at high risk of breast cancer.

Stereotactically Guided Automated Device Core Needle Biopsies

Women at average risk of cancer: Thirty-seven cohorts of 9,535 biopsies used stereotactic guidance and an automated biopsy device. The summary sensitivity was 0.97 (95% CrI, 0.95 to 0.98) and the summary specificity was 0.97 (95% CrI, 0.96 to 0.98), corresponding to a positive likelihood ratio of 33.6 (95% CrI, 22.6 to 50.9) and a negative likelihood ratio of 0.03 (95% CrI, 0.02 to 0.05). Eighteen studies provided information on DCIS underestimation; the summary rate...
was 0.26 (95% CrI, 0.19 to 0.36). Twenty-nine cohorts provided information on high risk lesion underestimation; the summary rate was 0.47 (95% CrI, 0.37 to 0.58).

**Women at high risk of cancer**: One study reported information on the test performance of stereotactically guided automated core needle biopsy methods. Using model-based results, sensitivity was 0.97 (0.82 to 1.00) and specificity was 0.97 (0.91 to 0.99). No studies provided information on the underestimation rate of stereotactically guided automated core needle biopsy methods in women at high risk of breast cancer.

**Stereotactically Guided Vacuum-Assisted Core Needle Biopsies**

**Women at average risk of cancer**: Forty-three cohorts of 14,667 biopsies used stereotactic guidance and a vacuum-assisted device to perform core needle biopsies. The summary sensitivity was 0.99 (95% CrI, 0.98 to 0.99) and the summary specificity was 0.92 (95% CrI, 0.89 to 0.94), corresponding to a positive likelihood ratio of 12.8 (95% CrI, 9.4 to 17.9) and a negative likelihood ratio of 0.01 (95% CrI, 0.01 to 0.02). Thirty-four studies provided information on DCIS underestimation; the summary rate was 0.11 (95% CrI, 0.08 to 0.14). Forty studies provided information on high risk lesion underestimation; the summary underestimation rate was 0.18 (95% CrI, 0.13 to 0.24).

**Women at high risk of cancer**: Two studies provided information on the test performance of stereotactically guided vacuum assisted core needle biopsies. The summary sensitivity was 0.99 (95% CrI 0.93 to 1.00) and summary specificity was 0.93 (0.79 to 0.98). One of the two studies also reported that two cancer cases were underestimated by the biopsy diagnosis, among a total of 17 high risk lesions (for an underestimation rate of 12%).

**MRI-Guided Automated Core Needle Biopsies**

**Women at average risk of cancer**: Two cohorts reported data on the accuracy of MRI-guided biopsies performed with automated biopsy devices. The summary sensitivity was 0.90 (95% CrI, 0.58 to 0.99) and the summary specificity was 0.99 (95% CrI, 0.92 to 1.00), corresponding to a positive likelihood ratio of 62.3 (95% CrI, 9.4 to 726.3) and a negative likelihood ratio of 0.10 (95% CrI, 0.01 to 0.44). None of the studies provided information on the DCIS underestimation rate. One study provided information on the high risk lesion underestimation rate (one biopsy-detected high risk lesion was found to be malignant).

**Women at high risk of cancer**: Two cohorts provided information on the test performance of MRI-guided automated core needle biopsies among women at high risk for cancer. The summary sensitivity was 0.90 (95% CrI 0.57 to 0.99) and summary specificity was 0.99 (0.91 to 1.00). One of the two studies also reported that no cancers were diagnosed in the two women considered to have high risk lesions on core needle biopsy (i.e. no underestimation was observed in the study).

**MRI-Guided Vacuum-Assisted Core Needle Biopsies**

**Women at average risk of cancer**: One cohort reported data on the accuracy of MRI-guided biopsies performed with automated biopsy devices. All malignant lesions (n=3) were identified on pathologic examination of biopsy samples (model-based sensitivity 1.00, 95% CrI, 0.98 to 1.00); none of the nonmalignant lesions (n=7) were false positives (model-based specificity 0.91, CrI, 0.54 to 0.99). No cancer was diagnosed in the one woman considered to have DCIS on core needle biopsy (i.e. no DCIS underestimation was observed in the study). The study did not provide information on the high risk lesion underestimation rate.
Women at high risk of cancer: One cohort provided information on the test performance of MRI-guided vacuum-assisted core needle biopsies among women at high risk for cancer. The model-based sensitivity was 1.00 (95% CrI 0.98 to 1.00) and summary specificity was 0.92 (0.61 to 0.99). The study also reported that no cancers developed in the seven women considered to have high risk lesions on core needle biopsy (i.e. no underestimation was observed in the study).

Populations Biopsied with Multiple Core Needle Methods (or Other Methods)

Women at average risk of cancer: An additional 33 cohorts reported results from populations of women undergoing core needle biopsy with diverse methods. The majority of these studies (31 of 33) did not stratify their results by biopsy method (with respect to imaging guidance or use of vacuum); this group also included one study using grid guidance, and one study that did not report information on the use of vacuum assistance. In this heterogeneous group of studies, the summary sensitivity was 0.99 (95% CrI, 0.98 to 0.99) and the summary specificity was 0.96 (95% CrI, 0.93 to 0.97), corresponding to a positive likelihood ratio of 22.2 (95% CrI, 15.1 to 32.9) and a negative likelihood ratio of 0.01 (95% CrI, 0.01 to 0.02). Eighteen studies provided information on the DCIS underestimation; the summary DCIS underestimation rate was 0.22 (95% CrI, 0.15 to 0.30). Twenty-five studies provided information on high risk lesion underestimation; the summary underestimation rate was 0.32 (95% CrI, 0.23 to 0.41).

Women at high risk of cancer: No studies of high risk women were included in this subgroup.

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 five 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 4 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 U.S.-guided methods, and MRI-guided vacuum assisted methods are expected to have fewer than ten false negative results (for every 1,000 women undergoing biopsy), even when prevalence is as high as 0.30.

To illustrate the dependence of the number of true positive results among patients who are test positive by breast biopsy on the prevalence of disease, we calculated positive predictive values over a range of prevalences for different biopsy methods (Figure 5). These results suggest that even in low breast cancer prevalence settings (of 5 to 10 percent), 70 to 80 percent of women who test positive will truly have breast cancer for all tests except stereotactically and MRI-guided, vacuum-assisted biopsy methods. These two methods are associated with
somewhat lower positive predictive values (approximately 50 to 60 percent) in low-prevalence settings, reflecting their lower specificity (compared to other tests). However, as the prevalence increases, the positive predictive value approaches 1 for all tests.

The above comparisons (test outcomes in a hypothetical cohort of known prevalence and positive predictive value calculations) can serve as aids for contextualizing the test performance meta-analysis results presented above. However, they do not reflect the uncertainty around the meta-analytic summary estimates. The following section presents the results of formal (indirect) comparisons among alternative core needle biopsy methods.

Figure 4. Outcomes of testing in a hypothetical cohort of 1,000 women

Lines correspond to different test modalities: grey dashed-dotted = freehand; black solid = stereotactically guided, automated; grey solid = stereotactically guided, vacuum-assisted; black dotted = US-guided, automated; grey dotted = U.S.-guided, vacuum-assisted; black dashed = MRI-guided, automated; grey dashed = MRI-guided, vacuum assisted.
Comparative Test Performance

To compare test performance across different biopsy methods we used indirect (meta-regression-based) comparisons. Table 5 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 percent 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.
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 populations. Overall, on the basis of both sources of information (within-study analyses and meta-regression analyses), we found that evidence was insufficient to support any specific factor as a modifier of test performance. Detailed results are presented below.

Within-Study Evidence

Twenty studies (14 identified by the original review and 6 included in the current update) of 11,280 patients provided information on factors that affect 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). The majority of studies (140 of 160) did not allow investigation of the impact of any factors on test performance. The 20 included studies were reported inconsistently and often lacked details necessary for formal statistical assessment of the impact of various factors on test performance. These findings raise concerns regarding selective analysis and outcome reporting with respect to modifiers of test performance. Table 6 summarizes the findings of individual studies.

---

Table 5. Differences in sensitivity between pairs of biopsy methods (meta-regression based indirect comparisons)

<table>
<thead>
<tr>
<th>Biopsy Methods Compared</th>
<th>Difference in Sensitivity</th>
<th>Difference in Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>US-guided, automated vs. vacuum-assisted</td>
<td>0.01 (-0.01, 0.06)</td>
<td>-0.01 (-0.03, 0.01)</td>
</tr>
<tr>
<td>Stereotactically guided, automated vs. vacuum-assisted</td>
<td>-0.02 (-0.04, -0.01)</td>
<td>0.05 (0.02, 0.08)</td>
</tr>
<tr>
<td>MRI-guided, automated vs. vacuum assisted</td>
<td>-0.10 (-0.43, -0.01)</td>
<td>0.07 (-0.03, 0.43)</td>
</tr>
</tbody>
</table>

All results are shown as medians of differences (95% CrI). Positive values denote that the method on the first-listed biopsy method has higher performance that 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.
<table>
<thead>
<tr>
<th>Author, Year</th>
<th>Biopsy Method</th>
<th>Factors Evaluated</th>
<th>Key Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cusick et al., 1990 [2183373]</td>
<td>Freehand</td>
<td>Lesion size</td>
<td>Smaller lesions (&lt;2 cm in diameter) were more likely to be misdiagnosed.</td>
</tr>
<tr>
<td>Barreto et al. 1991 [2044776]</td>
<td>Freehand</td>
<td>Lesion size; lesion location; Patient age</td>
<td>Tumor size did not affect the accuracy of the procedure. All lesions in the study were &gt; 2 cm in diameter. Lesions in the right breast were more likely to be misdiagnosed. Patient age was not related to accuracy.</td>
</tr>
<tr>
<td>Makkun et al. 2011 [no PMID]</td>
<td>Ultrasound-guided automated device</td>
<td>Lesion type</td>
<td>The accuracy of biopsy in palpable lesions was 100%, while the accuracy of biopsy in nonpalpable lesions was 79.16%</td>
</tr>
<tr>
<td>Povoski et al. 2011 [21835024]</td>
<td>Ultrasound-guided automated device</td>
<td>Size of lesion; BI-RADS classification;</td>
<td>There was no difference between the median size of the lesion in cases of false negative biopsies and the median size in all cases biopsied. Among women undergoing interval followup after biopsy, the rate of false negatives was 0% for lesions initially classified as BI-RADS 3, 0.6% for lesions classified BI-RADS 4, and 2.8% for lesions classified BI-RADS 5.</td>
</tr>
<tr>
<td>Wiratkapun et al. 2012 [22252182]</td>
<td>Ultrasound-guided automated device</td>
<td>Patient age; breast density; lesion type; BI-RADS classification; lesion location</td>
<td>There was no statistically significant relationship between underestimation and patient age, breast density, lesion size, lesion visibility on mammography, lesion type (pure mass vs. mass with calcification), lesion BI-RADS classification (4 vs. 5). Lesions in the lower outer quadrant of the breast were more often underestimated. There was a tendency for younger women with larger mass lesions located at the lower quadrants of the breast and with BI-RADS 5 lesions not seen on mammography to have underestimated lesions.</td>
</tr>
<tr>
<td>Dahlstrom et al. 1996 [8735717]</td>
<td>Stereotactically guided automated device</td>
<td>Lesion type</td>
<td>There was no difference in the number of cores needed for diagnosis of microcalcifications, densities, or stellate lesions.</td>
</tr>
<tr>
<td>Koskela et al. 2005 [16020555]</td>
<td>Stereotactically guided automated device</td>
<td>Lesion type</td>
<td>There were zero false-negatives out of 97 procedures performed on lesions detected as masses on mammography, but 4 false-negatives out of 108 procedures performed on lesions with microcalcifications.</td>
</tr>
<tr>
<td>Walker et al. 1997 [no PMID]</td>
<td>Stereotactically guided, vacuum-assisted</td>
<td>Lesion type</td>
<td>The sensitivity of core needle biopsy was much lower for microcalcifications than for any other type of lesion.</td>
</tr>
<tr>
<td>Lomoschitz et al. 2004 [15273332]</td>
<td>Stereotactically guided, vacuum-assisted</td>
<td>Lesion type</td>
<td>Biopsies were equally accurate for lesions with microcalcifications and lesions detected as masses on mammography.</td>
</tr>
<tr>
<td>Pfarl et al. 2002 [12438044]</td>
<td>Stereotactically guided, vacuum-assisted</td>
<td>Lesion type</td>
<td>Biopsies were equally accurate for lesions with microcalcifications and lesions detected as masses on mammography.</td>
</tr>
<tr>
<td>Author, Year</td>
<td>Biopsy Method</td>
<td>Factors Evaluated</td>
<td>Key Findings</td>
</tr>
<tr>
<td>--------------</td>
<td>---------------</td>
<td>------------------</td>
<td>--------------</td>
</tr>
<tr>
<td>Reiner et al. 2009 [19565246]</td>
<td>Stereotactically guided, vacuum-assisted</td>
<td>Lesion type</td>
<td>The agreement rate between core needle biopsy and surgery was higher in nonpapillary lesions than in papillary lesions, but this difference was not statistically significant. 5 of 6 cases of underestimation occurred in papillary lesions, and the one false negative occurred in a nonpapillary lesion.</td>
</tr>
<tr>
<td>Venkataraman et al. 2012 [22127375]</td>
<td>Stereotactically guided, vacuum-assisted</td>
<td>Patient age; lesion size</td>
<td>There was no correlation between patient age and upgrade. However, there was a positive correlation between size of the lesion and upgrade.</td>
</tr>
<tr>
<td>Abdsoleh et al. 2003 [12630998]</td>
<td>Multiple methods</td>
<td>Patient breast density</td>
<td>Technical failures were more likely to occur in women with very dense breast tissue.</td>
</tr>
<tr>
<td>Ciatto et al. 2007 [16823506]</td>
<td>Multiple methods</td>
<td>Lesion type</td>
<td>False negative results were 2.7% for palpable lesions, 2.2% for nonpalpable lesions, 2.3% for masses on mammography, 1.4% for distortions on mammography, and 2.5% for microcalcifications.</td>
</tr>
<tr>
<td>Cipolla et al. 2006 [16473738]</td>
<td>Multiple methods</td>
<td>Lesion type</td>
<td>Correspondence between core needle biopsy and surgical biopsy results was 100% for palpable lesions but only 88% for nonpalpable lesions.</td>
</tr>
<tr>
<td>Fajardo et al. 2004 [15035520]</td>
<td>Multiple methods</td>
<td>Lesion type</td>
<td>Sensitivity was 97.4% for biopsies of masses detected on mammography and 90.7% for biopsies of nonpalpable lesions and lesions with microcalcifications.</td>
</tr>
<tr>
<td>Cipolla et al. 2003 [16823506]</td>
<td>Multiple methods</td>
<td>Lesion type</td>
<td>Correspondence between core needle biopsy and surgical biopsy results was 100% for palpable lesions but only 88% for nonpalpable lesions.</td>
</tr>
<tr>
<td>Fajardo et al. 2004 [15035520]</td>
<td>Multiple methods</td>
<td>Lesion type</td>
<td>Sensitivity was 97.4% for biopsies of masses detected on mammography and 90.7% for biopsies of nonpalpable lesions and lesions with microcalcifications.</td>
</tr>
<tr>
<td>Procedural factors</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>de Lucena et al. 2007 [17663457]</td>
<td>Ultrasound-guided automated device</td>
<td>Number of cores</td>
<td>Taking &gt;2 cores did not improve accuracy. Taking &gt;2 cores did not reduce the rate of false negatives. The 6 tumors (out of 101) that were falsely diagnosed as benign by core needle biopsy would not have been correctly diagnosed even if up to six cores were taken.</td>
</tr>
<tr>
<td>Fishman et al. 2003 [12601206]</td>
<td>Ultrasound-guided automated device</td>
<td>Number of cores</td>
<td>Taking &gt;2 cores improved the accuracy of the biopsy, with 4 cores being the optimal number. 1 case of DCIS would have been missed if fewer than 4 cores had been taken; the other 13 tumors identified in the study would have been correctly diagnosed if only 2 cores had been taken.</td>
</tr>
<tr>
<td>Kirshenbaum et al. 2003 [484822]</td>
<td>Ultrasound-guided automated device</td>
<td>Number of cores</td>
<td>1 core was diagnostic in 82.6% of cases, 2 cores in 90.5% of cases, 3 cores in 97.9% of cases, 4 cores in 98.9% of cases, and 5 cores in 100% of cases. 100% of malignant lesions were diagnosed after the fourth core.</td>
</tr>
<tr>
<td>Wiratkapun et al. 2012 [22252182]</td>
<td>Ultrasound-guided automated device</td>
<td>Number of cores</td>
<td>There was no statistically significant relationship between underestimation and number of biopsy cores.</td>
</tr>
<tr>
<td>Dahlstrom et al. 1996 [8735717]</td>
<td>Stereotactically guided automated device</td>
<td>Number of cores</td>
<td>One core was diagnostic in 71% of cases, two cores in 84% of cases, three cores in 90% of cases, four cores in 91% of cases, and five cores in 93% of cases.</td>
</tr>
</tbody>
</table>
Table 6. Studies evaluating factors that may affect test performance (20 studies, reporting on multiple factors each) (continued)

<table>
<thead>
<tr>
<th>Author, Year [PMID]</th>
<th>Biopsy Method</th>
<th>Factors Evaluated</th>
<th>Key Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Koskela et al. 2005 [16020555]</td>
<td>Stereotactically guided automated device</td>
<td>Number of cores</td>
<td>Comment that more than three cores must be taken from lesions before an accurate diagnosis can be made.</td>
</tr>
<tr>
<td>Lomoschitz et al. 2004 [15273332]</td>
<td>Stereotactically guided, vacuum-assisted</td>
<td>Number of cores</td>
<td>12 cores were necessary for accurate diagnosis, but taking &gt;12 cores did not improve accuracy.</td>
</tr>
<tr>
<td>Venkataraman et al. 2012 [22127375]</td>
<td>Stereotactically guided, vacuum-assisted</td>
<td>Number of cores</td>
<td>There was no correlation between number of cores and upgrade.</td>
</tr>
<tr>
<td>Abdsaleh et al. 2003 [12630998]</td>
<td>Multiple methods</td>
<td>Number of cores; patient breast density</td>
<td>Taking 2 cores instead of one increased the accuracy of the procedure. Technical failures were more likely to occur in women with very dense breast tissue.</td>
</tr>
<tr>
<td>Helbich et al. 1997 [9169689]</td>
<td>Multiple methods</td>
<td>Patient position</td>
<td>Patients were randomly assigned to undergo stereotactic biopsies in either seated or prone position. The accuracy data were not reported separately for each group, but the authors commented that patient position did not affect the biopsy procedure.</td>
</tr>
</tbody>
</table>

**Clinician and facility factors**

<table>
<thead>
<tr>
<th>Author, Year [PMID]</th>
<th>Biopsy Method</th>
<th>Factors Evaluated</th>
<th>Key Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pfarl et al. 2002 [12438044]</td>
<td>Stereotactically guided, vacuum-assisted</td>
<td>Operator experience</td>
<td>For 6 of the seven false-negatives, the biopsy had been performed by an operator who had previously performed fewer than 15 stereotactic-guided biopsies.</td>
</tr>
<tr>
<td>Ciatto et al. 2007 [16823506]</td>
<td>Multiple methods</td>
<td>Operator experience</td>
<td>Sensitivity of core needle biopsies improved as the operators (radiologists) gained experience, from 88% in the first year of the study to 96% in the eighth year of the study.</td>
</tr>
</tbody>
</table>

BI-RADS = Breast Imaging-Reporting and Data System; PMID = PubMed identification number.
Meta-Regression Analyses

Meta-regression analyses were possible for the following factors: needle gauge, choice of reference standard, the 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. All models accounted for the biopsy method used (i.e., imaging guidance method and type of device) and the population’s risk of cancer (average vs. high). Table 7 summarizes the findings of the meta-regression analyses. The credible intervals for all examined factors included the null value (indicating the lack of difference in sensitivity or specificity), with the exception of increased 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, because they rely on indirect comparisons across studies. Furthermore, because these results represent odds ratios for test performance outcomes that are close to 1 (e.g., sensitivity and specificity for all tests were above 0.9), readers should keep in mind that small differences among subgroups of studies can result in (very) large odds ratio values. For example, if the summary sensitivities in two subgroups are 0.99 and 0.98, the odds ratio for sensitivity is approximately $2 \approx \frac{(0.99/0.01)}{(0.98/0.02)}$.

### Table 7. Meta-regression analysis for test performance outcomes

<table>
<thead>
<tr>
<th>Modifier Category</th>
<th>Potential Modifier</th>
<th>Comparison</th>
<th>Relative Odds for Sensitivity (95% CrI)</th>
<th>Relative Odds for Specificity (95% CrI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Biopsy procedure factors</td>
<td>Needle gauge</td>
<td>14G vs &lt;14G</td>
<td>0.55 (0.12, 2.26)</td>
<td>1.59 (0.57, 5.41)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>≥15G vs. &lt;14G</td>
<td>0.35 (0.07, 1.66)</td>
<td>0.46 (0.12, 1.74)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Unclear/NR vs. &lt;14G</td>
<td>0.92 (0.28, 2.71)</td>
<td><strong>2.69 (1.12, 6.63)</strong></td>
</tr>
<tr>
<td></td>
<td>Reference standard</td>
<td>≥2 yrs vs. open biopsy</td>
<td>1.84 (0.74, 4.35)</td>
<td>1.42 (0.75, 2.54)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>6mo to 2 yrs vs. open biopsy</td>
<td>1.59 (0.75, 3.49)</td>
<td><strong>1.84 (1.05, 3.35)</strong></td>
</tr>
<tr>
<td>Lesion factors</td>
<td>Lesion palpability*</td>
<td>&gt;80% palpable lesions vs. &lt;80%</td>
<td>1.04 (0.36, 2.88)</td>
<td>0.96 (0.37, 2.00)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>palpability NR vs. ≤80%</td>
<td>0.94 (0.36, 2.39)</td>
<td>0.93 (0.40, 1.88)</td>
</tr>
<tr>
<td>Clinician and facility factors</td>
<td>Country where the study was performed</td>
<td>United States vs. other countries</td>
<td><strong>1.79 (1.02, 3.15)</strong></td>
<td>0.96 (0.63, 1.49)</td>
</tr>
<tr>
<td></td>
<td>Multicenter study</td>
<td>≥1 centers vs. single center</td>
<td>1.34 (0.49, 3.65)</td>
<td>1.11 (0.56, 2.29)</td>
</tr>
<tr>
<td>Study design</td>
<td></td>
<td>Prospective vs. retrospective</td>
<td>1.36 (0.70, 2.73)</td>
<td><strong>2.11 (1.33, 3.37)</strong></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Unclear/NR vs. retrospective</td>
<td>0.95 (0.50, 1.88)</td>
<td><strong>2.07 (1.29, 3.30)</strong></td>
</tr>
<tr>
<td>Study risk of bias</td>
<td>Intermediate vs. high</td>
<td></td>
<td>0.56 (0.19, 1.52)</td>
<td>1.27 (0.61, 2.32)</td>
</tr>
<tr>
<td></td>
<td>Low vs. high</td>
<td></td>
<td>0.56 (0.17, 1.67)</td>
<td><strong>1.66 (0.78, 3.53)</strong></td>
</tr>
</tbody>
</table>

*Excluding studies of freehand biopsy (because all lesions in such studies were palpable).

Relative odds for sensitivity compare the odds of a positive test result among patients with cancer over the levels of the modifier.

Relative odds for specificity compare the odds of a negative test among patients without cancer over the levels of the modifier.

Both metrics are obtained from the bivariate meta-analysis model and are exponentiated coefficients from logistic regression; thus, they can be interpreted as odds ratios. Results were adjusted for biopsy technique and baseline risk of breast cancer (high vs. average). CrI = credible interval; NR = not reported; yrs = years.
Risk of Bias Assessment for Studies Addressing Key Question 1

Overall, on the basis of 14 items related to risk of bias, we deemed 12 studies to be at low risk of bias, 111 to be at moderate risk of bias, and 37 to be at high risk of bias. Given our relatively strict selection criteria related to study design and completeness of followup, it is not surprising that the majority of studies reported enrolling consecutive or randomly selected patients (66 percent), were successful in enrolling 85 percent of all eligible patients (63 percent), and reported complete data on at least 85 percent of all enrolled patients (71 percent). However, only 41 percent of studies were judged to be free of spectrum bias, 68 percent were conducted retrospectively (or did not report relevant information), and 84 percent did not apply a “gold” standard reference test on all patients. In most studies (86 percent) the index test was interpreted by readers blinded to the reference standard test results. However, the vast majority of studies (99 percent) either did not provide information on whether index test results were available to interpreters of the reference standard or reported that blinding was not used. Finally, information on the incorporation of clinical information in the interpretation of the index and reference standard tests was judged inadequate in the majority of studies (99 percent for both items).

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

This section summarizes findings from a total of 144 studies (70 new studies and 74 from the original evidence report) reporting information on at least one of the outcomes relevant to Key Question 2. Overall, studies were considered to be of low to moderate risk of bias. Of note, 76 of the 160 core needle biopsy studies included in Key Question 1 did not provide any information on adverse events, and thus do not allow us to determine whether any adverse events were observed (and not reported). As such, they are uninformative for this Key Question. Further, 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 of the included studies reported information about complications associated with open surgical biopsy. The original evidence report reported findings from a study published in 1993 and a narrative review published in 2007. The study found that 10.2 percent of a series of 425 wire-localized open biopsy procedures were complicated by vasovagal reactions. The narrative review reported that 2 to 10 percent of breast surgeries are complicated by hematoma formation, and that 3.8 percent are complicated by infections. The original evidence report also identified two studies on core needle biopsy with information on adverse events of open biopsy. One study reported that 6.3 percent of open surgical biopsies were complicated by infections. A second study reported that 2.1 percent of open biopsy procedures were complicated by the development of an abscess, but none of the 234 ultrasound-guided vacuum-assisted core needle procedures had abscess development. Finally, we found one study
reporting that four of 100 surgical biopsies required repeat biopsy compared to two of 100 vacuum-assisted core needle biopsies and one study reporting low levels of pain with open biopsy when local lidocaine was used.

**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 (69 new studies and 72 from the original evidence report). Of these studies, 26 reported information related to the dissemination or displacement of cancerous cells during the biopsy procedure, and 112 allowed for the calculation of event rates for hematomas, bleeding, vasovagal reactions, and infections. Table 8 summarizes information for the incidence of these adverse events. Overall, their incidence was low: in more than 50 percent of studies reporting information, the percentage of patients experiencing each of the aforementioned outcomes was less than 2 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. Results for these outcomes, stratified by biopsy technique, are discussed below. Information on less commonly reported adverse events (including seeding) is summarized narratively in the following sections.

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Number of Studies*</th>
<th>Number of Procedures</th>
<th>Median % of Procedures Where an Event Was Observed (25th – 75th percentile)</th>
<th>Minimum-Maximum Percentage of Procedures Where an Event Was Observed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hematoma</td>
<td>58</td>
<td>32,584</td>
<td>1.44 (0.25-8.57)</td>
<td>0.00-100.00</td>
</tr>
<tr>
<td>Bleeding</td>
<td>46</td>
<td>21,545</td>
<td>1.21 (0.33-3.97)</td>
<td>0.00-100.00</td>
</tr>
<tr>
<td>Bleeding requiring treatment</td>
<td>47</td>
<td>22,600</td>
<td>0.00 (0.00-0.14)</td>
<td>0.00-6.45</td>
</tr>
<tr>
<td>Infection</td>
<td>40</td>
<td>25,688</td>
<td>0.00 (0.00-0.33)</td>
<td>0.00-2.91</td>
</tr>
<tr>
<td>Vasovagal reaction</td>
<td>39</td>
<td>14,482</td>
<td>1.27 (0.37-3.88)</td>
<td>0.00-10.90</td>
</tr>
</tbody>
</table>

*Number of studies providing information on the outcome
max = maximum; min = minimum; perc. = percentile.

**Hematomas and Bleeding**

Fifty-eight studies including 32,584 core needle biopsy procedures reported information on hematoma formation. In 50 percent of these studies the event rate for hematomas was less than 1.44 percent, and in 75 percent the event rate was less than 8.57 percent. The highest rates of hematoma formation were observed in studies of vacuum-assisted procedures. For example, in 75 percent of studies of ultrasound-guided vacuum-assisted procedures, the event rate for hematomas was 23.24 percent or greater, while no hematomas were reported in three studies of ultrasound-guided biopsies without vacuum assistance. The median hematoma event rate for studies of stereotactic-guided vacuum-assisted biopsies was 1.59 percent, whereas the maximum event rate in studies of stereotactic-guided biopsies without vacuum-assistance was 1.25 percent. Due to incomplete (and potentially selective) reporting, these percentages should be interpreted with caution; however, vacuum-assisted procedures do appear to have a higher rate of hematoma formation than other core needle biopsy methods. Eight cohorts (reported in a total of 5 studies) of ultrasound-guided vacuum-assisted biopsy identified for this update reported that in 1,487
procedures only one hematoma required surgical intervention.\textsuperscript{219} The event rates in these studies was similar to the event rate for hematomas requiring treatment calculated across 24 studies included in the original evidence report. No other newly identified studies reported information on the number of hematomas requiring treatment.

Forty-six studies of 21,545 core needle biopsy procedures reported information on bleeding. In 50 percent of these studies the event rate for bleeding was less than 1.21 percent, and in 75 percent the event rate was less than 3.97 percent. In 25 percent of studies of stereotactic-guided vacuum-assisted procedures the event rate for bleeding was 3.75 percent or greater, while the maximum event rate reported in studies of stereotactic-guided biopsies without vacuum assistance was 1.55 percent. The highest event rate in studies of ultrasound-guided vacuum-assisted biopsy was just under 8 percent, while the single study of ultrasound-guided biopsy without vacuum assistance that contained information on bleeding reported an event rate of 5.26 percent. With the same caveats as for hematoma formation, vacuum-assisted procedures appeared to be associated with bleeding more often than nonvacuum-assisted procedures. Overall, bleeding was a rare complication. In addition to the studies reporting bleeding, we identified one study in which 19 percent of 1177 patients undergoing ultrasound guided vacuum-assisted biopsy had skin ecchymosis without hematoma.\textsuperscript{219} One study of stereotactic-guided vacuum-assisted biopsy identified in our updated searches reported that of 485 women biopsied, one patient was observed in the hospital for one day due to persistent bleeding. A second study of stereotactic-guided vacuum-assisted biopsy identified in our updated searches reported that of 64 women undergoing stereotactic-guided vacuum-assisted biopsy, one patient required surgery to stop bleeding.\textsuperscript{170} The event rates in these two studies are consistent with the 0.34 percent of vacuum-assisted procedures reported in the previous report to be complicated by bleeding that required treatment. No other newly identified studies reported information on bleeding events that required treatment. Overall, 47 studies provided information on bleeding events that required additional treatment; more than half of the studies reported than no bleeding events requiring treatment were observed and the rate was lower than 0.14 percent in 75 percent of the studies.

Nine studies of various core needle techniques that were included in the original report specified that bruising occurred after core needle biopsy procedures. Three of the nine reported that bruising was a common event, two reported that approximately 50 percent of patients had bruising, and four studies reported that 45 out of 976 patients (4.6\%) had severe bruising. We identified two additional studies that reported information on bruising. One study of stereotactic-guided biopsy without vacuum assistance noted that 1.2 percent of 200 patients reported tenderness, swelling or bruising at the biopsy site following the biopsy.\textsuperscript{67} The second study reported that 2 of 101 patients undergoing stereotactic vacuum-assisted biopsy experienced mild bruising that resolved without treatment.\textsuperscript{207}

Table 9 summarizes information on hematomas and bleeding (total and cases requiring treatment), stratified by biopsy technique.
Table 9. Core needle biopsy procedures and rates of hematoma formation and bleeding

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Biopsy Technique</th>
<th>N Studies</th>
<th>Number of Procedures</th>
<th>Median % of Procedures Where an Event Was Observed (25th to 75th percentile)</th>
<th>Minimum-Maximum Percentage of Procedures Where an Event Was Observed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hematoma formation</td>
<td>all devices</td>
<td>58</td>
<td>32,584</td>
<td>1.44 (0.25-8.57)</td>
<td>0.00-100.00</td>
</tr>
<tr>
<td></td>
<td>freehand, automated</td>
<td>2</td>
<td>1,487</td>
<td>0.00-0.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>US, automated</td>
<td>3</td>
<td>937</td>
<td>0.00-3.24</td>
<td></td>
</tr>
<tr>
<td></td>
<td>US, vacuum-assisted</td>
<td>8</td>
<td>3,291</td>
<td>13.20 (3.76-23.24)</td>
<td>0.99-36.27</td>
</tr>
<tr>
<td></td>
<td>stereotactic, automated</td>
<td>5</td>
<td>1,706</td>
<td>0.97 (0.77-1.00)</td>
<td>0.00-1.25</td>
</tr>
<tr>
<td></td>
<td>stereotactic, vacuum-assisted</td>
<td>22</td>
<td>12,345</td>
<td>1.59 (0.69-8.57)</td>
<td>0.00-100.00</td>
</tr>
<tr>
<td></td>
<td>MRI, automated</td>
<td>2</td>
<td>116</td>
<td>NA</td>
<td>1.33-4.88</td>
</tr>
<tr>
<td></td>
<td>MRI, vacuum-assisted</td>
<td>1</td>
<td>58</td>
<td>NA</td>
<td>43.10</td>
</tr>
<tr>
<td></td>
<td>other</td>
<td>15</td>
<td>12,644</td>
<td>1.07 (0.00-7.20)</td>
<td>0.00-79.12</td>
</tr>
<tr>
<td>Bleeding</td>
<td>all devices</td>
<td>46</td>
<td>21,545</td>
<td>1.21 (0.33-3.97)</td>
<td>0.00-100.00</td>
</tr>
<tr>
<td></td>
<td>freehand, automated</td>
<td>3</td>
<td>1,732</td>
<td>NA</td>
<td>0.14-3.97</td>
</tr>
<tr>
<td></td>
<td>US, automated</td>
<td>1</td>
<td>190</td>
<td>NA</td>
<td></td>
</tr>
<tr>
<td></td>
<td>US, vacuum-assisted</td>
<td>6</td>
<td>2,251</td>
<td>2.52 (1.14-4.95)</td>
<td>0.70-7.84</td>
</tr>
<tr>
<td></td>
<td>stereotactic, automated</td>
<td>5</td>
<td>1,144</td>
<td>0.48 (0.00-1.29)</td>
<td>0.00-1.55</td>
</tr>
<tr>
<td></td>
<td>stereotactic, vacuum-assisted</td>
<td>19</td>
<td>13,584</td>
<td>0.78 (0.30-3.75)</td>
<td>0.14-26.94</td>
</tr>
<tr>
<td></td>
<td>MRI, automated</td>
<td>0</td>
<td>0</td>
<td>NA</td>
<td></td>
</tr>
<tr>
<td></td>
<td>MRI, vacuum-assisted</td>
<td>1</td>
<td>10</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td></td>
<td>other</td>
<td>11</td>
<td>2,634</td>
<td>2.10 (0.44-4.24)</td>
<td>0.00-100.00</td>
</tr>
</tbody>
</table>

We only report the minimum and maximum percentage of events when three or fewer studies were available for a biopsy technique. When a single study reported information we simply list the percentage of procedures associated with complications. MRI = magnetic resonance imaging; NA = not applicable; US = ultrasound.

**Infections**

Across 40 studies, including 25,688 core needle procedures, the median percentage of infectious complications was 0.00 percent. One study, identified by the original evidence report, reported that a patient developed an abscess that required surgical treatment in a series of 268 stereotactically guided vacuum-assisted procedures. One study identified by this update reported that 2 of 118 patients undergoing stereotactic vacuum-assisted biopsy developed infections that required antibiotics. Table 10 summarizes information on infections, stratified by biopsy technique.
Table 10. Core needle biopsy procedures and rates of infectious complications

<table>
<thead>
<tr>
<th>Biopsy Technique</th>
<th>N Studies</th>
<th>Number of Procedures</th>
<th>Median % of Procedures Where an Event Was Observed (25th – 75th percentile)</th>
<th>Minimum-Maximum Percentage of Procedures Where an Event Was Observed</th>
</tr>
</thead>
<tbody>
<tr>
<td>All devices</td>
<td>40</td>
<td>25,688</td>
<td>0.00 (0.00-0.33)</td>
<td>0.00-2.91</td>
</tr>
<tr>
<td>Freehand, automated</td>
<td>3</td>
<td>1,637</td>
<td>NA</td>
<td>0.00-2.00</td>
</tr>
<tr>
<td>US, automated</td>
<td>4</td>
<td>1,675</td>
<td>0.05 (0.00-0.92)</td>
<td>0.00-1.74</td>
</tr>
<tr>
<td>US, vacuum-assisted</td>
<td>4</td>
<td>1,962</td>
<td>0.00 (0.00-0.99)</td>
<td>0.00-1.98</td>
</tr>
<tr>
<td>Stereotactic, automated</td>
<td>10</td>
<td>2,321</td>
<td>0.00 (0.00-0.63)</td>
<td>0.00-2.91</td>
</tr>
<tr>
<td>Stereotactic, vacuum-assisted</td>
<td>9</td>
<td>4,803</td>
<td>0.00 (0.00-0.20)</td>
<td>0.00-0.89</td>
</tr>
<tr>
<td>MRI, automated</td>
<td>0</td>
<td>0</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>MRI, vacuum-assisted</td>
<td>0</td>
<td>0</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Other techniques</td>
<td>10</td>
<td>13,290</td>
<td>0.10 (0.00-0.26)</td>
<td>0.00-2.20</td>
</tr>
</tbody>
</table>

We only report the minimum and maximum percentage of events when three or fewer studies were available for a biopsy technique. MRI = magnetic resonance imaging; NA = not applicable; US = ultrasound.

Pain and Use of Pain Medications

The original report identified three vacuum-assisted biopsy procedures reported to have been terminated after patients complained of severe pain, and we identified one study of 4086 stereotactic-guided vacuum-assisted procedures in which four biopsies were suspended due to pain.233 No studies reported procedure termination due to patient complaints of pain in any other types of biopsy procedures. Twenty-five studies of a wide variety of biopsy methods reported information about patient pain during the procedure (pain was assessed heterogeneously and for that reason we did not calculate overall event rates).

Eleven studies reported information on the use of pain medications. One of these studies reported that 100 percent of patients were sent home with narcotics after an open biopsy procedure, and only three patients required narcotics after a core needle procedure.102 Twenty patients were reported to have required acetaminophen after a core needle procedure.125 Note that being sent home with a medication may not necessarily mean the patients required or used the medication.

Vasovagal Reactions

Thirty-nine studies with 14,482 procedures reported information about the occurrence of vasovagal reactions (fainting or near-fainting) during core needle biopsy. The median event rate in these studies was 1.27 percent, although one study reported an event rate of nearly 11 percent. More than 40 percent of the vasovagal reactions occurred in patients who were reported to have been positioned sitting upright for the biopsy procedure (many of the studies did not report patient position so the other 60 percent of vasovagal reactions could have occurred in patients positioned in a variety of positions, or could have occurred primarily in seated patients).

Table 11 summarizes information on vasovagal reactions, stratified by biopsy technique.
Table 11. Core needle biopsy procedures and rates of vasovagal reactions

<table>
<thead>
<tr>
<th>Biopsy Technique</th>
<th>N Studies</th>
<th>Number of Procedures</th>
<th>Median % of Procedures Where an Event Was Observed (25th – 75th percentile)</th>
<th>Minimum-Maximum Percentage of Procedures Where an Event Was Observed</th>
</tr>
</thead>
<tbody>
<tr>
<td>All devices</td>
<td>39</td>
<td>14,482</td>
<td>1.27 (0.37-3.88)</td>
<td>0.00-10.90</td>
</tr>
<tr>
<td>Freehand, automated</td>
<td>1</td>
<td>1,431</td>
<td>NA</td>
<td>0.00-0.00</td>
</tr>
<tr>
<td>US, automated</td>
<td>2</td>
<td>235</td>
<td>NA</td>
<td>0.53-8.89</td>
</tr>
<tr>
<td>US, vacuum-assisted</td>
<td>3</td>
<td>2,532</td>
<td>NA</td>
<td>0.43-1.43</td>
</tr>
<tr>
<td>Stereotactic, automated</td>
<td>12</td>
<td>1,978</td>
<td>1.44 (0.51-3.47)</td>
<td>0.00-8.33</td>
</tr>
<tr>
<td>Stereotactic, vacuum-assisted</td>
<td>13</td>
<td>5,843</td>
<td>1.90 (0.34-5.41)</td>
<td>0.00-10.90</td>
</tr>
<tr>
<td>MRI, automated</td>
<td>0</td>
<td>0</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>MRI, vacuum-assisted</td>
<td>1</td>
<td>10</td>
<td>NA</td>
<td>0.00</td>
</tr>
<tr>
<td>Other techniques</td>
<td>7</td>
<td>2,453</td>
<td>1.78 (0.99-2.20)</td>
<td>0.00-3.47</td>
</tr>
</tbody>
</table>

We only report the minimum and maximum percentage of events when three or fewer studies were available for a biopsy technique. When a single study reported information we simply list the percentage of procedures associated with complications. MRI = magnetic resonance imaging; NA = not applicable; US = ultrasound.

**Impact of Biopsy Procedure on Usual Activities and Time to Recovery**

Three studies provided information on the impact of biopsy procedures on usual activities. The first study reported that of 34 women undergoing ultrasound-guided vacuum-assisted breast biopsy, 16 (47%) women stated that the procedure did not interfere with usual activity, 14 (41%) stated that there was minor interference, and four (12%) felt that there was mild interference. The second study reported four cases in which the patient felt constrained in her daily life due to the procedure. The third study reported vacuum-assisted biopsy results in less psychological/physical stress when compared to surgical procedures.

A single study provided information regarding time to recovery, measured by asking patients how long it had taken for them to return to their normal activities after the biopsy procedure. This study reported that the average time of recovery was 3.5 days for open biopsy procedures and 1.5 days for stereotactically guided automated gun core needle biopsy procedures.

**Impact of Biopsy Procedure on Subsequent Mammographic Procedures**

Five studies reported information about the impact of core needle biopsies on subsequent mammographic examinations. Three studies reported on stereotactic-guided vacuum-assisted core needle procedures. These studies enrolled 3,748 patients, of whom 3,345 (89.2%) were reported to have no mammographically visible scarring after the biopsy procedure. Only seven of the patients were reported to have scars that were potentially diagnostically confusing on subsequent mammographic procedures. In the fourth study, 91 patients underwent stereotactic- or ultrasound-guided vacuum-assisted core needle biopsy. The researchers reported that at 6-month followup there was no evidence of scarring, architectural distortion, alterations of the skin, fat necrosis, or other changes that are frequently observed after surgical breast biopsy. In the fifth study, patients underwent mammography at 6 or 12 months, and the authors reported that mammograms showed structural distortions at the biopsy site in the 100 women who underwent surgical biopsy, and no sequelae in the 100 women who received vacuum-assisted core needle biopsy.
Miscellaneous Reported Adverse Events

The original report identified eight studies with information on pneumothorax, seizures, vomiting, or acute inflammation, and we identified one additional study reporting vomiting and one additional study reporting inflammation. Four studies of 2,600 patients reported that four cases of pneumothorax, none of which required treatment, had occurred. None of these four studies used the same core needle biopsy method. Two studies reported that one patient per study (out of 3,487 patients in total) had suffered a seizure during a stereotactic-guided vacuum-assisted procedure. One study of 268 patients undergoing stereotactic-guided vacuum-assisted biopsies reported that three patients developed acute inflammation at the biopsy site after the procedure. One study of 485 women undergoing stereotactic-guided vacuum-assisted biopsies reported that two patients developed signs of inflammation judged to be mastitis. Two studies reported that a patient vomited during the procedure; one of these studies was of 185 stereotactic-guided vacuum-assisted procedures and the second was of 236 vacuum-assisted procedures using either stereotactic or ultrasound guidance. We did not identify any new studies reporting any other significant adverse events associated with core biopsy procedures.

Dissemination and Displacement of Cancerous Cells During the Biopsy Procedure

To address the potential dissemination or displacement of cancerous cells by breast biopsy we did not use the study-design evaluation criteria for Key Questions 1 and 2; instead, we considered any clinical study that addressed the topic (including case reports and case series). Full details of the included studies are available in SRDR.

We reviewed 14 studies that used histopathology to demonstrate dissemination or displacement of cells by core needle biopsy procedures (four new studies and 10 studies included in the original report). Nine studies had a cohort design, and five were case series or case reports.

The percentage of needle tracks previously reported to contain displaced cancerous cells ranged from 0 to 65 percent. The original report observed that the risk of finding displaced cancerous cells was increased by greater duration of the biopsy procedure, multiple needle passes, and a short interval between core needle biopsy and surgical excision, while the risk was decreased by diagnosis of invasive lobular carcinoma and the use of vacuum-assisted core needle biopsy. The incidence of positive cytological findings in needle wash material was also greater with multiple needle passes and automated device (versus vacuum-assisted) biopsy.

Although the clinical significance of these displaced cancerous cells is debated, we found four case reports of patients developing tumors at the site of prior core needle biopsies, which supplement the six case reports previously identified for this review. Four of these ten women were reported to have not received radiation therapy for the primary tumor; for the other six women it was not reported whether they had received radiation therapy.

The previous evidence report found four studies with 1,879 women that explored the risk of tumor recurrence following biopsy. Three of these four studies reported that women who did not have a preoperative needle biopsy had a higher rate of tumor recurrence than women who did receive a preoperative needle biopsy; the fourth study reported the opposite. We identified an additional cohort study, published in 2011, that reported no development of tumors along the needle track among more than a thousand women receiving a core needle
biopsy diagnosis of cancer in early 2008 through 2009. The majority of the women in the original four studies were treated with breast-conserving surgery and radiation therapy; the newly identified fifth study did not report whether women received radiation therapy.

The original evidence report found three studies with 3,103 women that investigated the risk of seeding the lymph nodes with cancerous cells after biopsy procedures. Two of the three studies reported that the method of biopsy did not affect the rate of positive sentinel lymph nodes; the third study reported that the rate of metastases to the sentinel lymph node was higher in women who underwent some form of preoperative biopsy. We found two new studies examining the topic of epithelial cell displacement into lymph nodes after biopsy. One study described 15 cases of epithelial cell displaced into the lymph node subcapsular sinus in a series of axillary lymph node dissections taken approximately 2 weeks after either core needle or open breast biopsy. The authors stated that this was probably the result of mechanical transport of cells during biopsy and that the clinical implications are likely not significant. The second study examined epithelial displacement into lymphovascular spaces in the breast core needle biopsy specimens of seven women who were diagnosed with pure DCIS after core needle biopsy and surgical excision. These women did not have recurrences or metastases after 24 to 84 months followup. The authors suggest that because this epithelial displacement is seen in the initial core biopsy sample, the presence of tumor cell clusters in lymphovascular spaces may not reflect lymphovascular invasion.

The original evidence report identified a case series report of 25 cases of false-positive sentinel lymph nodes, in which the false-positives appeared to be caused by displacement of benign epithelial cells during a biopsy procedure. Twelve of the false-positive cases had undergone core needle biopsy prior to the sentinel lymph procedure, 12 had undergone wire-localization open biopsy, and one had undergone a fine-needle aspiration procedure. Findings of false-positive sentinel lymph nodes are clinically important because the findings are likely to lead to adverse events from unnecessary treatment. Because 22 of the 25 cases had intraductal papilloma at the biopsy site, the authors of the case series report suggested using caution when interpreting sentinel lymph node histopathology in cases where intraductal papilloma was noted during the initial biopsy procedure.

Factors That Modify the Association of Biopsy Procedures With Adverse Events

Due to the small number of studies providing information on any of the factors of interest and the poor reporting of adverse events across studies, we believe that the evidence is insufficient to establish any specific factor (other than patient positioning for vasovagal events and the use of vacuum for bleeding, as discussed in preceding sections) as a determinant of the rate of adverse events among women undergoing biopsy for breast cancer diagnosis. Information extracted from individual studies is summarized in Table 12.
<table>
<thead>
<tr>
<th>Author, Year [PMID]</th>
<th>Biopsy Technique</th>
<th>Factors Evaluated</th>
<th>Key Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Patient and lesion factors</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lin et al., 2000 [not indexed]</td>
<td>Ultrasound guided vacuum-assisted</td>
<td>Breast density</td>
<td>Among 8 women with hematomas and pre-biopsy mammograms, 75% had breasts classified as dense. No patients with breasts classified as fatty developed hematomas.</td>
</tr>
<tr>
<td>Wang et al., 2012 [21300503]</td>
<td>Ultrasound guided vacuum-assisted</td>
<td>Lesion size</td>
<td>No statistically significant difference was observed in mean lesion size for cases with and without hematoma.</td>
</tr>
<tr>
<td>Zografos et al. 2008 [18814132]</td>
<td>Stereotactic-guided vacuum-assisted</td>
<td>BI-RADS classification, patient age</td>
<td>There was no statistically significant association between hematoma formation and BI-RADS classification or patient age.</td>
</tr>
<tr>
<td>Frank et al., 2007 [17661855]</td>
<td>Stereotactic-guided automated gun</td>
<td>Patient age</td>
<td>Pain was not associated with patient age (p=0.11).</td>
</tr>
<tr>
<td>Chetlen et al., 2013 [23789678]</td>
<td>Multiple methods</td>
<td>Medications received at the time of biopsy</td>
<td>Non-clinically significant hematomas developed in 22 of 102 (21.6%) procedures performed on patients taking antithrombotic medications vs. 67 of 515 procedures (13.0%) performed on patients not on antithrombotic therapy. The probability of development of a non–clinically significant hematoma was 21.6% in association with antithrombotic therapy and 13.0% without antithrombotic therapy (p = 0.025). This finding was confirmed in multivariable logistic regression analysis. The mean log volume of hematoma in patients taking antithrombotics did not differ significantly from that in patients not taking antithrombotics (p = 0.126). In analyses adjusted for needle size, the association of antithrombotic treatment with log volume remained nonstatistically significant (p = 0.07).</td>
</tr>
<tr>
<td><strong>Procedural factors</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>McMahon et al. 1992 [1422715]</td>
<td>Freehand</td>
<td>Needle gauge</td>
<td>18G core needle procedure were associated with significantly less pain than 14G core needle procedures, but there was no significant difference in pain between 14G and 16G procedures.</td>
</tr>
<tr>
<td>Wong and Hisham 2003 [484085]</td>
<td>Freehand</td>
<td>Needle gauge</td>
<td>No difference in the amount of pain experienced by patients undergoing a 14G core needle procedure vs. a 16G core needle procedure.</td>
</tr>
<tr>
<td>Zagouri et al., 2011 [21709018]</td>
<td>Stereotactic-guided vacuum-assisted</td>
<td>Number of cores</td>
<td>In women who underwent additional sampling (96 cores vs. the standard 24-36), the rate of clinically significant hematomas doubled from 3.5% to 7.5%.</td>
</tr>
<tr>
<td>Frank et al., 2007 [17661855]</td>
<td>Stereotactic-guided automated gun</td>
<td>Number of cores, duration of procedure</td>
<td>Pain was associated with the number of biopsy cores (p=0.032) and the duration of the procedure (p=0.046).</td>
</tr>
</tbody>
</table>
Table 12. Studies evaluating factors that may affect the incidence of adverse events (continued)

<table>
<thead>
<tr>
<th>Author, Year [PMID]</th>
<th>Biopsy Technique</th>
<th>Factors Evaluated</th>
<th>Key Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Schaefer et al., 2012 [22381441]</td>
<td>Multiple methods</td>
<td>Needle gauge; biopsy device</td>
<td>There were significantly higher rates of bleeding (p&lt;0.001) and hematoma (p=0.029) in the Mammotome 8G than in the Mammotome 11G group. There were no significant differences in bleeding rates (p=0.799) or hematoma rates (p=0.596) between the ATEC 12G and the ATEC 9G group. There were no significant differences in bleeding or hematoma rates in the Mammotome 8G group and the ATEC 9G group, but there was less bleeding (p=0.015) and fewer hematomas (p=0.001) in the Mammotome 11G group than in the ATEC 12G group.</td>
</tr>
<tr>
<td>Seror et al., 2012 [21310570]</td>
<td>Multiple methods</td>
<td>Needle gauge/probe</td>
<td>There was no difference in pain with different probe sizes (12 mm, 15 mm, and 20 mm).</td>
</tr>
<tr>
<td>Szynglarewicz et al., 2011 [21367573]</td>
<td>Multiple methods</td>
<td>Vacuum-assistance; biopsy device</td>
<td>Biopsy with an automated device was significantly more painful than biopsy with a vacuum-assisted hand-held device (p&lt;0.01).</td>
</tr>
<tr>
<td>Chetlen et al., 2013 [23789678]</td>
<td>Multiple methods</td>
<td>Needle gauge; number of cores</td>
<td>The proportion of hematoma formation after biopsy with 9G needles was 29.5% vs. 3.6% after biopsy with 12G or 14G needles (the difference was statistically significant and remained significant in multivariable logistic regression analysis). The mean log volume of hematoma comparing larger versus smaller gauge needles was not statistically significantly different (p = 0.08). In analyses adjusted for antithrombotic treatment, the association of needle size with log volume became statistically significant (p = 0.048). The paper stated that the mean and median numbers of tissue samples obtained from patients who developed and those who did not develop a hematoma were not statistically significantly different. However the reported p-values were both statistically significant (p &lt; 0.001). The authors noted that the number of tissue samples was strongly correlated with needle gauge.</td>
</tr>
</tbody>
</table>

**Clinic and facility factors**

<table>
<thead>
<tr>
<th>Author, Year [PMID]</th>
<th>Biopsy Technique</th>
<th>Key Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kirshenbaum et al., 2003 [12876040]</td>
<td>Multiple methods</td>
<td>The majority of vasovagal reactions occurred when inexperienced operators performed the biopsy procedures.</td>
</tr>
</tbody>
</table>

PMID = PubMed identification number.

**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 identified 59 new studies\(^3\) that addressed various aspects of KQ3. Together with the 84 studies\(^2\) that...
include in the original evidence report, this section synthesizes evidence from 143 studies. Generally, our findings confirmed those of the original evidence report. In the following subsections, we first discuss aspects of diagnostic biopsy important to patients, followed by economic factors that may influence the choice of a particular technique, and then proceed to summarize information on other factors, including the availability of equipment, procedure duration time, time to complete tumor removal, wait time for test results, and recurrence rates. Because of the nature of this Key Question and the heterogeneity of the sources of information used to address each outcome of interest, we did not attempt to grade the strength of evidence for most outcomes considered for this Key Question (this is consistent with the original evidence report).

**Anxiety and Distress**

We identified 12 studies\(^3\)\(^5\), \(^1\)\(^2\)\(^1\), \(^2\)\(^1\)\(^1\), \(^2\)\(^2\)\(^6\), \(^2\)\(^6\)\(^6\), \(^2\)\(^6\)\(^8\), \(^2\)\(^7\)\(^6\), \(^2\)\(^8\)\(^1\), \(^2\)\(^8\)\(^2\), \(^2\)\(^8\)\(^6\), \(^2\)\(^8\)\(^8\) that looked at levels of anxiety and distress related to biopsy procedures. This outcome was not specifically examined in the original report, and we base our conclusions on the studies retrieved for this update. Overall, patients reported increased levels of anxiety and distress immediately before or during the procedure, and these levels were reduced after the procedure. One study reported mean anxiety levels just before the procedure to be well above normal on State Trait Anxiety Inventory (STAI) (mean 48; normal=35.9), Impact of Event Scale (mean 26; normal < 8.5), Center for Epidemiological Studies-Depression Scale (mean 16; normal 8), and Perceived Stress Scale (mean 19; normal=12.6).\(^2\)\(^6\)\(^8\) This was corroborated by a second study that reported participants prebiopsy STAI-S and STAI-T T scores were two standard deviations higher than the mean T score (T-score mean 50, SD 10).\(^2\)\(^8\)\(^1\) Yet another study reported that one procedure out of 602 could not be completed because of patient anxiety.\(^1\)\(^2\)\(^1\) One study found greater anxiety in surgical biopsy patients than in those receiving core needle vacuum-assisted biopsies.

Four studies,\(^\)\(^2\)\(^2\)\(^6\), \(^2\)\(^6\)\(^6\), \(^2\)\(^8\)\(^2\), \(^2\)\(^8\)\(^8\) three of which were randomized controlled trials, looked at a range of options to ameliorate stress during core needle biopsy procedures, with relaxation, medication, empathy, and hypnosis all showing reductions in stress either just before or during the procedure. One randomized controlled trial reported on stress levels in three groups of patients (those receiving usual care, relaxation, or medication to reduce anxiety). All three groups had preprocedural state anxiety levels that were significantly higher than normal and reported significant reductions in anxiety 24 hours after the procedure. Patients in the medication group reported significantly less anxiety during the procedure, when compared with the usual care and relaxation groups.\(^2\)\(^6\)\(^6\) They also reported that there was no statistically significant difference in anxiety levels during the procedure for those who underwent stereotactically guided versus ultrasound-guided procedures.\(^2\)\(^6\) A second randomized controlled trial looked at the use of empathy and hypnosis in relieving anxiety. The authors found that standard care patients experienced an increase in anxiety during the procedure, patients who were given empathy experienced no change in anxiety during the procedure, and patients receiving hypnosis experienced a decrease in anxiety during the procedure. A final randomized controlled trial reported that the main effect of an education intervention on anxiety was that those in the control group tended to have lower postconsultation anxiety than those in the education group.
Procedure Preference

We found two studies\textsuperscript{198, 286} that specifically addressed procedure preference in addition to the 20\textsuperscript{30, 46, 58, 73, 91, 125, 181, 182, 299, 303, 304, 306, 309, 311, 312, 317, 321, 322, 331, 347} in the original evidence report. Both of the new studies reported a positive experience with core needle biopsy, relative to surgical biopsy. One study reported that women who had previously experienced only core needle or surgical biopsy were willing to wait a median of 3.2 weeks longer to avoid surgical than to avoid core needle biopsy; while women who had experienced both were willing to wait 2.4 weeks longer to avoid surgical than to avoid core needle biopsy.\textsuperscript{286} This supports the findings of the original report: the majority of studies reported core needle biopsies to be preferable to open biopsies. However, a single study reported the reverse: a survey of 59 patients (20 open biopsy, 20 fine needle aspiration, and 19 core needle biopsy) from Detroit, Michigan in 1997 and 1998 found that 90 percent were satisfied with their open surgical biopsy compared to only 80 percent satisfied with a vacuum-assisted core needle biopsy, though the authors reported that this difference was not statistically significant at the p=0.05 level.\textsuperscript{299} The original evidence report also noted that the majority of the studies reported such information as that the patients tolerated the procedure well or would recommend it to others in the future. One study reported that 99 percent of image-guided core needle biopsy patients rated their overall experience as positive and 97 percent said they would recommend the center to a family member or friend if they needed a biopsy.\textsuperscript{198} Another study reported that patients preferred the decubitus position to the prone position.\textsuperscript{182} Two studies reported that vacuum-assisted procedures were more comfortable than other types of core needle biopsies.\textsuperscript{317} Two other studies reported that patients lost less time to core needle procedures than to open procedures.\textsuperscript{309}

Surgical Procedures Avoided

We identified 12 new studies\textsuperscript{35, 49, 86, 96, 118, 137, 170, 263, 269, 274, 289, 290} providing information on the number of surgical procedures avoided by the use of core needle biopsy methods for breast cancer diagnosis. Including the 30 studies\textsuperscript{42, 47, 50, 58, 115-117, 143, 167, 293-295, 297, 304, 305, 308, 319, 320, 324, 325, 327, 329, 332, 335, 340, 342, 344-346, 349} considered in the original report, a total of 42 studies provide information on this outcome. In general, studies found that core needle biopsy obviated the need for surgery for a substantial proportion of women, ranging from 29 to 87 percent. Of the 42 studies, ten reported comparisons against open surgical biopsy with respect to the number of patients requiring only one surgical procedure (vs. more than one). 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 of the possibility of confounding by indication. Women may have been selected for a specific diagnostic approach on the basis of clinical or other factors, which may also be associated with the need for additional surgical interventions.

Cosmetic Results

We identified three new studies that addressed cosmetic results\textsuperscript{62, 239, 255} with core needle or open biopsy. Two reported that core needle biopsy produced minimal scars that were acceptable to the patients, and the third reported that vacuum-assisted core needle biopsy produced "better cosmetic effects compared to open excision."\textsuperscript{255} The original evidence report identified 10 other
studies that included information on cosmetic results for vacuum-assisted core needle biopsy and reported that patients were generally satisfied with the cosmetic results. Only one of the 10 studies included in the original report compared a group of patients undergoing core needle biopsy to a group of patients undergoing open biopsy. This study reported a greater satisfaction with appearance of the breast 2 years after surgery in core needle patients (95 percent very satisfied) than in open biopsy patients (25 percent very satisfied).

Resource Utilization and Costs

We found two additional studies on the relative costs of core needle biopsy. The results below reflect a total of ten studies, including eight studies identified in the original report. The original report concluded that the costs of surgical biopsy are considerably greater than those of core needle biopsy. In this update we identified one study (2008) reporting average charges for core needle biopsy at $10,500 and excision biopsy at $11,500. The authors based their costs on the calculation of mean patient charges for initial diagnostic procedure and subsequent necessary surgeries, which were compared for patients undergoing biopsy for BI-RADS-5 lesions between 1998 and 2002. The authors recommend core needle biopsy as the initial diagnostic approach for highly suspicious lesions, based upon improved pathologic margins and fewer surgical procedures rather than significant cost savings.

Another study compared per-procedure costs of core needle biopsy and fine needle biopsy. Based on reimbursements for facility fees, but excluding professional fees, the costs were $477.92 versus $166.34, respectively.

The original evidence report reported on the relative costs of open surgical biopsy and various core needle biopsy techniques in six studies. The studies reviewed factors that included purchase price of devices, personnel time and costs, the costs of processing and analyzing samples, patient volume, whether the device is used as a complementary procedure, and what mammography results determine the use of a core needle biopsy technique. The original report also noted that MRI-guidance is the most expensive method of performing core needle biopsies. We did not find any new studies comparing the costs or cost-effectiveness of different core needle or imaging techniques.

We did not identify any new studies for resource utilization. The two studies discussed in the original evidence report stated that vacuum-assisted procedures and procedures that required dedicated prone tables required more physician and room time.

Physician Experience

We identified seven new studies, which, together with the 10 studies included in the original report, support the conclusions that greater experience with particular devices improves accuracy, shortens procedure duration times, and leads to a decrease in the number of open biopsies. One study reported a trend that indicated that in a training program, the fellows were able to establish an accurate diagnosis with fewer core biopsy samples in their later cases (i.e. as the training progressed and they gained experience). A second study introduced a training program for breast lesion excision system biopsy, for which they reported that fellows who had previous experience in vacuum-assisted biopsy could perform the new procedure after four procedures (median), while those without previous exposure showed proficiency after nine procedures (median). This was compared to the
12 procedures required for a new user to become proficient with vacuum-assisted biopsy. A survey of 79 fellows who had graduated from approved breast fellowships between 2005 and 2009 reported that many physicians feel poorly prepared to do ultrasound-guided (41 poorly prepared; 16 moderately prepared; 22 well prepared) or stereotactic (57 poorly prepared; 7 moderately prepared; 15 well prepared) core needle biopsies. A report of data from the National Accreditation Program for Breast Centers showed that the two most common deficiencies for breast centers were in standards for ultrasound-guided biopsy (24 of 238 centers failed) and for stereotactic core needle biopsy (17 of 238 centers failed). Two new studies reported on the effects of training programs. The first reported that residents performed 83 percent of vacuum-assisted biopsies and 86 percent of core needle biopsies successfully after a training program and that their comfort level increased at least one level. The other showed that surgeon-directed, multi-year, quality improvement workshops across 12 hospitals improved preoperative core biopsy rates.

Availability of a Qualified Pathologist

We did not identify any new studies for this outcome. The two studies included in the original report showed conflicting results, with one reporting that whether the specimen was read by a local or central pathologist had little effect as agreement rates were very high, and the second reporting that the pathologist’s lack of experience with the TruCut device explains its poor performance.

Availability of Equipment/Utilization

The original report identified three studies reporting on the impact of equipment availability and utilization, to which we added four more for a total of seven. The original report concluded that wait times are longer for open procedures and dedicated prone biopsy tables. We found a randomized controlled trial that reported that patients who waited 4 days or more for a core needle biopsy procedure were less satisfied than patients who waited 3 days or less (p=0.007). We did not find any new studies reporting the overall wait times for core needle biopsies or comparing wait times for core needle vs. open biopsy procedures. Other studies looked at utilization rates of core needle biopsies over time. One study reported that the nonoperative diagnosis rates in core needle biopsy had increased from 49 percent in 1995/96 to 87 percent in 2000/01 to 94 percent in 2005/06. A second study reported that with a stable total patient population and constant number of open and needle-localized procedures, stereotactic breast biopsies had increased from 56 in 1995 to 68 in 1996, 118 in 1997, and 172 in 1998. They further reported that diagnostic yield had increased in the stereotactic era. A third study reported a similar increase in core needle biopsy utilization between January 1992 and March 1998, with a corresponding decrease in open biopsies.

Procedure Duration Time

We identified an additional 17 studies that reported results for procedure duration across various types of biopsy. When these studies are added to the 40 studies...
identified in the original evidence report, reported procedure times range between 3 and 128 minutes. This large range is probably the result of different definitions for procedure time. For example, one study reported times for “total procedure” (from signing of informed consent to end of preparation for next patient) as 26.7 minutes for ultrasound guided core biopsy and 47.5 minutes for stereotactic core biopsy; “room time” (from signing of informed consent to end of procedure) as 23.1 minutes for ultrasound guided core biopsy and 36.5 minutes for stereotactic core biopsy; and “physician time” (time radiologist located lesion to time enough samples had been obtained) as 12.3 minutes for ultrasound guided core biopsy and 18.6 minutes for stereotactic core biopsy.287

Mean procedure times for ultrasound-guided core needle biopsies ranged from 3 to 60 minutes, based on 11 original39, 50, 91, 143, 188, 303, 309, 316, 330, 337, 343 and five new studies,206, 213, 226, 270, 287 while stereotactically guided core needle procedures tended to take longer, with mean procedure times ranging from 10 to 100 minutes (15 original42, 46, 54, 58, 75, 91, 141, 175, 181, 182, 296, 316, 317, 330, 349 plus five new studies137, 174, 211, 236, 287). Twelve studies (four new41, 43, 78, 163 and eight from the original report149, 249, 323, 326, 328, 333, 334, 338 studies) gave mean times for MRI-guided procedures, ranging from 8 to 70 minutes.

Vacuum-assisted core biopsies had reported mean or median durations of 3 to 70 minutes, based on 28 studies (19 original39, 42, 46, 58, 143, 181, 230, 249, 298, 302, 303, 323, 326, 328, 330, 331, 333, 334, 347 and 9 new41, 43, 121, 137, 138, 206, 213, 256, 287). Two studies gave mean times for open procedures, ranging from 40 to 45 minutes depending on tumor size.259, 298

### Time to Complete Tumor Removal

We identified nine studies142, 229, 262, 264, 272, 273, 280, 290 that reported results for time in days from biopsy to surgery for tumor removal. There were no studies addressing this specific outcome in the original report. Overall times from biopsy to tumor removal ranged from 5 to 153 days. One study directly compared wait times for core needle and surgical biopsies, reporting an average time from initial procedure to final surgical procedure for core needle biopsy as 27 days and excisional biopsy as 22 days.290 The rest of the studies gave results for core needle biopsy only, with means ranging from 14 to 62 days and medians ranging from 9 to 83 days. One study reported that the implementation of a Rapid Diagnosis and Support Program reduced the time from biopsy to surgery from 51.54 to 33.36 days.262 A second study reported that excisional biopsy is a factor in the increase in delays between first physician encounter and surgery from 1992 to 2005.264

### Wait Time for Test Results

We found seven studies that discussed wait times for core needle biopsy results.142, 198, 262, 272, 273, 291, 292 There were two studies included in the original report, for a total of nine studies addressing this outcome. Overall, core needle wait times ranged from 1 to 114 days, with most reported as between 1 and 1.3 days. The two studies in the original report that compared wait times after core needle and open biopsies showed that wait times for core needle biopsy results are shorter by an average of 7 to 10 days. One study reported that using a microwave processor (a nonstandard processing method that is not in widespread use) to reduce wait times for test results reduced the average wait for results (P<0.001).142 Another study reported that the implementation of a Rapid Diagnosis and Support Program reduced wait times from 3.92 to 3.35 days.262 One study looked specifically at the reasons for diagnostic delays that exceeded 90 days
and found that many diagnostic delays were the result of false negative results that were caused by sampling errors. Two studies assessed patient satisfaction with wait times. One found that most participants (88 percent) thought the wait for test results (usually the day after the biopsy by phone) was reasonable. The other reported an improvement in patient satisfaction scores sense of timeliness of provision of diagnostic test results from 4.5 (of 5) before the implementation of the Rapid Diagnosis and Support Program to 4.75 after its implementation.

Recurrence Rates

We found one study that discussed recurrence rates among core needle biopsy patients. In this study, 143 lesions in 86 patients were completely excised using an ultrasound-guided Mammotome system. Excision was considered complete when a fluid-filled cavity or air bubbles were demonstrated by ultrasound. Of these 143 lesions, only one lesion recurred within six months. A second biopsy showed breast adenosis, the same diagnosis as the original biopsy.
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 E-G 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). Interested readers should consult Appendix D for the detailed assessment of 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 (0.98). 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 13 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. Of note, we rated the strength of evidence on both absolute and comparative test performance, whereas the original report considered absolute test performance only.

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>
<thead>
<tr>
<th>Outcome</th>
<th>Comparison or Biopsy Method</th>
<th>Overall Rating</th>
<th>Key Findings and Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Test performance of individual biopsy methods</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
| | 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 | Moderate | – 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] |
| **Modifiers of test performance for women at average and high risk of breast cancer** | All biopsy methods | Insufficient | – 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) |

CrIs = 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 14.
### Table 14. Strength of evidence for underestimation rates in women at average risk of cancer

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Comparison or Biopsy Method</th>
<th>Overall Rating</th>
<th>Key Findings and Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>DCIS underestimation</td>
<td>Ultrasound-guided, automated</td>
<td>Low</td>
<td>– Average underestimation probability: 0.38 (0.26 to 0.51) [14 studies]</td>
</tr>
<tr>
<td></td>
<td>Ultrasound-guided, vacuum-assisted</td>
<td>Low</td>
<td>– Average underestimation probability: 0.09 (0.02 to 0.26) [5 studies]</td>
</tr>
<tr>
<td></td>
<td>Stereotactically guided, automated</td>
<td>Low</td>
<td>– Average underestimation probability: 0.26 (0.19 to 0.36) [18 studies]</td>
</tr>
<tr>
<td></td>
<td>Stereotactically guided, vacuum-assisted</td>
<td>Low</td>
<td>– Average underestimation probability: 0.11 (0.08 to 0.14) [34 studies]</td>
</tr>
<tr>
<td></td>
<td>Other biopsy methods</td>
<td>Insufficient</td>
<td>No available studies or few studies with small numbers of lesions.</td>
</tr>
<tr>
<td>High risk lesion</td>
<td>Ultrasound-guided, automated</td>
<td>Low</td>
<td>– Average underestimation probability: 0.25 (0.16 to 0.36) [21 studies]</td>
</tr>
<tr>
<td>underestimation rate</td>
<td>Ultrasound-guided, vacuum-assisted</td>
<td>Low</td>
<td>– Average underestimation probability: 0.11 (0.02 to 0.33) [9 studies]</td>
</tr>
<tr>
<td></td>
<td>Stereotactically guided, automated</td>
<td>Low</td>
<td>– Average underestimation probability: 0.47 (0.37 to 0.56) [29 studies]</td>
</tr>
<tr>
<td></td>
<td>Stereotactically guided, vacuum-assisted</td>
<td>Low</td>
<td>– Average underestimation probability: 0.18 (0.13 to 0.24) [40 studies]</td>
</tr>
<tr>
<td></td>
<td>Other biopsy methods</td>
<td>Insufficient</td>
<td>No available studies or few studies with small numbers of lesions.</td>
</tr>
</tbody>
</table>

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 or adverse events requiring additional treatment. 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 15.

Among core needle biopsy methods, vacuum-assisted methods appeared to be associated with increased bleeding and hematoma formation. 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. Cancer cell seeding along the needle tract was a rare outcome. 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 almost 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 15. Strength of evidence assessment for adverse events of biopsy

<table>
<thead>
<tr>
<th>Outcomes</th>
<th>Test or Comparison</th>
<th>Overall Rating</th>
<th>Key Findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bleeding (any severity)</td>
<td>Alternative core needle biopsy methods</td>
<td>Low</td>
<td>– Median %: 1.21 (25&lt;sup&gt;th&lt;/sup&gt; perc. = 0.33; 75&lt;sup&gt;th&lt;/sup&gt; perc. = 3.97)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>– Selective outcome and analysis reporting likely</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>– Few studies reported bleeding requiring treatment; the event rate was low (&lt;0.40 perc.) in those studies</td>
</tr>
<tr>
<td>Bleeding events that require treatment</td>
<td>Alternative core needle biopsy methods</td>
<td>Low</td>
<td>– Median %: 0 (25&lt;sup&gt;th&lt;/sup&gt; perc. = 0; 75&lt;sup&gt;th&lt;/sup&gt; perc. = 0.14)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>– Selective outcome and analysis reporting likely</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>– Few studies reported bleeding requiring treatment; the event rate was low</td>
</tr>
<tr>
<td>Hematoma formation</td>
<td>Alternative core needle biopsy methods</td>
<td>Low</td>
<td>– Median %: 1.44 (25&lt;sup&gt;th&lt;/sup&gt; perc. = 0.25; 75&lt;sup&gt;th&lt;/sup&gt; perc. = 8.57)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>– Selective outcome and analysis reporting likely</td>
</tr>
<tr>
<td>Infectious complications</td>
<td>Alternative core needle biopsy methods</td>
<td>Low</td>
<td>– Median %: 0 (25&lt;sup&gt;th&lt;/sup&gt; perc. = 0; 75&lt;sup&gt;th&lt;/sup&gt; perc. = 0.33)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>– Selective outcome and analysis reporting likely</td>
</tr>
<tr>
<td>Vasovagal reactions:</td>
<td>Alternative core needle biopsy methods</td>
<td>Low</td>
<td>– Median %: 1.27 (25&lt;sup&gt;th&lt;/sup&gt; perc. = 0.37; 75&lt;sup&gt;th&lt;/sup&gt; perc. = 3.88)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>– Potential for selective outcome and analysis reporting</td>
</tr>
<tr>
<td>Pain and severe pain</td>
<td>Alternative core needle biopsy methods</td>
<td>Low</td>
<td>25 studies of a wide variety of biopsy methods reported information about patient pain during the procedure (pain was assessed heterogeneously across studies).</td>
</tr>
<tr>
<td>Other adverse events</td>
<td>Alternative core needle biopsy methods</td>
<td>Insufficient</td>
<td>– Most events were reported by a single study precluding assessment of consistency</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>– Individual studies did not provide adequate information for precise estimation of the event rate</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>– Only informal indirect comparisons among biopsy methods were possible</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>– Selective outcome and analysis reporting likely</td>
</tr>
<tr>
<td></td>
<td>Sitting upright during the biopsy procedure</td>
<td>Low</td>
<td>– Vasovagal reactions were more common among patients sitting during the biopsy procedure</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>– Results were reported in few studies (11 studies; 8 from the original evidence report and 3 from this update)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>– Selective outcome and analysis reporting likely</td>
</tr>
<tr>
<td></td>
<td>Vacuum-assisted versus nonvacuum assisted biopsy methods</td>
<td>Low</td>
<td>– Vacuum-assisted procedures were generally associated with increased rates of bleeding and hematoma formation</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>– Bleeding events were generally uncommon</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>– Comparisons among biopsy methods were based on informal indirect comparisons (across studies)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>– Selective outcome and analysis reporting likely</td>
</tr>
<tr>
<td></td>
<td>Comparisons among alternative core needle biopsy methods</td>
<td>Insufficient</td>
<td>– Most factors assessed by a single study limiting our ability to assess consistency</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>– Selective outcome and analysis reporting likely</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>– Within-study comparisons provided direct evidence</td>
</tr>
</tbody>
</table>

perc. = percentile.

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:

- Published evidence on the test performance and adverse events of open surgical biopsy was sparse.
- Available studies, particularly for Key Questions 1 and 2, were at moderate to high risk of bias and the publications we reviewed did not follow the Standards for Reporting of Diagnostic Accuracy (STARD) guidelines. Information on patient selection criteria, patient or lesion characteristics (e.g., granular reporting of pathology results), was often missing or inconsistently reported. Information on adverse events and patient-relevant outcomes was often incomplete, potentially selectively reported. Studies did not use standardized definitions and ascertainment methods for adverse events. Pathology results
were not reported with adequate granularity in the majority of cases.

- Studies typically used lesions (or biopsy procedures) as the unit of analysis, instead of patients. This way, patients with multiple lesions contributed multiple observations to the analyses. Lesions belonging to the same patient are likely to have similar characteristics (i.e. they are correlated). Unfortunately, studies reported results in a way that did not allow for the correlation to be accounted for in our statistical models. As such, our analyses (and those of the original report) assume independence among lesions. If the correlation among lesions in the same patient is high (positive and close to one) individual study and meta-analytic results will underestimate uncertainty and may also be biased (the direction of bias is unpredictable). However, unless each patient contributes large numbers of lesions that are highly correlated, the underestimation of uncertainty will not be large. Further, bias is unlikely unless patients contributing large numbers of lesions also have lesions that are substantially harder (or easier) to diagnose compared to those of other patients. Without additional data on the test performance on individual lesions within patients it is not possible to ascertain the impact of this factors on our results.

- Studies provided limited information to assess the impact of various patient-, lesion-, procedure-, or system-related factors on the outcomes of breast biopsy. For example, the impact of patient age, breast density, lesion type, training and experience of the operators, and error rates of pathologists who read the samples, on test performance, adverse events, or clinical outcomes could not be assessed.

- We found very few studies on MRI-guided biopsy for women at average or high risk of cancer. Because MRI-guided biopsy is likely reserved for diagnostically challenging cases (e.g., when lesions cannot be visualized by other modalities) and may be available in specialized care settings indirect (i.e. across studies) comparisons between MRI-guided and other biopsy procedures may be confounded by factors unrelated to the diagnostic value of the tests compared.

- There is limited information 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.

**Strengths and Limitations of This Review**

We conducted an up-to-date review of the benefits and risks of breast biopsy methods for breast cancer diagnosis, with respect to test performance, underestimation rates, adverse events, and patient-relevant outcomes. Previous reviews on this topic have focused on special patient populations (e.g., patients with nonpalpable lesions), selected outcomes (e.g. DCIS underestimation\textsuperscript{352} or seeding\textsuperscript{353}), or biopsy methods (e.g., ultrasound-guided biopsy\textsuperscript{354}). Nonetheless, our work has several limitations, which – to a large extent – reflect the limitations of the underlying evidence base. Studies were deemed to be of moderate to high risk of bias because of characteristics related to their design and conduct, limiting our ability to draw strong conclusions. Information for several outcomes of interest was not reported from all available studies (e.g., underestimation rates, adverse events) raising concerns about selective outcome and analysis reporting. Information on study- or population level characteristics that could be modifiers of test performance, adverse events, or clinical outcomes, was inadequate. Thus, our ability to explore between-study heterogeneity was limited. Further, because we relied on published information and did not have access to individual patient data, we were unable to evaluate the impact of patient- or lesion-level factors on outcomes of interest.
The reference standard in the reviewed studies was a combination of clinical followup and pathologic confirmation (following open biopsy or excisional surgery). 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 (e.g., some degree of diagnostic error is possible for pathologic examination, and clinical followup may provide less than perfectly accurate information). 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.

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. Studies enrolled patients with an average age similar to that of women undergoing breast biopsy in the United States, and for indications that represent the most prevalent indications in U.S. clinical practice (i.e. mammographic findings of suspicious lesions). While fewer than half of the studies in this review were conducted in the United States, almost all 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. However, the applicability of our findings to women at high risk of breast cancer may be limited because we found few studies explicitly reporting on groups of patients at high baseline risk of breast cancer on the basis of factors such as genetic testing, or family history of disease. Of note, this may be an instance of incomplete reporting rather than a true characterization of the baseline risk of included populations (i.e. some high risk populations may have been misclassified as “average risk”).

Evidence Gaps and Ongoing Research

Table 16 summarizes the evidence gaps with regards to the Key Questions of diagnostic test performance and adverse events. A search on ClinicalTrials.gov for randomized trials comparing alternative biopsy methods did not identify trials examining biopsy techniques for breast cancer diagnosis (last search: Dec 5, 2013; 141 records retrieved).

Table 16. Evidence gaps for biopsy methods for the diagnosis of breast cancer

<table>
<thead>
<tr>
<th>Key Question</th>
<th>Category</th>
<th>Evidence Gap</th>
</tr>
</thead>
<tbody>
<tr>
<td>Comparative effectiveness of core needle biopsy and open surgical biopsy</td>
<td>General</td>
<td>Limited information on the diagnostic test performance of open surgical biopsy was available. However, expert opinion and research studies consider open biopsies to have very low measurement error (but not exactly zero).</td>
</tr>
<tr>
<td></td>
<td>Population</td>
<td>Limited information for women specified to be at high baseline risk of breast cancer.</td>
</tr>
<tr>
<td></td>
<td>Interventions &amp; Comparators</td>
<td>Limited information on MRI-guided biopsy methods (all patient populations). For other biopsy methods a large body of evidence was available; however studies were at moderate to high risk of bias and poorly reported.</td>
</tr>
<tr>
<td></td>
<td>Outcomes</td>
<td>Information on underestimation rates was relatively limited. Pathology results were not reported using consistent or sufficiently granular classification schemes.</td>
</tr>
<tr>
<td></td>
<td>Modifiers of test performance</td>
<td>Optimal core needle biopsy method for specific subgroups of patients, lesion characteristics.</td>
</tr>
</tbody>
</table>
Table 16. Evidence gaps for biopsy methods for the diagnosis of breast cancer (continued)

<table>
<thead>
<tr>
<th>Key Question</th>
<th>Category</th>
<th>Evidence Gap</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adverse events of core needle biopsy and open surgical biopsy</td>
<td>General</td>
<td>Information for adverse events of interest was incompletely and (potentially) selectively reported.</td>
</tr>
<tr>
<td>Interventions &amp; Comparators</td>
<td></td>
<td>Evidence comparing the adverse events of open and alternative core needle biopsy methods was limited.</td>
</tr>
<tr>
<td>Outcomes</td>
<td></td>
<td>Limited information was available for key adverse events of interest. Reporting in existing studies was inconsistent and potentially selective. Outcome ascertainment was not standardized.</td>
</tr>
<tr>
<td>Modifiers of adverse events</td>
<td></td>
<td>Information on factors that affect the incidence of adverse events is sparse. Unclear what subgroups of patients and lesions may be most likely to experience adverse events.</td>
</tr>
<tr>
<td>Patient-relevant and resource-related outcomes</td>
<td>General</td>
<td>Comparative effectiveness information among alternative biopsy techniques (both open and core needle) was very sparse and indirect. Comparisons between methods are susceptible to confounding and selection bias.</td>
</tr>
<tr>
<td>Population</td>
<td></td>
<td>Evidence is limited both for women at average and high risk of breast cancer.</td>
</tr>
<tr>
<td>Outcomes</td>
<td></td>
<td>The balance of benefits and risks associated with alternative breast biopsy with respect to clinical outcomes, quality of life, and resource use has not been comprehensively assessed.</td>
</tr>
</tbody>
</table>

MRI = magnetic resonance imaging.

**Future Research Needs**

- Studies of test performance are needed to evaluate MRI-guided biopsy methods. Ideally, these studies will be large (powered to achieve adequate precision), prospectively designed, multicenter investigations enrolling patients representative of those seen in clinical practice. Patient selection criteria and the characteristics of included populations should be reported in detail. Studies should use standardized histological classification systems for pathological classification the specialty and experience of those performing the biopsy procedure should be reported. The reference standard for test negative cases should be regular monitoring for an adequate period of time (e.g., 2 years).
- Although a large number of studies were available for other core needle biopsy methods we believe that 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. Given that a large number of core needle biopsies are performed annually in diverse settings, such studies could be conducted at relatively low cost.
- Large-scale databases of prospectively-collected observational data on breast biopsy procedures and outcomes could be used to evaluate the effectiveness of alternative biopsy methods with respect to short and long term outcomes, and potential modifying factors. Such studies would need to collect detailed information on baseline factors that may be associated with both the choice of biopsy method and the outcomes of interest (e.g., lesion size, palpability, imaging characteristics, etc.), to adjust for potential confounding factors. Comparisons across methods should be performed only among patients that would be candidates for assessment with all methods being compared.
- 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 is unlikely to be fruitful because existing studies indicate that biopsy procedures have sensitivities and specificities that are fairly similar and close to 1. Under these conditions randomized trials comparing alternative biopsy methods would need to enroll very large numbers of participants to allow reliable comparisons between tests.

• Additional information is also needed to define what patient and lesion factors may correspond with accuracy or adverse events of specific techniques. Future research needs to be better reported for progress to be made on these questions.

**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 are more likely to have their cancer treated with a single surgical procedure, compared with women diagnosed by open surgical biopsy.
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADH</td>
<td>Atypical ductal hyperplasia</td>
</tr>
<tr>
<td>AHRQ</td>
<td>Agency for healthcare Research and Quality</td>
</tr>
<tr>
<td>CrI</td>
<td>Credible interval</td>
</tr>
<tr>
<td>DCIS</td>
<td>Ductal carcinoma in situ</td>
</tr>
<tr>
<td>EPC</td>
<td>Evidence-based Practice Center</td>
</tr>
<tr>
<td>FN</td>
<td>False negative</td>
</tr>
<tr>
<td>FP</td>
<td>False positive</td>
</tr>
<tr>
<td>MRI</td>
<td>Magnetic resonance imaging</td>
</tr>
<tr>
<td>PICOTS</td>
<td>Populations-Interventions-Comparators-Outcomes-Timing-Setting</td>
</tr>
<tr>
<td>PRISMA</td>
<td>Preferred Reporting Items for Systematic Reviews and Meta-Analyses</td>
</tr>
<tr>
<td>QUADAS</td>
<td>Quality Assessment of Diagnostic Accuracy Studies</td>
</tr>
<tr>
<td>ROC</td>
<td>Received operating characteristic</td>
</tr>
<tr>
<td>SRDR</td>
<td>Systematic Review Data Repository</td>
</tr>
<tr>
<td>STARD</td>
<td>Standards for Reporting of Diagnostic Accuracy</td>
</tr>
<tr>
<td>TEP</td>
<td>Technical expert panel</td>
</tr>
<tr>
<td>TOO</td>
<td>Task Order Officer</td>
</tr>
<tr>
<td>TN</td>
<td>True negative</td>
</tr>
<tr>
<td>TP</td>
<td>True positive</td>
</tr>
<tr>
<td>US</td>
<td>Ultrasound</td>
</tr>
</tbody>
</table>


## Appendix A. Search Strategy

Search strategy for CINAHL/Embase/Medline, adapted from the 2009 Comparative Effectiveness Review, “Comparative Effectiveness of Core Needle Biopsy and Open Surgical Biopsy for the Diagnosis of Breast Lesions”

<table>
<thead>
<tr>
<th>Set Number</th>
<th>Concept</th>
<th>Search statement</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Breast biopsy</td>
<td>(breast biopsy or stereotactic breast biopsy or directional vacuum assisted biopsy).de.</td>
</tr>
<tr>
<td>2</td>
<td>Breast</td>
<td>Breast</td>
</tr>
<tr>
<td>3</td>
<td>Breast diseases</td>
<td>Exp breast cancer/di or exp breast neoplasms/di or exp breast disease/di or exp breast diseases/di</td>
</tr>
<tr>
<td>4</td>
<td>(breast or mammar$) and (Papilloma or calcification$ or calcinosis or tum?or$ or lesion$ or cancer or carcinoma$ or lump$)</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Combine sets</td>
<td>or/2-4</td>
</tr>
<tr>
<td>6</td>
<td>Biopsy</td>
<td>5 and ((Biopsy or tumor biopsy).de. or biops$)</td>
</tr>
<tr>
<td>7</td>
<td>Large core needle biopsy</td>
<td>6 and ((needle biopsy or biopsy needle or percutaneous biopsy).de. or (large core or needle or mammotome or mammatome or vacuum))</td>
</tr>
<tr>
<td>8</td>
<td>Open biopsy</td>
<td>6 and (breast/su or breast tumor/su)</td>
</tr>
<tr>
<td>9</td>
<td></td>
<td>6 and (su.fs. or open or excision$ or incision$ or surgical)</td>
</tr>
<tr>
<td>10</td>
<td>Combine sets</td>
<td>8 or 9</td>
</tr>
<tr>
<td>11</td>
<td>Combine sets</td>
<td>or/1,7,10</td>
</tr>
<tr>
<td>12</td>
<td>Limit by publication type</td>
<td>11 not ((letter or editorial or news or comment or case reports or note or conference paper).de. or (letter or editorial or news or comment or case reports).pt.)</td>
</tr>
<tr>
<td>13</td>
<td>Diagnostics filter</td>
<td>12 and (exp prediction and forecasting/ or (predictive value of tests or receiver operating characteristic or ROC curve or sensitivity and specificity or accuracy or diagnostic accuracy or precision or likelihood).de. or ((false or true) adj (positive or negative)))</td>
</tr>
<tr>
<td>14</td>
<td>Clinical trials filter</td>
<td>13 and ((Randomized controlled trials or random allocation or double-blind method or single-blind method or placebos or crossover studies or crossover procedure or double blind procedure or single blind procedure or placebos or latin square design or crossover design or double-blind studies or single-blind studies or triple-blind studies or random assignment or exp controlled study/ or exp clinical trial/ or exp comparative study/ or cohort analysis or follow-up studies.de. or intermethod comparison or parallel design or control group or prospective study or retrospective study or case control study or major clinical study).de. or Case control studies/ or Cohort/ or Longitudinal studies/ or Evaluation studies/ or Follow-up studies/ or Prospective studies/ or Retrospective studies/ or Case control study/ or Cohort analysis/ or Longitudinal study/ or Follow up/ or Cohort analysis/ or Followup studies/ or random$.hw. or</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>15</td>
<td>Combine sets</td>
<td>13 or 14</td>
</tr>
<tr>
<td>16</td>
<td>Eliminate overlap</td>
<td>random$ or placebo$.mp. or ((singl$ or doubl$ or tripl$ or trebl$) and (dummy or blind or sham or mask)).mp. or latin square.mp. or (time adj series) or (case adj (study or studies) or ISRCTN$.mp. or ACTRN$.mp. or (NCT$ not nctc$))</td>
</tr>
<tr>
<td>17</td>
<td>Seeding</td>
<td>12 and seeding.ti,ab.</td>
</tr>
<tr>
<td>18</td>
<td>Patient satisfactionQOL</td>
<td>12 and ((patient satisfaction or pain measurement or pain assessment or visual analog scale or quality of life).de. or satisf$ or QOL or preference$)</td>
</tr>
<tr>
<td>19</td>
<td>Adverse events</td>
<td>12 and ((ae or co).fs. or cross infection or drainage or surgical wound infection).de.)</td>
</tr>
<tr>
<td>20</td>
<td>Disfigureation</td>
<td>12 and (disfigur$ or deform$)</td>
</tr>
<tr>
<td>21</td>
<td>Combine sets</td>
<td>Or/16-20</td>
</tr>
</tbody>
</table>
## Appendix B. Excluded Studies

<table>
<thead>
<tr>
<th>Reason for Exclusion</th>
<th>Excluded Studies (see list of references on the next page)</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt; 10 patients, no seeding</td>
<td>1-33</td>
</tr>
<tr>
<td>&gt; 15% current or previous breast cancer</td>
<td>36-110</td>
</tr>
<tr>
<td>Case control or retrospective case study</td>
<td>111-260</td>
</tr>
<tr>
<td>Incomplete reference standard</td>
<td>261-315</td>
</tr>
<tr>
<td>Instrument no longer available</td>
<td>316</td>
</tr>
<tr>
<td>Less than 50% follow-up</td>
<td>317-372</td>
</tr>
<tr>
<td>No CNB or CNB not for diagnosis of breast cancer in women</td>
<td>373-573 574-769</td>
</tr>
<tr>
<td>No primary data</td>
<td>770-825</td>
</tr>
<tr>
<td>Non-English full text</td>
<td>826-914</td>
</tr>
<tr>
<td>Not outcome of interest (no KQ1, KQ2, KQ3)</td>
<td>915-1107 1108-1281 1282-1363 1364-1439</td>
</tr>
<tr>
<td>Selected on the basis of CNB or final outcomes</td>
<td>1480-1649 1650-1896 1897-2018 2019-2184</td>
</tr>
</tbody>
</table>

CNB = core-needle biopsy; KQ = Key Question.
Citations to Excluded Studies


228. RJ GEaGJaPJ. Significance of extent of lobular neoplasia on breast core biopsy. J Womens Imaging.


235. Shin SJ, Rosen PP. Excisional biopsy should be performed if lobular carcinoma in situ is seen on needle core biopsy. Arch Pathol Lab Med. 2002 Jun;126(6):697-701. PMID: 12033958.


B-26


B-37


Sewell RF, Brenchley PE. Coupling of mono- and polyamino ligands to solid phases by nucleophilic attack on coupling groups that may be spontaneously hydrolysed. J Chromatogr. 1990 Jan 26;525(1):246-8. PMID: 2338446.


708. SONUÇ MLSB-R, İÇİN K, DEĞERLERİ PÖ. POSITIVE PREDICTIVE VALUES OF THE SONOGRAPHIC BI-RADS FINAL ASSESSMENT CATEGORIES FOR BREAST LESIONS.


824. Z. PaSaNaSaSaCaHaCaLaAaaRa. The effect of a hypnotic intervention on anxiety and psychological distress in the breast biopsy suite: A pilot study [conference abstract].
845. Ding Jianhui eaSCwwbsrotfrotiaR-. Shanghai Community 8234 women with breast cancer screening results of the first round of the image analysis.


1041. E. FaBDaCaCJaFaKaGaTaaMaS. Core needle biopsy: A useful adjunct to fine-needle aspiration in select patients with palpable breast lesions. Cancer.


1337. Renshaw AA. Minimal (< or =0.1 cm) invasive carcinoma in breast core needle biopsies. Incidence, sampling, associated findings, and follow-up. Arch Pathol Lab Med. 2004 Sep;128(9):996-9. PMID: 15335264.


Publication Type: journal article.


1776. Kimmick GG, Cirrincione C, Duggan DB, et al. Fifteen-year median follow-up results after neoadjuvant doxorubicin, followed by mastectomy, followed by adjuvant cyclophosphamide, methotrexate, and fluorouracil (CMF) followed by radiation for stage III


1808. Lee AH, Denley HE, Pinder SE, et al. Excision biopsy findings of patients with breast needle core biopsies reported as suspicious of malignancy (B4) or lesion of uncertain malignant potential (B3). Histopathology. 2003 Apr;42(4):331-6. PMID: 12653944.


Appendix C. Sensitivity Analysis to the Exclusion of High Risk Lesions on Core Needle Biopsy That Were Confirmed to be High Risk Lesions in Subsequent Open Biopsy or Surgical Excision

Table C1. Summary estimates of test performance for alternative core-needle biopsy methods – women at average risk of cancer

<table>
<thead>
<tr>
<th>Biopsy method or device</th>
<th>Sensitivity</th>
<th>Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Freehand, automated</td>
<td>0.91 (0.80 to 0.96)</td>
<td>0.99 (0.97 to 1.00)</td>
</tr>
<tr>
<td>US-guided, automated</td>
<td>0.99 (0.98 to 0.99)</td>
<td>0.99 (0.98 to 1.00)</td>
</tr>
<tr>
<td>US-guided, vacuum-assisted</td>
<td>0.97 (0.92 to 0.99)</td>
<td>0.99 (0.97 to 1.00)</td>
</tr>
<tr>
<td>Stereotactically guided, automated</td>
<td>0.97 (0.95 to 0.98)</td>
<td>0.98 (0.97 to 0.99)</td>
</tr>
<tr>
<td>Stereotactically guided, vacuum-assisted</td>
<td>0.99 (0.98 to 0.99)</td>
<td>0.96 (0.94 to 0.98)</td>
</tr>
<tr>
<td>MRI-guided, automated</td>
<td>0.90 (0.57 to 0.99)</td>
<td>0.99 (0.89 to 1.00)</td>
</tr>
<tr>
<td>MRI-guided, vacuum-assisted</td>
<td>1.00 (0.98 to 1.00)</td>
<td>0.93 (0.42 to 1.00)</td>
</tr>
<tr>
<td>Multiple methods/other</td>
<td>0.99 (0.98 to 0.99)</td>
<td>0.97 (0.96 to 0.99)</td>
</tr>
</tbody>
</table>

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

Note: All numbers are medians with 95% CrIs. ‘Other’ denotes one study using grid guidance and one study that did not report information on the use of vacuum assistance.

Table C2. Summary estimates of test performance for alternative core-needle biopsy methods – women at high risk of cancer

<table>
<thead>
<tr>
<th>Biopsy method or device</th>
<th>Sensitivity (95% CrI)</th>
<th>Specificity (95% CrI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stereotactically guided, automated</td>
<td>0.97 (0.82, 1.00)</td>
<td>0.98 (0.91, 1.00)</td>
</tr>
<tr>
<td>Stereotactically guided, vacuum-assisted</td>
<td>0.99 (0.93, 1.00)</td>
<td>0.97 (0.83, 0.99)</td>
</tr>
<tr>
<td>MRI-guided, automated</td>
<td>0.90 (0.58, 0.98)</td>
<td>0.99 (0.91, 1.00)</td>
</tr>
<tr>
<td>MRI-guided, vacuum-assisted</td>
<td>1.00 (0.98, 1.00)</td>
<td>0.94 (0.52, 1.00)</td>
</tr>
</tbody>
</table>

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

Note: All numbers are medians with 95% CrIs. 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 bivariate model with risk group as a covariate.
## Appendix D. Assessment of the Strength of Evidence

<table>
<thead>
<tr>
<th>Key Question</th>
<th>Population</th>
<th>Outcome</th>
<th>Comparison or biopsy method</th>
<th>Risk of Bias</th>
<th>Consistency</th>
<th>Precision</th>
<th>Directness</th>
<th>Overall Rating</th>
<th>Key Findings and Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Key Question 1:</strong> Comparative effectiveness of core-needle biopsy and open surgical biopsy</td>
<td>Average risk women</td>
<td>Test performance</td>
<td>Freehand</td>
<td>Moderate to high</td>
<td>Consistent</td>
<td>Precise</td>
<td>Direct (studies investigating a given technique)</td>
<td>Low</td>
<td>– Sensitivity: 0.91 (0.80 to 0.96) – Specificity: 0.98 (0.95 to 1.00)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>US-guided, automated</td>
<td>Moderate to high</td>
<td>Consistent</td>
<td>Precise</td>
<td>Direct (studies investigating a given technique)</td>
<td>Moderate</td>
<td>– Sensitivity: 0.99 (0.98 to 0.99) – Specificity: 0.97 (0.95 to 0.98)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Stereotactically guided, automated</td>
<td>Moderate to high</td>
<td>Consistent</td>
<td>Precise</td>
<td>Direct (studies investigating a given technique)</td>
<td>Moderate</td>
<td>– Sensitivity: 0.97 (0.95 to 0.98) – Specificity: 0.97 (0.96 to 0.98)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>MRI-guided, automated</td>
<td>Moderate to high</td>
<td>Consistent</td>
<td>Precise</td>
<td>Direct (studies investigating a given technique)</td>
<td>Insufficient</td>
<td>– Sensitivity: 0.90 (0.57 to 0.99) – Specificity: 0.99 (0.91 to 1.00)</td>
</tr>
<tr>
<td>Average risk women</td>
<td>Comparative test performance</td>
<td>US-guided, automated vs. vacuum-assisted</td>
<td>US-guided, automated</td>
<td>Moderate to high</td>
<td>Consistent</td>
<td>Precise</td>
<td>Indirect (regression based comparisons across studies)</td>
<td>Low</td>
<td>– Difference in sensitivity: 0.01 (-0.01 to 0.06) [no difference] – Difference in specificity: -0.01 (-0.03 to 0.01) [no difference]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Stereotactically-guided, automated vs. vacuum-assisted</td>
<td>Moderate to high</td>
<td>Consistent</td>
<td>Precise</td>
<td>Indirect (regression based comparisons across studies)</td>
<td>Low</td>
<td>– 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]</td>
</tr>
</tbody>
</table>

D-1
<table>
<thead>
<tr>
<th>Key Question</th>
<th>Population</th>
<th>Outcome</th>
<th>Comparison or biopsy method</th>
<th>Risk of Bias</th>
<th>Consistency</th>
<th>Precision</th>
<th>Directness</th>
<th>Overall Rating</th>
<th>Key Findings and Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>DCIS underestimation</td>
<td>Ultrasound-guided, automated</td>
<td>Moderate to high</td>
<td>Imprecise</td>
<td>Low</td>
<td>– Average underestimation probability: 0.38 (0.26 to 0.51) [14 studies]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Stereotactically guided, vacuum-assisted</td>
<td>Moderate to high</td>
<td>Imprecise</td>
<td>Low</td>
<td>– Average underestimation probability: 0.11 (0.08 to 0.14) [34 studies]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Other biopsy methods</td>
<td>Moderate to high</td>
<td>Imprecise</td>
<td>Insufficient</td>
<td>No available studies or few studies with small numbers of lesions.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High risk lesion underestimation rate</td>
<td>Ultrasound-guided, automated</td>
<td>Moderate to high</td>
<td>Imprecise</td>
<td>Low</td>
<td>– Average underestimation probability: 0.25 (0.16 to 0.36) [21 studies]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Stereotactically guided, vacuum-assisted</td>
<td>Moderate to high</td>
<td>Imprecise</td>
<td>Low</td>
<td>– Average underestimation probability: 0.11 (0.02 to 0.33) [9 studies]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- MRI-guided, automated vs. vacuum-assisted

Risk of Bias: Moderate to high
Consistency: Reliance on indirect comparisons does not allow assessment of consistency (effects are not estimable within studies)
Precision: Very imprecise
Directness: Indirect (regression based comparisons across studies)
Overall Rating: 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]

- Ultrasound-guided, automated

Risk of Bias: Moderate to high
Consistency: Moderately inconsistent
Precision: Imprecise
Directness: Comparisons between tests were indirect (across studies)
Overall Rating: Low

- Stereotactically guided, automated

Risk of Bias: Moderate to high
Consistency: Moderately inconsistent
Precision: Imprecise
Directness: Comparisons between tests were indirect (across studies)
Overall Rating: Low

- Stereotactically guided, vacuum-assisted

Risk of Bias: Moderate to high
Consistency: Moderately inconsistent
Precision: Imprecise
Directness: Comparisons between tests were indirect (across studies)
Overall Rating: Low

- Other biopsy methods

Risk of Bias: Moderate to high
Consistency: Moderately inconsistent
Precision: Imprecise
Directness: Comparisons between tests were indirect (across studies)
Overall Rating: Insufficient

- High risk lesion underestimation rate

Risk of Bias: Moderate to high
Consistency: Moderately inconsistent
Precision: Imprecise
Directness: Comparisons between tests were indirect (across studies)
Overall Rating: Low

- Ultrasound-guided, vacuum-assisted

Risk of Bias: Moderate to high
Consistency: Moderately inconsistent
Precision: Imprecise
Directness: Comparisons between tests were indirect (across studies)
Overall Rating: Low

- Stereotactically guided, automated

Risk of Bias: Moderate to high
Consistency: Moderately inconsistent
Precision: Imprecise
Directness: Comparisons between tests were indirect (across studies)
Overall Rating: Low

- Other biopsy methods

Risk of Bias: Moderate to high
Consistency: Moderately inconsistent
Precision: Imprecise
Directness: Comparisons between tests were indirect (across studies)
Overall Rating: Insufficient

- Ultrasound-guided, vacuum-assisted

Risk of Bias: Moderate to high
Consistency: Moderately inconsistent
Precision: Imprecise
Directness: Comparisons between tests were indirect (across studies)
Overall Rating: Low

- Stereotactically guided, automated

Risk of Bias: Moderate to high
Consistency: Moderately inconsistent
Precision: Imprecise
Directness: Comparisons between tests were indirect (across studies)
Overall Rating: Low

- Other biopsy methods

Risk of Bias: Moderate to high
Consistency: Moderately inconsistent
Precision: Imprecise
Directness: Comparisons between tests were indirect (across studies)
Overall Rating: Insufficient

No available studies or few studies with small numbers of lesions.
<table>
<thead>
<tr>
<th>Key Question</th>
<th>Population</th>
<th>Outcome</th>
<th>Comparison or biopsy method</th>
<th>Risk of Bias</th>
<th>Consistency</th>
<th>Precision</th>
<th>Directness</th>
<th>Overall Rating</th>
<th>Key Findings and Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test performance</td>
<td>Women at high risk of cancer</td>
<td>All biopsy methods</td>
<td>Stereotactically guided, vacuum-assisted</td>
<td>Moderate to high</td>
<td>Moderately inconsistent</td>
<td>Imprecise</td>
<td>Comparisons between tests were indirect (across studies)</td>
<td>Low</td>
<td>– Average underestimation probability: 0.18 (0.13 to 0.24) [40 studies]</td>
</tr>
<tr>
<td>Test performance</td>
<td>Women at high risk of cancer</td>
<td>Other biopsy methods</td>
<td>Moderate to high</td>
<td>Moderately inconsistent</td>
<td>Imprecise</td>
<td>Comparisons between tests were indirect (across studies)</td>
<td>Insufficient</td>
<td>No available studies or few studies with small numbers of lesions.</td>
<td></td>
</tr>
<tr>
<td>Comparative test performance</td>
<td>Women at high risk of cancer</td>
<td>All biopsy methods</td>
<td>Comparisons of biopsy methods using the same imaging guidance</td>
<td>Moderate to high</td>
<td>Number and sample size of studies does not allow assessment of consistency</td>
<td>Imprecise</td>
<td>Direct (studies investigating a given technique)</td>
<td>Insufficient</td>
<td>No available studies or few studies with small numbers of lesions.</td>
</tr>
<tr>
<td>Comparative test performance</td>
<td>Women at high risk of cancer</td>
<td>All biopsy methods</td>
<td>DCIS underestimation</td>
<td>Moderate to high</td>
<td>Number and sample size of studies does not allow assessment of consistency</td>
<td>Imprecise</td>
<td>Comparisons between tests were indirect (across studies) or not possible</td>
<td>Insufficient</td>
<td>No available studies or few studies with small numbers of lesions.</td>
</tr>
<tr>
<td>Comparative test performance</td>
<td>Women at high risk of cancer</td>
<td>All biopsy methods</td>
<td>High risk lesion underestimation</td>
<td>Moderate to high</td>
<td>Number and sample size of studies does not allow assessment of consistency</td>
<td>Imprecise</td>
<td>Comparisons between tests were indirect (across studies) or not possible</td>
<td>Insufficient</td>
<td>No available studies or few studies with small numbers of lesions.</td>
</tr>
</tbody>
</table>

Women at average and high risk of breast cancer

Modifiers of test performance

All biopsy methods

Moderate to high

Unclear

Imprecise

Direct

Insufficient

– 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
<table>
<thead>
<tr>
<th>Key Question</th>
<th>Population</th>
<th>Outcome</th>
<th>Comparison or biopsy method</th>
<th>Risk of Bias</th>
<th>Consistency</th>
<th>Precision</th>
<th>Directness</th>
<th>Overall Rating</th>
<th>Key Findings and Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Key Question 2: Adverse events of core-needle biopsy and open surgical biopsy</td>
<td>All patient populations</td>
<td>Any complications</td>
<td>Open vs. core-needle biopsy</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>Not rated</td>
<td>– In meta-regression analyses CrIs were wide and extreme odds ratio values were often observed because sensitivity and specificity for all tests were very close to 1 (see Results for additional details)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Severe complications</td>
<td>Open vs. core-needle biopsy</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>Not rated</td>
<td>– Few studies provided information on open biopsy; comparisons of methods are indirect and based on limited empirical evidence and expert opinion – Open surgical biopsy is associated with an increased incidence of adverse events compared to core-needle biopsy</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Bleeding (any severity)</td>
<td>Comparisons among all core-needle biopsy methods</td>
<td>Moderate to high</td>
<td>Consistent</td>
<td>Imprecise</td>
<td>Indirect</td>
<td>Low</td>
<td>– Few studies provided information on open biopsy; comparisons of methods are indirect and based on limited empirical evidence and expert opinion – Open surgical biopsy is associated with an increased incidence of serious adverse events compared to core-needle biopsy</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Bleeding events that require treatment</td>
<td>Comparisons among all core-needle biopsy methods</td>
<td>Moderate to high</td>
<td>Consistent</td>
<td>Imprecise</td>
<td>Indirect</td>
<td>Low</td>
<td>– 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 (&lt;0.40 perc.) in those studies</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Hematoma formation</td>
<td>Comparisons among all core-needle biopsy methods</td>
<td>Moderate to high</td>
<td>Consistent</td>
<td>Imprecise</td>
<td>Indirect</td>
<td>Low</td>
<td>– 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</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Infectious complications</td>
<td>Comparisons among all core-needle biopsy methods</td>
<td>Moderate to high</td>
<td>Consistent</td>
<td>Imprecise</td>
<td>Indirect</td>
<td>Low</td>
<td>– Median %: 0 (25th perc. = 0; 75th perc. = 0.33) – Selective outcome and analysis reporting likely</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Vasovagal reactions</td>
<td>Comparisons among all core-needle</td>
<td>Moderate to high</td>
<td>Consistent</td>
<td>Imprecise</td>
<td>Indirect</td>
<td>Low</td>
<td>– Median %: 1.27 (25th perc. = 0.37; 75th perc. = 3.88)</td>
</tr>
<tr>
<td>Key Question</td>
<td>Population</td>
<td>Outcome</td>
<td>Comparison or biopsy method</td>
<td>Risk of Bias</td>
<td>Consistency</td>
<td>Precision</td>
<td>Directness</td>
<td>Overall Rating</td>
<td>Key Findings and Comments</td>
</tr>
<tr>
<td>--------------</td>
<td>------------</td>
<td>---------</td>
<td>-----------------------------</td>
<td>--------------</td>
<td>-------------</td>
<td>-----------</td>
<td>------------</td>
<td>----------------</td>
<td>--------------------------</td>
</tr>
<tr>
<td>Pain and severe pain</td>
<td>Comparisons among all core-needle biopsy methods</td>
<td>Moderate to high</td>
<td>Consistent</td>
<td>Imprecise</td>
<td>Indirect</td>
<td>Low</td>
<td>– Potential for selective outcome and analysis reporting</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other adverse events</td>
<td>Comparisons among all core-needle biopsy methods</td>
<td>Moderate to high</td>
<td>Unclear</td>
<td>Imprecise</td>
<td>Indirect</td>
<td>Insufficient</td>
<td>– Most events were reported by a single study precluding assessment of consistency</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>– Individual studies did not provide adequate information for precise estimation of the event rate</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>– Only informal indirect comparisons among biopsy methods were possible</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>– Potential for selective outcome and analysis reporting</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Modifiers of test adverse events – vasovagal reactions</td>
<td>Sitting upright during the biopsy procedure</td>
<td>Moderate to high</td>
<td>Unclear (few available studies; heterogeneous reporting)</td>
<td>Imprecise</td>
<td>Direct</td>
<td>Low</td>
<td>– Vasovagal reactions were more common among patients sitting during the biopsy procedure</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>– Results were reported in few studies (11 studies; 8 from the original evidence report and 3 from this update)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>– Potential for selective outcome and analysis reporting</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Modifiers of test adverse events – bleeding</td>
<td>Vacuum-assisted versus non-vacuum assisted biopsy methods</td>
<td>Moderate to high</td>
<td>Fairly consistent</td>
<td>Imprecise</td>
<td>Indirect</td>
<td>Low</td>
<td>– Vacuum-assisted procedures were generally associated with increased rates of bleeding and hematoma formation</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>– Bleeding events were generally uncommon</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>– Comparisons among biopsy methods were based on informal indirect comparisons (across studies)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>– Potential for selective outcome and analysis reporting</td>
<td></td>
<td></td>
</tr>
<tr>
<td>All other modifiers of adverse events</td>
<td>Comparisons among all core-needle biopsy methods</td>
<td>Moderate to high</td>
<td>Unclear</td>
<td>Imprecise (and sometimes impossible to assess due to incomplete information)</td>
<td>Direct</td>
<td>Insufficient</td>
<td>– Most factors assessed by a single study limiting our ability to assess consistency</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>– Potential for selective outcome and analysis reporting</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>– Within-study comparisons provided direct evidence</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

CrI = credible interval; DCIS = ductal carcinoma in situ; NA = not applicable; perc. = percentile