Australian Government
Cancer Australia

High level review of evidence and national and international guidelines to inform the *Position Statement on the use of fine needle aspiration (FNA) and core biopsy of the breast in the BreastScreen Australia program*

January 2020

Contents

List of Tables

List of Figures

Abbreviations and acronyms

AGO	Austrian Arbeitsgemeinschaft für Gynäkologische Onkologie
AHRQ	Agency for Healthcare Research and Quality
AICR	American Institute for Cancer Research
AMSTAR	Assessing the Methodological Quality of Systematic Reviews
ASCO	American Society of Clinical Oncology
AUC	area under the curve
CADTH	Canadian Agency for Drugs and Technologies in Health
BSA	BreastScreen Australia
CA	Cancer Australia
CB	core biopsy
CI	confidence interval
CRD	Centre for Reviews and Dissemination at the University of York
DARE	Database of Abstracts of Reviews of Effect
DCIS	ductal carcinoma *in situ*
DNA	deoxyribonucleic acid
EORTC	European Organisation for Research and Treatment of Cancer
ER	oestrogen receptor
ESMO	European Society for Medical Oncology
EuroScan	European Information Network on New and Changing Health Technologies
EUSOBI	European Society of Breast Imaging
FN	false negative
FNA	fine needle aspiration
FNAC	fine needle aspiration cytology
FP	false positive
GRADE	Grading of Recommendations Assessment Development and Evaluation
HealthPACT	Health Policy Advisory Committee on Technology
HER2	human epidermal growth factor receptor 2
HR	hazard ratio
HTA	Health Technology Assessment
HQO	Health Quality Ontario
IARC	International Agency for Research on Cancer
JBCS	Japanese Breast Cancer Society
JBI	Joanna Briggs Institute
KCE	Belgian Health Care Knowledge Centre
LoE	Level of Evidence
MOHM	Ministry of Health Malaysia
MRI	magnetic resonance imaging
MSAC	Medical Services Advisory Committee
NA	not applicable
NCCN	National Comprehensive Cancer Network
NCI	National Cancer Institute
NHMRC	National Health and Medical Research Council
NHS	National Health Service
NICE	National Institute for Health and Care Excellence
NLR	negative likelihood ratio
NPV	negative predictive value
NR	not reported
OR	odds ratio
PICOS	population, intervention, comparator, outcomes, study type
PIRDS	population, index test, reference standard, diagnosis of interest, study type
PLR	positive likelihood ratio

PPV	positive predictive value
PR	progesterone receptor
QoL	quality of life
QUADAS	Quality Assessment of Diagnostic Accuracy Studies
RCT	randomised controlled trial
RNA	ribonucleic acid
RoB	risk of bias
SBU	Swedish Agency for Health Technology Assessment and Assessment of Social Services
SCOS	Standing Committee on Screening
SG	stereotactic-guided
SIGN	Scottish Intercollegiate Guidelines Network
SR	systematic review
SROC	summary receiver operating characteristic
TN	true negative
TP	true positive
UG	ultrasound-guided
UK	United Kingdom
US	United States
USPSTF	United States Preventive Services Task Force
VACB	vacuum-assisted core biopsy
WHO	World Health Organization

Executive summary

The use of biopsy techniques for the assessment of mammographic screening-detected lesions varies across the BreastScreen Australia program. In this context, there is uncertainty about the appropriate use and limitations of non-excisional biopsy techniques. Accurate biopsy information is essential to the BreastScreen Australia Program, as women with screen-detected lesions and their health professionals make decisions on their management options using this information.

The objective of this high-level review is to identify and appraise the highest level clinical evidence and national and international guidelines on the use of fine needle aspiration (FNA) and core biopsy of the breast. The findings of this high-level review were used to inform the development of an up-to-date *Position Statement on the use of fine needle aspiration (FNA) and core biopsy of the breast in the BreastScreen Australia program*. The Position statement is supported by high quality evidence and data and includes recommendations to support best practice in the BreastScreen Australia Program.

Methods

Two separate reviews were undertaken to identify evidence and guidance: a review of clinical evidence and a review of recent national and international clinical guidance. Each are described below.

Clinical evidence review

Two research questions were defined for the clinical evidence review:

1. *In women with suspicious lesions detected by breast cancer screening programs[1], what is the diagnostic accuracy, reliability and reproducibility of FNA compared with core biopsy for the detection of (a) DCIS or (b) invasive breast cancer?*
2. *In women with suspicious lesions detected by breast cancer screening programs, what are the clinical consequences of FNA compared with core biopsy for the detection of (a) DCIS or (b) invasive breast cancer?*

A single literature search was conducted that encompassed the Embase and MEDLINE databases, selected databases from the Cochrane library, and selected health technology assessment (HTA) and cancer agency websites.

Study eligibility criteria included: (i) had to be in a population with suspected breast cancer; (ii) had to compare FNA and core biopsy; and (iii) had to include data on one of the outcomes defined by the Expert Advisory Group (EAG).[2] For each outcome, the most recent study with the highest level of evidence (based on the National Health and Medical Research Council [NHMRC] levels of evidence hierarchies related to diagnostic accuracy and intervention questions) was selected. Where multiple studies of the same level were available, studies assessing the use of FNA versus core biopsy in a mammographic screening population were given preference. The risk of bias of included studies was determined, and data extracted and described for each outcome.

[1] Breast cancer screening includes routine and population-based breast cancer screening

[2] Outcomes of interest were diagnostic accuracy (rate of false positive tests, sensitivity, specificity, negative likelihood ratio, positive predictive value, negative predictive value, underestimation rate), reproducibility (concordance), management (repeat tests, cascade testing, change in surgical technique or treatment decisions), adverse events (complications associated with biopsy technique, unnecessary surgery), Quality of Life (measures of worry or anxiety related to breast cancer risk, quality of life using validated instruments), and effectiveness (recurrence, progression-free survival, overall survival).

Clinical guidance review

A single literature search was conducted that encompassed clinical practice guidelines databases, the Embase and MEDLINE databases, and selected cancer agency websites. Guideline eligibility criteria included: (i) had to be related to the identification of breast cancer; and (ii) had to include recommendations on FNA, core biopsy or both.

The literature search identified 870 unique records, of which:

- one systematic review/meta-analysis (Wang et al 2017) and two prospective diagnostic cohort studies (Lieske 2006 and Bonifacino 2005) (one specifically in a screening population and one that compared FNA with VACB) met the eligibility criteria for the review of diagnostic accuracy of testing (Question 1), and
- Seven original studies (a mix of prospective and retrospective studies) met the eligibility criteria for the review of clinical impact of testing (Question 2).

During the course of determining the eligibility of studies, a number of comprehensive narrative reviews of the use of FNA and core biopsy were identified. These were also retrieved and were used to provide information on other considerations of the use of FNA and core biopsy.

Results

Diagnostic accuracy of FNA and core biopsy

Three studies provided data on the comparative diagnostic accuracy of FNA and core biopsy: a systematic review and meta-analysis comparing FNA with core biopsy (Wang 2017), a prospective cohort study comparing FNA with core biopsy (Lieske 2006), and a prospective cohort study comparing FNA with vacuum-assisted core biopsy (VACB) (Bonifacino 2005). Only the Lieske 2006 study was conducted specifically in a mammography screening setting. The Wang 2017 and Bonifacino 2005 included both screening and diagnostic setting. All studies were considered to be at a low to moderate risk of bias.

Sensitivity was lower for FNA compared with core biopsy in the overall analyses across the three studies. In Lieske 2006, complete sensitivity (which includes lesions considered to be atypical/uncertain, suspicious or malignant on FNA or core biopsy) was 83% for FNA and 93% for core biopsy. The greatest differences in sensitivity between FNA and core biopsy were seen in the following subgroups: DCIS (74% versus 94%; –20%), lesions presenting as microcalcifications on mammography (75% versus 94%; –19%) and when stereotactic guidance is used (71% versus 95%; –24%). The smallest difference between FNA and core biopsy was for the clinical guidance subgroup (91% versus 95%; –4%). These findings reflect current clinical practice whereby FNA is not generally recommended for investigation of microcalcification seen on mammography, and stereotactic guidance is rarely used when performing FNA.

The overall pooled sensitivity in the Wang 2017 meta-analysis was 74% (95% CI 72% - 77%) for FNA and 87% (95% CI 84% - 88%) for core biopsy. The differences in sensitivity between FNA and core biopsy were greatest in the following subgroups: non-palpable (68% versus 84%; –16%), free-hand guidance (80% versus 93%; –13%), and stereotactic guidance (66% versus 84%; –18%). The smallest differences were seen for ultrasound guidance (84% versus 88%; – 4%) and the subgroup of studies published in 2007 or later (81% versus 88%; –7%). As noted above, smaller differences are seen between FNA and core biopsy in the ultrasound guidance subgroup and this is reflected in current clinical practice where ultrasound guidance is the modality of choice. It should be noted that the sensitivity rates were calculated over a range of different modalities of biopsy.

In the Bonifacino 2005 study, overall absolute sensitivity (where malignant, suspicious or indeterminate results were included) was 77% for FNA, while absolute sensitivity for VACB[3] was 97%. VACB had the highest absolute sensitivity among all the methods compared in this study.

The specificity of FNA and core biopsy/VACB were similar to each other in the Wang 2017 meta-analysis (the pooled specificity for core biopsy was 98% (95% CI, 96%-99%, I^2 = 76.2%) and for FNA was 96% (95% CI, 94%-98%, I^2 = 39.0%)) and the Bonifacino 2005 study (the specificity for vacuum-assisted core biopsy was 100% and for FNA was 99%). Specificity was not measured in the Lieske 2006 study.

The overall underestimation rate in the Lieske 2006 study (that is, the proportion of patients with a benign finding on FNA or core biopsy who were subsequently found to be cancer (invasive and DCIS) was 10.0% for FNA and 2.2% for core biopsy. The biggest differences in underestimation rate for FNA compared with core biopsy were for the following subgroups: stereotactic guidance (15.2% versus 3.9%;), microcalcification (12.5% versus 3.5%;), invasive lobular lesions (19.4% versus 0%;), other invasive lesions (12.8% versus 4.7%;) and DCIS (13.6% versus 3.4%;). The underestimation rate in the Bonifacino 2005 study was 3.4% for FNA compared with 0.7% for VACB.

The findings of the current high-level review of published clinical evidence suggest that core biopsy has greater sensitivity and similar specificity relative to FNA for identifying breast cancer in a mammographic screening population. However, there are two limitations of the body of evidence that need consideration when interpreting the findings and extrapolating to the BreastScreen Australia program: the studies were conducted as long ago as the late 1980s and are unlikely to reflect current clinical practice. As practice has changed over time, results based on non-current practice may be biased against FNA.

As noted above, the sensitivity of FNA compared with core biopsy was substantially lower when stereotactic guidance was used than when US guidance was used (observed in both the Lieske 2006 study and the Wang 2017 meta-analysis). The EAG noted that the use of a 21-gauge needle for FNA in Lieske 2006, in some studies included in the Wang 2017, and in some patients in the Bonifacino 2005 study is not consistent with current Australian practice where 22-25 gauge needles are used. These differences are likely to be a consequence of the fact that the populations examined in the included studies underwent breast screening up to 30 years ago.

The fact that the evidence comes from studies conducted up to 30 years ago, with no published evidence since 2016, also likely biases the findings. Based on the results of these previous studies assessing FNA and core biopsy, it has been determined that FNA does not perform well in certain lesion types (e.g. those presenting with microcalcifications on mammography) and if identified or suspected these should be investigated using core biopsy, as is generally current practice. Thus, determining the sensitivity of FNA using the data from studies that included testing of FNA-unsuitable lesions may bias the overall findings against FNA.

Clinical impact of FNA and core biopsy

Seven studies provided the evidence of the clinical impact of FNA versus core biopsy: the previously mentioned prospective cohort studies by Lieske 2006 and Bonifacino 2005, as well as two additional prospective cohort studies (Liikanen 2016; Satchithananda 2005) and three retrospective cohort studies (Mak 2012; Hukkinen 2008; Wong 2009). While data on ancillary testing, complications, breast cancer survival and recurrence were identified, the evidence was considered to be at a high risk of bias due to the likelihood of potential confounding. Therefore, definitive conclusions around FNA and core biopsy based on these outcomes are unreliable. While

[3] In the Bonifacino 2005 study, VA core biopsy was classified as malignant or benign only.

the analysis of overall survival in the Liikanen 2016 study was considered to be at a lower risk of bias due to the use of multivariate analysis to adjust for potential confounding variables, the analysis was underpowered, and thus the lack of difference in overall survival between FNA and core biopsy may be the result of a lack of precision rather than a true lack of effect.

Current guidance on the use of FNA and core biopsy

The high-level review of clinical guidance suggests the use of core biopsy over FNA in a mammographic screening setting, with the National Health Service (NHS; 2016) advises the use of core biopsy over FNA, the European Society of Breast Imaging (EUSOBI; 2017) indicating a preference towards core biopsy, and the National Comprehensive Cancer Network (NCCN; 2018) suggesting only core biopsy following mammographic evaluation (which may refer to screening or in a diagnostic setting), with no mention of FNA. The existing BreastScreen Australia Clinical Advisory Committee (BSA; 2017) guidance provides more detail, stating "the use of FNA in the screening setting should be limited to cysts, lymph nodes, and the rare situations where core biopsy is not possible", and "where possible core biopsy, including vacuum-assisted core biopsy, should be the procedure of choice". The BSA 2017 guidance was informed by a review of the evidence and the NHS 2016 guidance. It should be noted that the NHS in the UK does not routinely employ cytology for breast lesions and breast cytology expertise in that country and in many parts of the USA is lacking.

The guidance around the use of FNA and core biopsy in the diagnostic setting is more mixed: two sources clearly recommend the use of core biopsy over FNA (European Society for Medical Oncology [ESMO; 2015]; Albert 2008), one source recommends core biopsy in the intervention pathway and make no mention of FNA (AGO; 2013), while three sources present no clear preference between core biopsy and FNA and advise that the choice depends on the level of expertise available (Belgian Healthcare Knowledge Centre [KCE; 2013]; Japanese Breast Cancer Society [JBCS; 2013]; Ministry of Health Malaysia [MOHM; 2010]). Although not specifically stated, the Malaysian guidelines appear to preference the use of FNA over core biopsy as the initial method of pathological assessment for palpable breast lumps where facility and expertise are available. They also recommended that core biopsy should be used in combination with FNA.

Other considerations

The three selected narrative reviews outline a number of additional considerations regarding the use of FNA and core biopsy, including procedural advantages and disadvantages of the two techniques, diagnostic performance based on clinical/radiological or histological features, assessment of prognostic and predictive biomarkers and the time taken to perform.

The narrative reviews note that the success of FNA is operator-dependent. Lieske et al. (2006) note that FNA is more operator-dependent than core biopsy, and that an "adequately trained and experienced cytopathologist is necessary to reduce the number of non-diagnostic FNAC (fine-needle aspiration cytology), and the interpretation of cytology is relatively more dependent on the expertise of the cytopathologist than [core biopsy] is on the quality of histopathologist." This may help explain the greater variability in the diagnostic results of FNA compared with core biopsy across the studies that were included in the Wang 2017 meta-analysis.

The issue of the appropriateness of FNA and core biopsy for different clinical or radiological lesions was noted in the narrative reviews, with core biopsy considered to perform well for both palpable and non-palpable, smaller or larger lesions (< 10 mm or > 40 mm) and lesions showing microcalcification on mammography; FNA performs well for palpable lesions only between 10 mm and 40 mm only and is not useful where microcalcification is present. With regards to diagnostic performance based on histological features, unlike core biopsy, FNA is not able to distinguish

between *in situ* and invasive cancers and has poor performance when diagnosing pre-invasive lesions and borderline lesions such as papillary lesions.

The narrative reviews also state that FNA performs poorly compared with core biopsy in terms of the assessment of prognostic and predictive biomarkers. However, more recently developed techniques such as liquid-based cytology (LBC; which standardises cell fixation, concentrates epithelial cells and discards blood cells and/or cell debris that can obscure the smear, and cell block preparation(which involves centrifuging collected cells into a pellet, and then embedding the pellet in a synthetic polymer gel that is processed in a paraffin block that can be cut at 4 μm, as per classical histopathology biopsy slides)) allow for reasonably accurate determination of hormonal and HER2 status (Garbar and Curé, 2013). The Royal College of Pathologists of Australasia (RCPA) Guideline *ASCO CAP 2018 HER2 Testing for Breast Cancer Guidelines - Recommendations for Practice in Australasia* recommends that all biomarker testing be performed on core biopsies of the primary tumour.

Summary

Core biopsy and FNA are methods for the diagnosis of screen-detected breast lesions. Whilst the evidence demonstrates that FNA is less sensitive in comparison to core biopsy, its advantages are that it offers rapid assessment and is relatively non-invasive. FNA may be used for limited circumstances such as the investigation of cysts or rare situations where a core biopsy may not be possible. The evidence demonstrates that core biopsies are more suitable in the majority of lesions (e.g. for the definitive assessment of borderline and *in situ* lesions, as well as prognostic features and receptor assessment in invasive lesions). For the investigation of the majority of screen-detected lesions in the context of the BreastScreen Australia program, core biopsy under ultrasound or image guidance is a sensitive and specific technique and is the procedure of choice for the assessment of the majority of screen-detected breast lesions.

1 Background

In 2017, the Standing Committee on Screening (SCOS) of the Australian Health Ministers' Advisory Council – Principal Advisory Committee, recommended that an evidence-based Position Statement be developed to provide the BreastScreen Australia Program with guidance on the use of non-excisional biopsies (FNA or core biopsy) in the assessment of abnormalities identified through population screening for breast cancer. This Position Statement is not intended to provide guidance for the use of biopsy techniques outside of the BreastScreen Australia Program.

FNA is a commonly utilised breast biopsy procedure. A small needle (usually 22-25 gauge) is used to remove some fluid or cells from the breast lesion of interest. Multiple samples are often taken from the same breast lesion and these are then cytologically assessed. Depending on the experience of the operator and the staining process used, FNA results can be available within 5 minutes if rapid on-site evaluation (ROSE) by a cytopathologist is performed or up to one hour if transported to the laboratory. As FNA is minimally-invasive it does not require the use of local anaesthesia (however local anaesthesia is used by some facilities). FNA can be used for palpable or non-palpable lesions and ultrasound-guidance or stereotactic-guidance are both available to assist the procedure.

Core biopsy is also a common breast biopsy technique. The patient receives local anaesthetic and a small incision is made in the breast. A hollow needle is then inserted into the breast through this incision and is guided to the lesion either by palpation or with an imaging test. This biopsy method removes a cylindrical (core) sample of the lesion that is later histologically analysed. As a result, an intact core of tissue preserving the architectural features is obtained with core biopsy, compared with a FNA cytology sample where the architectural features are lost. The staining and processing for core biopsy samples takes longer compared with FNA where samples may be prepared and stained immediately for cytological examination. Histological analysis of core biopsy samples on average may take up to 24 hours (however, faster tissue processing procedures may shorten this to 4 hours). Core biopsy can be performed on palpable or non-palpable lesions and a range of imaging guidance techniques can be utilised, including ultrasound-guidance, stereotactic-guidance and Magnetic Resonance Imaging (MRI)-guidance.

A marker clip should be placed at the lesion biopsy site at the completion of the core biopsy, when a substantial proportion of the lesion has been removed by the biopsy process. The use of lesion marker clips is also recommended when multiple lesions are biopsied, and to assist multimodality correlation when undertaking biopsy on mammographically-occult lesions under ultrasound control.

Vacuum-assisted core biopsy (VACB) is another core biopsy technique that is used to investigate breast lesions. Through a small incision or cut in the skin, a biopsy needle is inserted into the breast and, using a vacuum-powered instrument, several tissue samples are taken. The vacuum draws tissue into the centre of the needle and a rotating internal trocar then cuts the tissue and the sample is retrieved. The biopsy procedure is performed under imaging guidance (mammography, ultrasound or MRI) to ensure the needle is correctly positioned within the lesion. As this method utilises a vacuum, multiple samples can be taken without needing to move the needle. It is performed with needles of various sizes (12 gauge to 7 gauge) and can retrieve greater amounts of tissue compared with conventional core biopsy. VACB is generally used to sample very small or very diffuse lesions.

1.1 Aim

The overall aim of the high-level evidence review is to provide up-to-date evidence-based information on:

1. the *diagnostic accuracy* and *reliability* of the use of FNA and core biopsy techniques in a breast screening program; and
2. the *clinical impact* of diagnostic testing with FNA or core biopsy (repeat biopsy, cascade ancillary testing, clinical management) in a mammographic breast screening program.

The objective of this high-level review is to:

- identify and appraise highest-level clinical evidence on the use of FNA and core biopsy of relevance to breast screening programs; and
- identify and appraise national and international clinical practice guidelines on the use of FNA and core biopsy in breast screening programs.

The findings of the high-level review were used to inform the development of a *Position Statement on the use of FNA and core biopsy of the breast in the BreastScreen Australia program*.

1.2 Scope of the review

The review includes the following components:

- a review of the highest-level clinical evidence on the use of FNA and core biopsy for the detection of breast cancer (published from 2000 onwards);
- a review of clinical practice guidelines, position statements, and other guidance from Australia and other countries on the use of FNA and core biopsy for the identification of breast cancer (published from 2008 onwards).

2 Methodology

This section of the Technical Report describes the methodology used to identify and review (i) clinical evidence on the diagnostic accuracy of FNA and core biopsy in detecting breast cancer and the clinical impact of diagnostic testing with FNA or core biopsy, and (ii) clinical guidance on the use of FNA and core biopsy for the identification of breast cancer.

2.1 Clinical evidence review

2.1.1 Clinical questions and PIRDS/PICOS criteria

2.1.1.1 Diagnostic accuracy of testing

For the review of the diagnostic accuracy of FNA and core biopsy, the research question is:

> *In women with suspicious lesions detected by breast cancer screening programs[4], what is the diagnostic accuracy, reliability and reproducibility of FNA compared with core biopsy for the detection of (a) DCIS or (b) invasive breast cancer?*

PIRDS criteria (population/index test/reference standard/diagnosis of interest/study type) were developed to assist with evidence selection using information from the literature. As shown in Table 2-1, these criteria define the following five elements in detail:

- the target population for the index tests
- the index tests being considered
- the appropriate reference standards
- the outcomes that are most relevant to assess diagnostic accuracy
- the study types that will be considered for inclusion.

Additional considerations are also noted (e.g. size of needle, number of samples). Each of the PIRDS elements are described in more detail in Table 2-1.

[4] Breast cancer screening includes routine and population-based breast cancer screening

Table 2-1 PIRDS criteria and additional considerations for the diagnostic accuracy of testing

PIRDS criterion	Description
Population	Women at average risk of breast cancer participating in breast cancer screening programs[5] (including mammography, clinical examination, and self-examination)
Index tests	FNA and core biopsy (including VACB) with lesion located via any method (i.e. freehand [by palpation] or using stereotactic mammography or ultrasonography or MRI)
Reference standards	1. Final pathological results by open surgical biopsy 2. A definitive diagnosis correlated with radiological findings 3. Long term follow-up in patients with a 'benign' result who do not undergo surgery
Diagnostic outcomes	**Diagnostic accuracy** • Rate of false-positive tests • Sensitivity • Specificity • Negative likelihood ratio • Positive Predictive Value (PPV) • Negative Predictive Value (NPV) • Underestimation rates (DCIS or atypical ductal hyperplasia) **Reproducibility** • Concordance Note: Important to capture (i) how positives and negatives are classified where available (particularly atypia on cytology) and (ii) whether they were separated by radiological type or lumped together
Study type	The highest-level evidence for each outcome will be selected; will be supplemented with findings from narrative reviews where required
Additional considerations	• Size of needle • Number of samples • Whether data are reported on a per lesion or per patient basis • Duration of follow-up • Lesion type • Guidance method (if any) • Whether lesions were removed with defined margins.

Abbreviations: DCIS, ductal carcinoma *in situ*; FNA, fine needle aspiration; NPV, negative predictive value; PPV, positive predictive value; VACB, vacuum-assisted core biopsy.

2.1.1.2 Clinical impact of testing

For the review of clinical consequences of FNA compared with core biopsy, the research question is:

> *In women with suspicious lesions detected by breast cancer screening programs[6], what are the clinical consequences of FNA compared with core biopsy for the detection of (a) DCIS or (b) invasive breast cancer?*

PICOS criteria (population/intervention/comparator/outcome/study type) were developed to assist with evidence selection using information from the literature. As shown in Table 2-2, these criteria define the following five elements in detail:

- the target population for the intervention
- the intervention being considered
- the appropriate comparator
- the outcomes that are most relevant to assess safety and effectiveness
- the study types that will be considered for inclusion.

Additional considerations are also noted (e.g. frequency of screening, duration of follow-up).

[5] Breast cancer screening includes routine and population-based breast cancer screening
[6] Breast cancer screening includes routine and population-based breast cancer screening

Table 2-2 PICOS criteria and additional considerations for the clinical impact of testing

PICOS criterion	Description
Population	Women at average risk of breast cancer participating in breast cancer screening programs[7] (including mammography, clinical examination, and self-examination)
Intervention	1. FNA techniques 2. Core biopsy techniques (including VACB) with lesion located freehand (by palpation) or using stereotactic mammography or ultrasonography or MRI
Comparator	1. Core biopsy techniques (including VACB) 2. FNA techniques with lesion located freehand (by palpation) or using stereotactic mammography or ultrasonography or MRI
Outcomes	**Management** • Repeat tests • Cascade ancillary testing • Change in surgical technique or treatment decisions **Adverse events** • Complications associated with biopsy technique (including pain, bleeding and hospitalisation) • Unnecessary surgery **QoL** • Measures of worry or anxiety related to breast cancer risk • Quality of life (assessed using validated instruments) **Effectiveness** • Recurrence • Progression-free survival • Overall survival
Study type	The highest-level evidence for each outcome will be selected; will be supplemented with findings from narrative reviews where required
Additional considerations	• Size of needle • Number of samples • Whether data are reported on a per lesion or per patient basis • Duration of follow-up

Abbreviations: FNA, fine needle aspiration; MRI, magnetic resonance imaging; QoL, quality of life; VACB, vacuum-assisted core biopsy.

2.1.2 Search strategy

A comprehensive search of peer-reviewed scientific literature, designed to identify relevant studies for both clinical questions, was undertaken on 31 July 2018. The search aimed to identify health technology assessments (HTAs), systematic reviews/meta-analyses and original studies providing clinical evidence comparing the diagnostic accuracy and clinical impact of testing with FNA versus core biopsy in women with lesions detected by breast cancer screening[8].

Briefly, the search strategy included:

- a search of the Embase and MEDLINE databases using Embase®;
- a search of selected databases within the Cochrane library (Cochrane Database of Systematic Reviews; Database of Abstracts of Reviews of Effect [DARE]; [9] Health Technology Assessment Database);
- a targeted search of the websites of peak body websites and HTA agencies; and
- scanning of the reference lists of included studies for additional relevant studies that might not have been identified in the formal literature search.

[7] Breast cancer screening includes routine personal or clinical assessment and population-based breast cancer screening
[8] Breast cancer screening includes routine personal or clinical assessment and population-based breast cancer screening
[9] DARE was produced by the Centre for Reviews and Dissemination at the University of York, UK, until April 2015. Funding for DARE ceased at the end of March 2015. The April 2015 updates contain bibliographic records from searches until the end of 2014.

The search strategy, including the websites targeted and the Embase/MEDLINE databases (via Embase®) search string for the clinical evidence searches, is shown in Appendix B.

2.1.3 Study eligibility

Study eligibility was based on the PIRDS/PICOS criteria outlined in Section 2.1.1. The exclusion criteria were as follows:

- Not a full publication/report of a systematic review/meta-analysis or original (primary) study (protocols and conferences abstract were excluded);
- Not in a population being assessed for breast cancer;
- Not comparing FNA with core biopsy;
- Not assessing one of the outcomes of interest; or
- Not in English.

Following application of the exclusion criteria, the remaining studies were examined to identify the highest level of evidence available for each outcome. The NHMRC Evidence Hierarchy for diagnostic accuracy and intervention studies is shown in Appendix A. The highest level of evidence for a diagnostic accuracy question (Level I) is a systematic review of level II studies, with a level II study being a study of test accuracy with an independent, blinded comparison with a valid reference standard, among consecutive persons with a defined clinical presentation. The highest level of evidence for an intervention/clinical impact question (Level I) is a systematic review of level II studies, with a level II study being a randomised controlled trial. It should be noted that a systematic review is only assigned a level of evidence as high as the studies it contains, except where those studies are of Level II evidence.

Where multiple studies were available for an outcome, preference was given to those specifically conducted in a population undergoing mammographic screening, or where smaller or non-palpable lesions were being examined.

During the course of determining the eligibility of studies, a number of comprehensive narrative reviews of the use of FNA and core biopsy were identified. These were also retrieved and were used to provide information on other considerations of the use of FNA and core biopsy.

2.1.4 Quality appraisal

As noted above, for the assessment of the diagnostic accuracy of testing (Question 1), levels of evidence relating to diagnostic accuracy are appropriate and the highest level of evidence is a systematic review of studies of test accuracy with an independent, blinded comparison with a valid reference standard, among consecutive persons with a defined clinical presentation. For the assessment of clinical impact of testing (Question 2), the levels of evidence relating to intervention questions are appropriate, as the testing is an intervention that affects subsequent clinical management and outcome. The highest level of evidence for an intervention question is a systematic review of randomised controlled trials.

Risk of bias of systematic reviews/meta-analyses was assessed using the AMSTAR (A MeaSurement Tool to Assess systematic Reviews) 2 tool (Shea et al., 2017), while risk of bias within the individual (primary) studies included in the systematic review were captured as reported in the included systematic review/meta-analysis. It should be noted that systematic reviews of lower level evidence present results that are likely to have poor internal validity and thus are rated on the likelihood that the results have been affected by bias, rather than whether the systematic review itself is of good

quality. In systematic reviews that include different study designs, the level of evidence relates to each individual outcome/result, as different studies (and study designs) might contribute to each different outcome.

Risk of bias of diagnostic cohort studies was assessed using the QUADAS 2 (Quality Assessment of Diagnostic Accuracy Studies 2) tool (Whiting et al., 2011). For the clinical impact of testing question, the risk of bias of original studies was assessed using the appropriate study-specific tool from the Joanna Briggs Institute (JBI).[10]

[10] http://joannabriggs.org/research/critical-appraisal-tools.html.

2.2 Clinical guidance review

2.2.1 Search strategy

A comprehensive search was undertaken on 23 August 2018 to identify national and international evidence-based clinical guidance (clinical practice guidelines and position statements) on the use of FNA and/or core biopsy in breast cancer screening[11].

Briefly, the search strategy included:

- a search of the EMBASE.com electronic database;
- a search of clinical practice guideline databases;
- a targeted search of the websites of peak cancer bodies; and
- the reference lists of identified guidance documents.

The search strategy, including the websites targeted and EMBASE.com search string, is shown in Appendix B.

2.2.2 Guideline eligibility

Clinical guidance documents were considered eligible if they provided guidance on the use of core biopsy and/or FNA in women with lesions detected by any method and were published in English.[11] Guidelines published prior to 2008 were not eligible for inclusion on the basis that they may not reflect current practice.

[11] Breast cancer screening includes routine and population-based breast cancer screening

3 Results

3.1 Clinical evidence review

The literature search identified 870 unique records, of which:

- one systematic review/meta-analysis and two prospective diagnostic cohort studies (one specifically in a screening population and one that compared FNA with VACB) met the eligibility criteria for the review of diagnostic accuracy of testing (Question 1), and
- Seven original studies (a mix of prospective and retrospective studies) met the eligibility criteria for the review of clinical impact of testing (Question 2).

The characteristics and results of the identified studies are presented below.

3.1.1 Diagnostic accuracy

3.1.1.1 Study characteristics

The literature search identified one relevant systematic review/meta-analysis (Wang et al, 2017). In addition, two prospective diagnostic cohort studies were selected: (i) Lieske 2006, because it assessed the role of FNA and core biopsy in a mammographic screening population and (ii) Bonifacino 2005, because it included a comparison between FNA and VACB. The key characteristics and level/quality of the Wang 2017 systematic review and Lieske 2006 and Bonifacino 2005 diagnostic cohort studies are summarised in Table 3-1.

The systematic review by Wang 2017 assessed the sensitivity and specificity of FNA and core biopsy. The literature search by Wang 2017 was conducted in February 2016 and 12 studies were included, all of which are prospective and assess both FNA and core biopsy against a reference standard in the same population. Importantly, in all studies included within Wang 2017, FNA and core biopsy were performed on the same lesions in each patient. The study eligibility criteria for the Wang 2017 systematic review were as follows:

1. Prospective study
2. Includes ≥ 10 patients
3. Comparison data on the accuracy of FNA and core biopsy obtained from the same lesions in the same setting
4. Sufficient data to construct a two-by-two contingency table
5. Compared FNA and core biopsy against a reference standard, including:
    a. Final pathological results by open surgical biopsy or frozen section
    b. A definitive diagnosis reported on core biopsy specimens was considered as gold standard if it correlated with the radiological findings
    c. For patients with lesions with a result of benign after the biopsy and not undergoing operations, long-term follow-up information was used as an indirect assessment of the presence of breast cancer.

6. The target condition was breast cancer with five pathological result groups: inadequate, benign, benign with atypia, suspicious and malignant. Malignant and suspicious diagnoses were counted as positive results; inadequate, benign and benign with atypia were counted as negative results.

The authors used the QUADAS-2 tool to assess risk of bias of the included studies. Every study was deemed by the authors to be of 'good quality.' This systematic review also provides the raw diagnostic data from each of the included studies (i.e. the true and false positive rates, and true and false negative rates).

The characteristics and raw results of the individual studies included in Wang 2017 are presented in Table 3-2. It should be noted that the number of samples per patient/lesion and the size of the needle was not reported in Wang 2017; these data were extracted for the current review directly from the individual studies. However, as Wu 2004 was not published in English, data for these outcomes could not be extracted from this publication.

There are a number of points to note about the studies included in Wang 2017:

- The studies were published between 1991 and 2016.
- The number of patients in each study ranged from 24 to 522.
- For studies reporting age, the mean age was 34-59.
- The more recent studies tend to report data on palpable lesions that undergo free hand and ultrasound (US)-guided FNA and core biopsy, whereas, the older studies more frequently reported data from non-palpable lesions that underwent stereotactically-guided FNA and core biopsy.
- In studies reporting the number of samples taken, the range was from 2 to 7.
- The sizes of the needles for FNA and for core biopsy were similar across the individual studies for each method. The needle gauge for core biopsy was most commonly 14 gauge (two studies used higher gauge needles), while the needle sizes for FNA ranged from 20 to 23 gauge.

The key characteristics of the prospective diagnostic cohort study by Lieske 2006 are also presented in Table 3-1. The study included 763 cancers in women who had attended mammographic screening, who had both FNA and core biopsy performed on the same lesion, and who had a malignancy confirmed by subsequent surgery. Following initial mammography, 231 cancers were classified as microcalcification and the remaining 532 cancers were classified as soft tissue lesion (mass, mass associated with microcalcification, asymmetrical density or stromal deformity). The final histology resulted in the identification of invasive ductal carcinoma (53%), invasive lobular carcinoma (12%), mixed ductal and lobular carcinoma (4%) or tubular carcinoma (4%), with or without DCIS. Twenty-three percent of patients had DCIS only, while the remaining 3% had invasive carcinomas of other types.

The prospective cohort study by Bonifacino 2005 compared the diagnostic results of FNA and VACB performed on 146 lesions in 135 patients, with both diagnostic tests performed on the same lesions. Key characteristics for this study can be found in Table 3-1. All patients had lesions with a clinical/instrumental controversial diagnosis or ultrasound-undetermined lesion without mammography. Lesions were either <1.5cm in diameter or deep non-palpable lesions over 1.5cm in diameter that were difficult to sample by core biopsy. Thirty-three percent of lesions were ≤1cm in

diameter, 34% of lesions were >1cm – ≤2cm and 29% lesions were >2cm. Of the 146 lesions, 65% were palpable and 35% were non-palpable.

Table 3-1 Diagnostic accuracy – key characteristics of included studies

Study ID *Risk of bias*	Study aim	Literature search	Study/patient eligibility	Included studies (no. of patients)	Index tests and reference standards	Diagnostic accuracy outcomes
Wang 2017 *Low*	To quantitatively summarise the sensitivity and specificity of FNA and CB for suspicious breast lesions and supply useful information for clinical practitioners.	Medline, Embase.com, PubMed, The Cochrane Central Register of Controlled Trials, reference lists, suggested additional studies by researchers in the field, Conference Proceedings Citation Indexes Searched up to February 29, 2016	Inclusion • prospective studies • directly compare FNA and CB on the same lesions in the same setting • lesions are suspicious and have been referred for breast biopsy • sufficient data to construct a two-by-two contingency table such that cells could be labelled as TP, FP, TN and FN • appropriate reference standard • English and Chinese language articles Exclusion • retrospective studies • a sample size below 10 patients	12 prospective studies (N=1802): • Saha 2016 (N=50; India; good quality) • Tikku 2016 (N=85; India; good quality) • Altaf 2015 (N=60; Pakistan; good quality) • Ahmed 2010 (N=80; Pakistan; good quality) • Leaver 2010 (N=54; UK; good quality) • Barra 2008 (N=264; Brazil; good quality) • Garg 2007 (N=50; India; good quality) • Wu 2004 (N=24; China; good quality) • Dennison 2003 (N=143; UK; good quality) • Leifland 2003 (N=522; Sweden; good quality) • Cangiarella 2000 (N=181; America; good quality) • Dowlatshahi 1991 (N=242; US; good quality)	Index tests FNA and CB (free hand, stereotactic and ultrasound guidance) Reference standards • Final pathological results by open surgical biopsy or frozen section • A definitive diagnosis reported on CB specimens that correlated with radiological findings • Long-term follow up	• Sensitivity • Specificity
Lieske 2006 Moderate	To assess the performance of FNA and CB in the pre-operative diagnosis of screen-detected breast carcinoma	NA	Inclusion • Attended the Bedfordshire and Hertfordshire Breast Screening Unit between April 1999 and March 2003 • Had both FNA and CB performed at the preoperative diagnostic assessment and had malignancy confirmed by subsequent surgery	N=763	Index tests FNA and CB (free hand, stereotactic and ultrasound guidance) Reference standard Reference standard Surgical biopsy	• Absolute sensitivity • Complete sensitivity

Study ID *Risk of bias*	Study aim	Literature search	Study/patient eligibility	Included studies (no. of patients)	Index tests and reference standards	Diagnostic accuracy outcomes
Bonifacino 2005 Moderate	To determine the accuracy of well-established techniques, such as mammography, ultrasound and FNA, and a recently introduced technique, VB, in the diagnosis of breast cancer.	NA	Inclusion • Had a lesion with clinical/instrumental controversial diagnosis or ultrasound undetermined lesion without mammography • Either of the following: (i) lesion < 1.5 cm or (ii) deep non-palpable lesions > 1.5 cm in diameter, difficult to sample by core biopsy	N=135 (146 lesions)	Index tests FNA and VACB (also mammography and ultrasound) Reference standard Surgical biopsy or clinical follow-up	• Absolute sensitivity • Complete sensitivity • Specificity • PPV (malignant) • PPV (suspicious and indeterminant) • NPV • False negative rate • False positive rate • Suspicious and indeterminate rate • PLR • NLR

Abbreviations: CB, core biopsy; FN, false negative; FNA. Fine needle aspiration; FP, false positive; NLR, negative likelihood ratio; NPV, negative predictive value; PLR, positive likelihood ratio; PPV, positive predictive value; TN, true negative; TP, true positive, UK, United Kingdom; US, United States of America; VACB, vacuum-assisted core biopsy.

Table 3-2 Diagnostic accuracy – key characteristics of studies included in Wang 2017

Author, year	Study type	N	Mean age (y)	Lesion size	Index test guidance	Number of passes/samples per patient		Size of needle (gauge)		Results FNA				Results CB			
						FNA	CB	FNA	CB	TP	FP	FN	TN	TP	FP	FN	TN
Wang 2017																	
Saha 2016	Prospective cohort	50	47	Palp	Free hand	NR	NR	NR	14	29	0	13	8	35	0	7	8
Tikku 2016	Prospective cohort	85	31-50	Palp	Free hand	2-4	3-5	NR	14	31	0	17	37	46	0	2	37
Altaf 2015	Prospective cohort	60	34	Palp	US	NR	3-4	22	14	15	0	1	44	11	3	5	41
Ahmed 2010	Prospective cohort	80	44	Non-palp	US	2-3	NR	21	NR	53	1	4	22	53	2	4	21
Leaver 2010	Prospective cohort	54	NR	Palp & non-palp	US	NR	NR	NR	NR	33	0	21	39	45	0	9	39
Barra 2008	Prospective cohort	264	NR	Palp & non-palp	US	2-3	5	21	14	190	0	32	42	196	0	26	42
Garg 2007	Prospective cohort	50	45	Palp & non-palp	US	2-3	2-3	23	18	30	0	2	18	29	0	3	18
Wu 2004	Prospective cohort	24	45	Non-palp	ST	NA	NA	NA	NA	5	2	4	13	8	1	1	14
Dennison 2003	Prospective cohort	143	59	Palp	Free hand	NR	4	21	14	95	10	10	28	100	0	5	38
Leifland 2003	Prospective cohort	522	NR	Non-palp	ST	3	3	20/22	14	302	1	146	73	406	4	42	70
Cangiarella 2000	Prospective cohort	181	55	Non-palp	ST	5-7	5	22	14	27	2	9	143	23	3	13	142
Dowlatshahi 1991	Prospective cohort	242	53	Non-palp	ST	2-3	2-3	20	20	40	9	36	165	39	2	37	172
Lieske 2006	Prospective cohort	763	NR	NR	Clinical, US and ST	Variable[12]	Variable[13]	21	14	627	-	135	-	709	NR	53	NR
Bonifacino 2005	Prospective cohort	135 (146 lesions)	50	Palp & non-palp	US	3+	4-10	21-23	11 (VACB)	28	1	5	83	34	0	1	111

Source: Wang et al (2017): Table 1, page 161 with additional data extraction from individual included studies; Bonifacino 2005: Text, page 2466 and Tables IV and V, page 2467.
Abbreviations: CB, core biopsy; FN, false negative; FNA, fine needle aspiration; FP, false positive; NPV, negative predictive value; NA, not available; NR, not reported; palp, palpable; PPV, positive predictive value; SB, surgical biopsy; ST, stereotactic; TN, true negative; TP, true positive; US, ultrasound; VACB, vacuum-assisted core biopsy.

[12] 5 when performed under stereotaxis, 1-2 when performed under clinical guidance or ultrasound.
[13] 5+ when performed user stereotaxis (until calcium retrieved), 2-3 when performed under clinical guidance or ultrasound.

3.1.1.2 Study results

The results for the three included studies (the systematic review/meta-analysis by Wang 2017, the prospective cohort study conducted specifically in a population screening setting by Lieske 2006 and the prospective cohort study comparing FNA with VACB by Bonifacino 2005) are presented separately below. The following diagnostic accuracy-related outcomes were specified for inclusion in the review: rate of false positives, sensitivity, specificity, negative likelihood ratio (NLR), positive predictive value (PPV), negative predictive value (NPV), underestimation rates and concordance. It should be noted that data was available from at least one of the three included studies for all outcomes except concordance of test results between multiple operators or multiple samples.

Wang 2017

The diagnostic accuracy results from Wang 2017 (low risk of bias for all outcomes) are presented in Table 3-3. Analyses were conducted for all studies combined, and for subgroups based on study and patient characteristics (i.e. size of lesion, method of guidance and year of publication). Wang 2017 carried out pooled analyses only for sensitivity and specificity, and also calculated the area under the curve (AUC) of the summary receiver operating characteristics (ROC) curve (which has been included below). An additional post hoc pooled analysis was conducted for the purpose of this review for NLR, using Meta-Disc software[14], as well as calculation of individual included study values for NLR, PPV and NPV.

Rate of false positive tests

The rate of false positive tests indicates the probability of a positive test when the patient does not have the disease.

Based on their analysis, Wang 2017 calculated that the false positive rate for FNA was 14 per 1000 patients tested (95% CI 7–22) compared with 7 per 1000 tested for core biopsy (95% CI 4–14), resulting in 7 additional false positives per 1000 tested for FNA compared with core biopsy.[15] It should be noted that the false positive rate in the Wang study is largely related to two studies with outlying high rates. One is from 1991, and the other one from the UK and was published in 2003, with low FNA usage.

Sensitivity

The sensitivity of a diagnostic test quantifies its ability to correctly identify subjects with the disease. It is the proportion of true positives that are correctly identified by the test.

As shown in Table 3-3 and Figure 3-1, the pooled sensitivity of core biopsy from the Wang 2017 (0.87; 95% CI 0.84–0.88) was substantially higher than that of FNA (0.74; 95% CI 0.72–0.77); there was no overlap in the 95% CIs of the two results. For each test there was statistically significant heterogeneity in the individual study results for sensitivity. Visual inspection of the forest plots suggests that there is less variability associated with core biopsy results than with FNA results.

Subgroup analyses based on palpability, guidance method or publication date were also conducted. A lack of overlap in the 95% CIs for sensitivity for FNA and core biopsy was seen for the following subgroups:

- non-palpable (0.68 for FNA versus 0.84 for core biopsy),
- free-hand guidance (80% for FNA versus 93% for core biopsy),
- stereotactic guidance (66% for FNA versus 84% for core biopsy),

[14] http://www.hrc.es/investigacion/metadisc_en.htm.
[15] Source: Wang et al. (2017): Appendix 2.

- published ≥ 2007 (81% for FNA versus 88% for core biopsy), and
- published < 2007 (70% for FNA versus 86% for core biopsy).

Of particular interest to this current review, sensitivity was substantially higher for core biopsy (84%) compared with FNA (68%) when assessing non-palpable lesions, which would likely make up a substantial proportion of lesions identified via mammographic screening.

It should be noted that there is no subgroup analysis for comparison of like lesion types.

Compared with the overall results, the difference in sensitivity between FNA and core biopsy was substantially smaller in the US-guidance subgroup (74% versus 87%; –13% for overall and 84% versus 88%; –4% for subgroup).

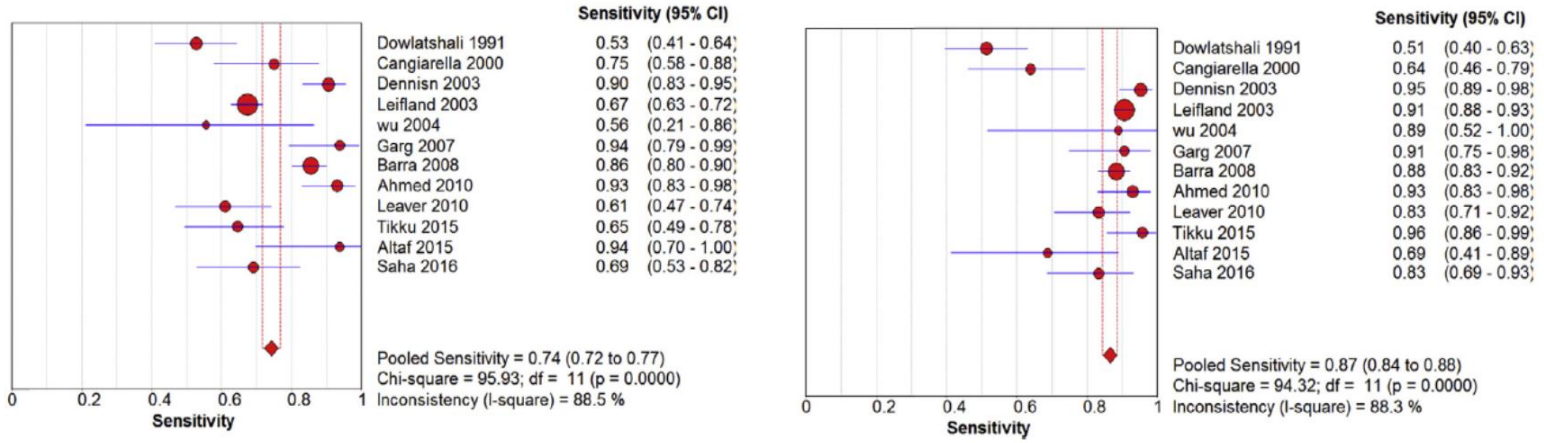
Table 3-3 Diagnostic accuracy results – Wang 2017 and post hoc calculations

Author, year	Category	No. studies (No. participants)	Sensitivity		Specificity		NLR[16]	
			FNA% (95% CI) [p value, I^2]	CB% (95% CI) [p value, I^2]	FNA% (95% CI) [p value, I^2]	CB% (95% CI) [p value, I^2]	FNA (95% CI) [p value, I^2]	CB (95% CI) [p value, I^2]
Wang 2017	Overall	12 (1802)	74 (72-77) [<0.001, 88.3]	87 (84-88) [<0.001, 88.5]	96 (94-98) [<0.001, 76.2]	98 (96-99) [0.081, 39]	0.26 (0.19-0.34) [<0.001, 84.1]	0.15 (0.9-0.25) [<0.001, 90.7]
	Palpable	4 (338)	81 (75-86) [<0.001, 84.9]	91 (86-95) [0.004, 77.2]	92 (86-96) [<0.001, 88.6]	98 (93-99) [0.09, 53.8]	0.23 (0.12-0.43) [0.004, 77.2]	0.12 (0.05-0.31) [0.001, 80.9]
	Non-palpable	5 (1057)	68 (64-72) [<0.001, 86.8]	84 (81-87) [<0.001, 94.6]	96 (94-98) [0.105, 47.8]	97 (95-99) [0.164, 38.5]	0.32 (0.22-0.47) [<0.001, 81.7]	0.19 (0.08-0.46) [<0.001, 95]
	Free hand guidance	3 (278)	80 (73-85) [<0.001, 88.6]	93 (88-96) [0.049, 66.9]	88 (79-94) [<0.001, 88.4]	100 (96-100) [1, 0]	0.26 (0.15-0.47) [0.012, 77.5]	0.08 (0.03-0.24) [0.011, 77.8]
	Ultrasound guidance	5 (547)	84 (80-89) [<0.001, 84.1]	88 (84-91) [0.141, 42]	99 (97-100) [0.407, 0]	98 (96-99) [0.184, 35.6]	0.14 (0.07-0.3) [<0.001, 86.5]	0.15 (0.01-0.23) [0.070, 53.8]
	Stereotactic guidance	4 (977)	66 (62-70) [0.047, 62.2]	84 (80-87) [<0.001, 95.7]	97 (94-98) [0.055, 60.6]	97 (93-99) [0.22, 32]	0.38 (0.28-0.52) [0.009, 74.2]	0.23 (0.09-0.63) [<0.001, 95.9]
	Published in 2007 or later	7 (682)	81 (77-84) [<0.001, 84.4]	88 (85-91) [0.079, 47]	99 (97-100) [0.613, 0]	98 (95-99) [0.066, 49.2]	0.20 (0.12-0.33) [<0.001, 84]	0.14 (0.01-0.21) [0.064, 49.6]
	Published before 2007	5 (1112)	70 (66-73) [<0.001, 89.6]	86 (83-88) [<0.001, 95.1]	95 (92-97) [<0.001, 86.1]	98 (96-99) [0.184, 35.6]	0.32 (0.22-0.46) [<0.001, 81.3]	0.17 (0.07-0.43) [<0.001, 95.5]

Source: Wang et al. (2017): Tables 3, 4 and 5, pages 163-164; post hoc analysis for NLR.
Abbreviations: CI, confidence interval; CB, core biopsy; FNA, fine needle aspiration; NLR, negative likelihood ratio.

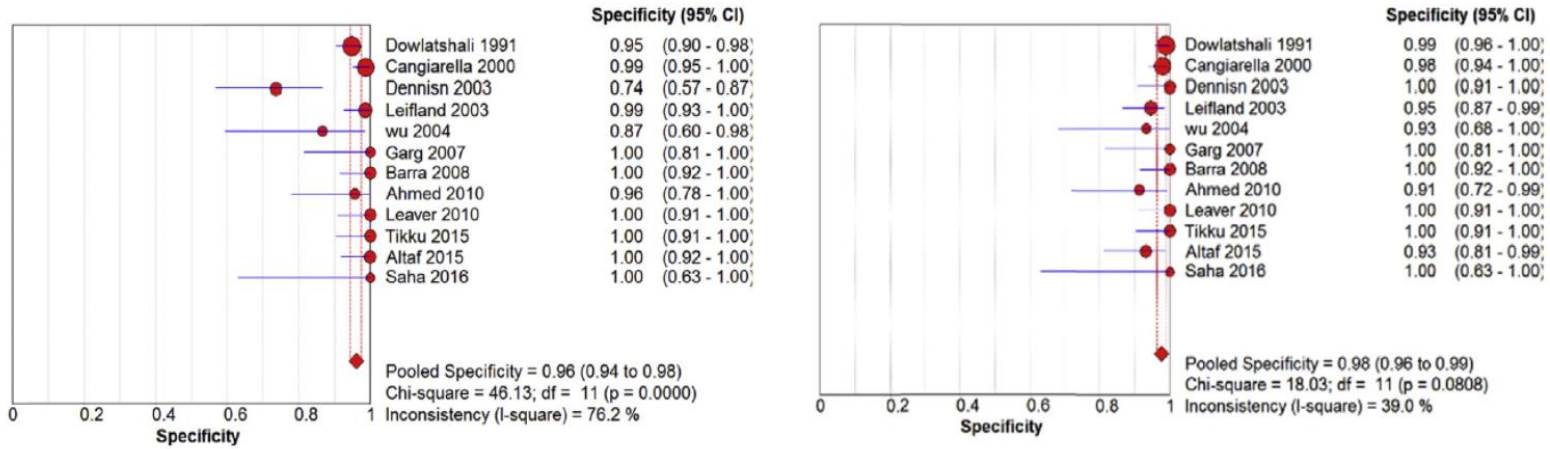
[16] Not reported by Wang 2017. Post hoc calculation for this review using Meta-DiSc from raw data values included in Wang 2017.

Figure 3-1 Sensitivity from Wang 2017 of FNA (left) versus CB (right)



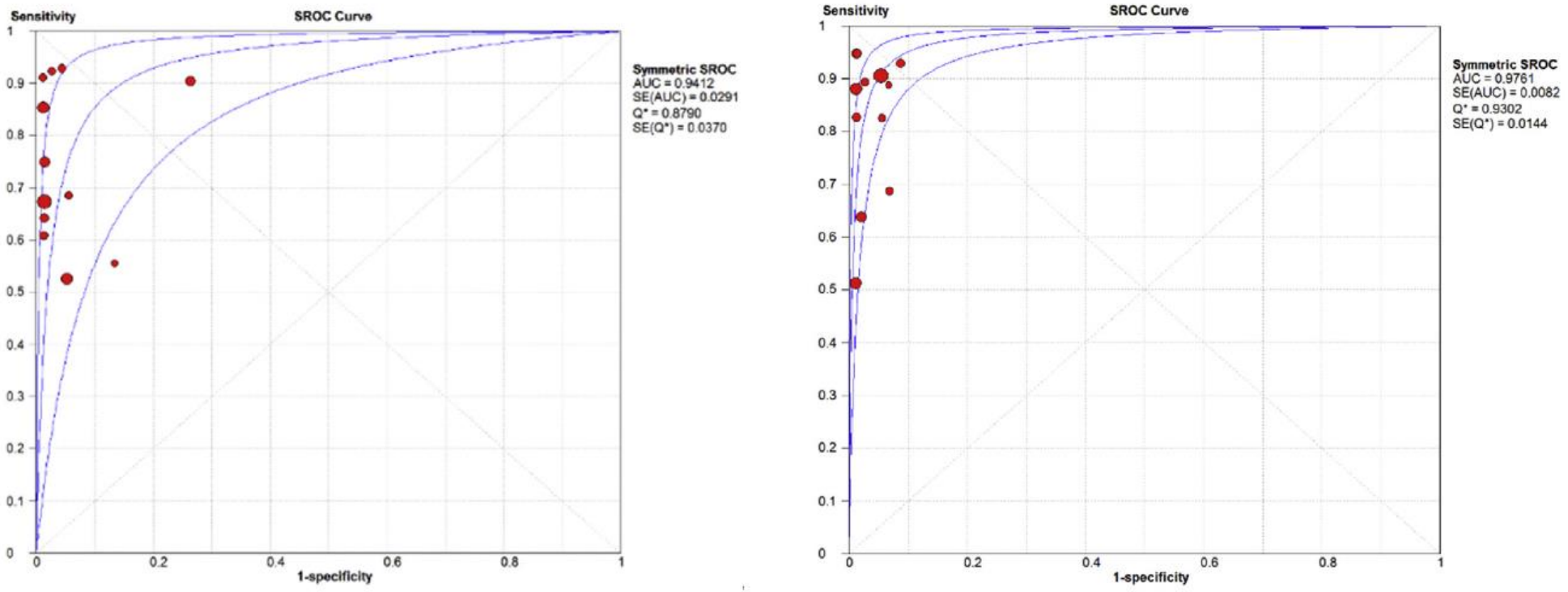
Source: Wang et al. (2017): Figure 5, page 163.

Figure 3-2 Specificity from Wang 2017 of FNA (left) vs CB (right)



Source: Wang et al. (2017): Figure 5, page 163.

Figure 3-3 SROC plots from Wang 2017 for FNA (left) and CB (right)



Source: Wang et al. (2017): Figure 4, page 162.
Abbreviations: AUC, area under the curve; SE, standard error; SROC, summary receiver operating characteristic.

Specificity

The specificity is the ability of a test to correctly identify subjects without the disease. It is the proportion of true negatives that are correctly identified by the test.

As shown in Table 3-3 and Figure 3-2, the pooled specificity of core biopsy (0.98; 95% CI 0.96–0.99) was similar to that of FNA (0.96; 95% CI 0.94–0.98). For each test there was statistically significant heterogeneity in the individual study results for specificity. As for sensitivity, visual inspection of the forest plots supports that there is less variability associated with core biopsy compared with FNA.

Subgroup analyses of specificity based on palpability, guidance method or publication date were also conducted. A lack of overlap in the 95% CIs for FNA and core biopsy was seen for one subgroup only: free-hand guidance (specificity 88% for FNA versus 100% for core biopsy). Similar specificities were seen for FNA and core biopsy in the non-palpable subgroup: 96% and 97%, respectively.

Area under the summary receiver operating characteristic curve

The area under the curve (AUC) of the summary receiver operating characteristic (SROC) curve is a global measure of diagnostic accuracy. A higher AUC indicates higher specificity and sensitivity along all the available cut-offs.

The findings from Wang 2017 are presented in Figure 3-3. The authors note that FNA and core biopsy had similar AUC values (0.94 versus 0.98), although core biopsy had a slightly higher value, indicating it is better at discriminating whether a breast lesion is breast cancer or not.

Negative likelihood ratio

The NLR is the ratio of the probability that a negative result will occur in subjects with the disease to the probability that the same result will occur in subjects without the disease. Therefore, the NLR tells us how much less likely the negative test result is to occur in a subject with the disease than in a healthy subject.

As shown in Table 3-3, the post hoc analysis showed the NLR was lower for core biopsy compared with FNA (0.12 and 0.23, respectively). The NLR was also lower for core biopsy compared with FNA for most subgroups examined. The NLR results for each individual study included in Wang 2017 are presented in Figure 3-4.

Positive predictive value

The positive predictive value (PPV) is the probability that the disease is present when the test is positive.

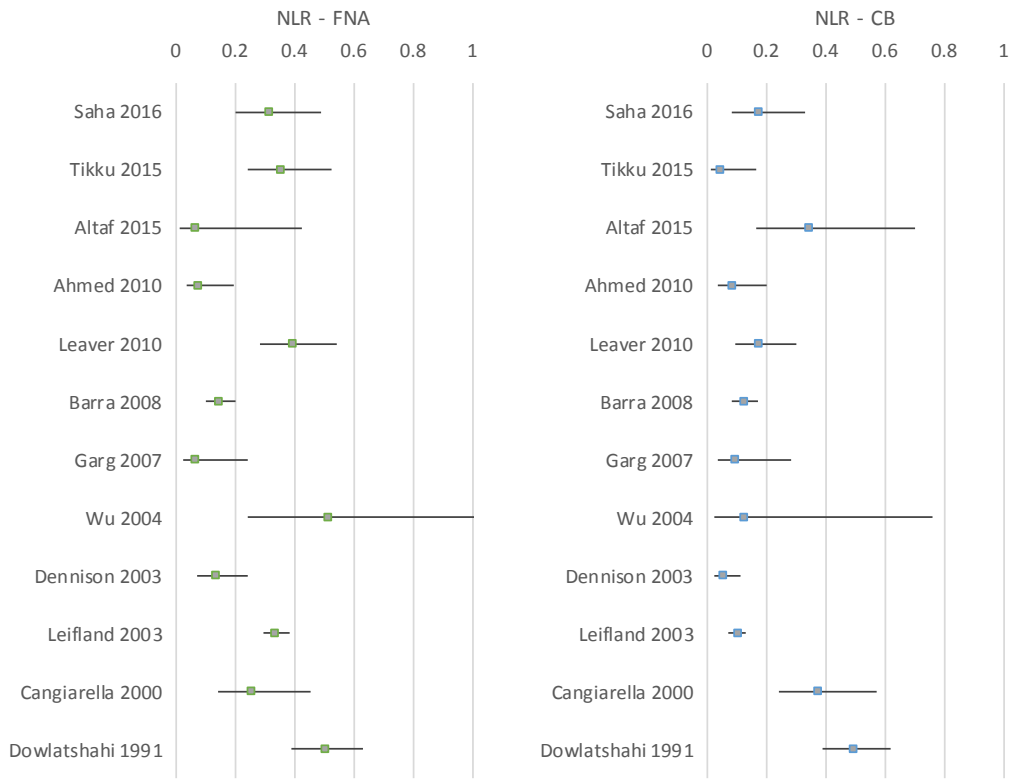
A pooled value for PPV was not reported by Wang 2017. As shown in Figure 3-5, however, the PPVs by study for FNA and core biopsy appeared to be similar.

Negative predictive value

The negative predictive value (NPV) is the probability that the disease is not present when the test is negative.

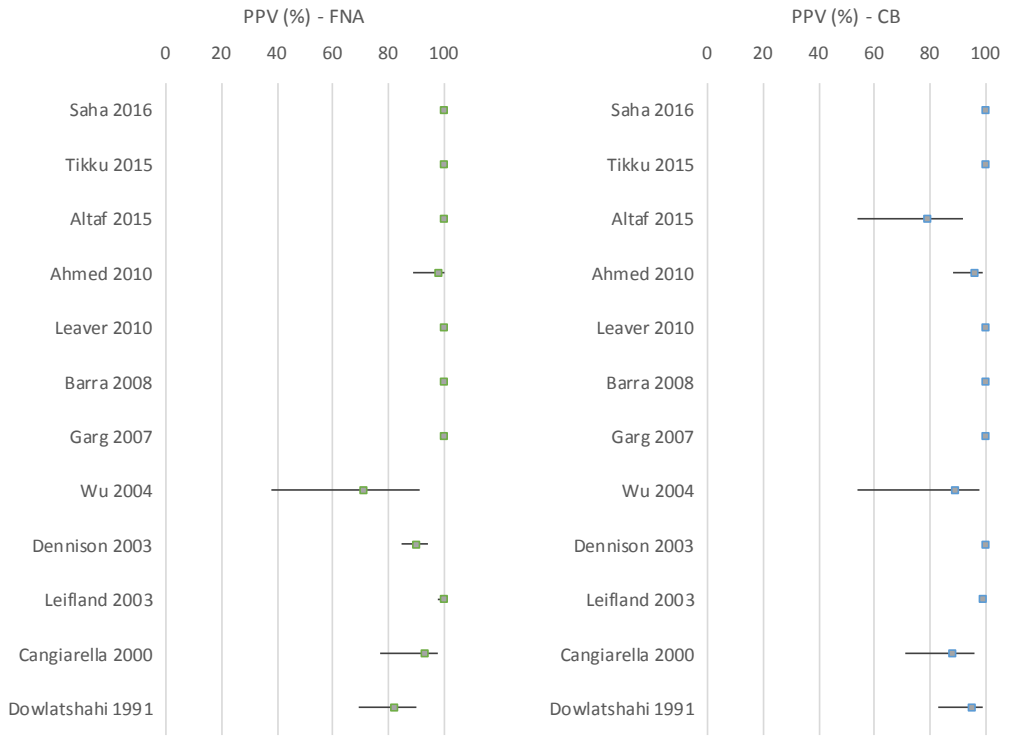
A pooled value for NPV was not reported by Wang 2017. As shown in Figure 3-6, the NPVs by study for FNA and core biopsy appeared to be higher for core biopsy (9/12 studies had an NPV between 0.8 and 1.0) than for FNA (5/12 studies had a NPV between 0.8 and 1.0).

Figure 3-4 Negative likelihood ratios from the individual studies included in Wang 2017



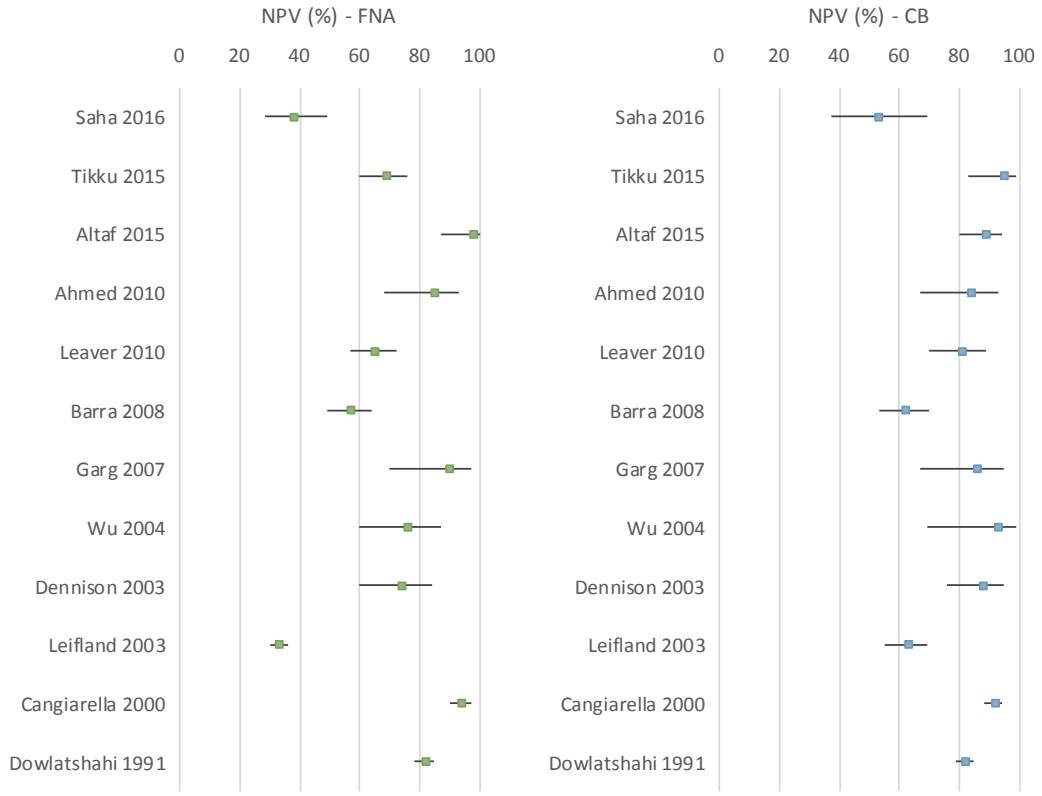
Abbreviations: CB, core biopsy; FNA, fine needle aspiration; NLR, negative likelihood ratio.

Figure 3-5 Positive predictive values from the individual studies included in Wang 2017



Abbreviations: CB, core biopsy; FNA, fine needle aspiration; PPV, positive predictive value.

Figure 3-6 Negative predictive values from the individual studies included in Wang 2017



Abbreviations: CB, core biopsy; FNA, fine needle aspiration; NPV, negative predictive value.

Lieske 2006

The diagnostic accuracy results from the prospective cohort study by Lieske 2006 that was conducted in a population screening setting are presented in Table 3-5 and Table 3-6 (moderate risk of bias for the outcomes below). Lieske 2006 presented data on sensitivity and underestimation rate only. The reporting categories used in the Lieske 2006 study for FNA and core biopsy are presented in Table 3-4.

Table 3-4 Reporting categories for FNA and core biopsy from Lieske 2006

	Cytology reporting		Core biopsy reporting
C1	Unsatisfactory	B1	Unsatisfactory/normal tissue only
C2	Benign	B2	Benign
C3	Atypia probably benign	B3	Benign but of uncertain malignant potential
C4	Suspicious of malignancy	B4	Suspicious of malignancy
C5	Malignant	B5	Malignant B5a Noninvasive cancer B5b Invasive cancer B5c Cancer of non-assessable invasiveness

Abbreviations: B, biopsy; C, cytology.

Sensitivity

The sensitivity results from Lieske 2006 are presented in Table 3-5. Analyses for sensitivity were conducted in two ways:

- Absolute sensitivity – where only C5 and B5 results were included
- Complete sensitivity – where C3/C4/C5 and B3/B4/B5 results were included.

Analyses were conducted for the overall population, as well as for subgroups based on histology (invasive or DCIS only), mammographic presentation (microcalcification or soft tissue lesion) and mode of biopsy (clinical, US or ST guidance).

The results show that when used in a mammography screening population, core biopsy is substantially more sensitive than FNA (93% versus 83%, respectively). This improved sensitivity was consistent across different histology types, mammographic presentations and modes of biopsy. The greatest differences in sensitivity between FNA and core biopsy were seen in the following subgroups: DCIS (74% versus 94%; –20%), lesions presenting as microcalcifications on mammography (75% versus 94%; –19%) and when stereotactic guidance is used (71% versus 95%; –24%). Across the subgroups, sensitivity ranged from 71% to 91% for FNA, and from 93% to 95% for core biopsy. The smallest difference between FNA and core biopsy was for the clinical guidance subgroup (91% versus 95%; –4%).

Table 3-5 Absolute and complete sensitivity from Lieske 2006

	Histology			Mammographic presentation		Mode of biopsy		
	All cancers (Invasive and DCIS)	Invasive only	DCIS only	Microcalcification	Soft tissue lesion	Clinical	US	ST
Absolute sensitivity								
FNA	65%	-	-	-	-	-	-	-
CB	80%	-	-	-	-	-	-	-
Complete sensitivity								
FNA	83	85	74	75	85	91	86	71
CB	93	93	94	94	93	95	94	95

Abbreviations: CB, core biopsy; DCIS, ductal carcinoma *in situ*; FNA, fine needle aspiration; NC, not calculable; ST, stereotactic; US, ultrasound.

Underestimation rate

Underestimation rate, as presented in Lieske 2006, measures the proportion of patients with a benign finding on testing (C2 for FNA or B2 for core biopsy) who were subsequently found to be cancer (invasive and DCIS).

In patients shown to have malignancy on surgical biopsy, FNA was benign (category C2) in 10% of patients compared with 2% for core biopsy (category B2). The results presented in Table 3-6 show that underestimation occurred substantially more frequently for FNA compared with core biopsy across all subgroups examined, and particularly for invasive lobular lesions (19.4% versus 0%), other invasive lesions (12.8% versus 4.7%) and DCIS (13.6% versus 3.4%).

Table 3-6 Underestimation rate

Lieske 2006	Cases of underestimation n/N (%)	
	FNA n/N (%)	CB n/N (%)
Test result showing benign lesion[17]	74/763 (10.0)	17/763 (2.2)
Final histology		
Invasive ductal	NR (5.1)	NR (1.7)
Invasive lobular	NR (19.4)	NR (0)
Other invasive	NR (12.8)	NR (4.7)
DCIS	24/176 (13.6)	6/176 (3.4)
Mammographic presentation		
Microcalcification	29/231(12.5)	8/231(3.5)
Soft tissue lesion	45/532 (8.5)	9/532 (1.7)
Mode of biopsy		
Clinical	10/208 (4.8)	1/86 (1.2)
US	21/270 (7.8)	2/313 (0.6)
ST	43/283 (15.2)	14/359 (3.9)

Abbreviations: CB, core biopsy; DCIS, ductal carcinoma *in situ*; FNA, fine needle aspiration; ST, stereotactic; US, ultrasound.

Bonifacino 2005

The diagnostic accuracy results from the prospective cohort study by Bonifacino 2005 that included a comparison of FNA with VACB are presented below (moderate risk of bias for the outcomes

[17] C2 for FNA or B2 for core biopsy.

below). The reporting categories for FNA and VACB are presented in Table 3-7. Specific detail on how vacuum-assisted biopsy samples were classified as benign or malignant is not provided.

Table 3-7 Reporting categories for FNA and VACB from Bonifacino 2005

FNA reporting	VACB reporting
Non-diagnostic	
Benign	Benign
Atypia probably benign	
Suspicious of malignancy	
Malignant	Malignant

Abbreviations: FNA, fine needle aspiration; VACB, vacuum-assisted core biopsy.

Rate of false-positive tests

As shown in Table 3-8, the rate of false positive tests was similar and low for FNA (0.9%) and VACB (0%).

Sensitivity

The sensitivity results from Bonifacino 2005 are presented in Table 3-8. Analyses of sensitivity for FNA were conducted in two ways:

- Absolute sensitivity – where only malignant results were included
- Complete sensitivity – where malignant, suspicious or indeterminate results were included.

For VACB, analyses of absolute sensitivity only were available, because samples were only classified as malignant or benign.

The results show that the absolute sensitivity is substantially higher for VACB than FNA (97% versus 77%, respectively).

Specificity

As shown in Table 3-8, the specificity of FNA was similar to that of VACB (99% versus 100%, respectively). The authors note that specificity was not impacted by palpability/lesion size.

Negative likelihood ratio

As shown in Table 3-8, the NLR of FNA was higher than that of VACB (0.20 versus 0.03, respectively).

Positive predictive value

For FNA, the PPV for malignant and special or indeterminate findings were calculated separately. As shown in Table 3-8, the PPV was similar for FNA and VACB findings of malignant, while the PPV of an FNA finding of suspicious or indeterminate was 50%.

Negative predictive value

As shown in Table 3-8, the NPV of FNA was lower than that of VACB (94% versus 99%, respectively).

Table 3-8 Diagnostic accuracy outcomes from Bonifacino 2005

	FNA	VACB
False-positive rate	0.9%	0%
Absolute sensitivity	77.1%	97.1%
Non-palpable and/or < 1 cm	67%	-
Palpable or ≥ 1 cm	84%	-

Complete sensitivity	80.0%	-
Non-palpable and/or < 1 cm	70%	-
Palpable or ≥ 1 cm	89%	-
Specificity	99.1%	100%
NLR	0.20	0.03
PPV malignant	100%	100%
PPV suspicious or indeterminate	50.0%	-
NPV	94.0%	99.1%

Abbreviations: NLR, negative likelihood ratio; NPV, negative predictive value; PPV, positive predictive value; VACB, vacuum-assisted core biopsy.

Underestimation rate

As shown in Table 3-9, five of the FNA benign findings were shown to be malignant based on the reference standard, resulting in an underestimation rate of 3.4%. This is compared with the VACB findings where only one case was underestimated (underestimation rate 0.7%).

Table 3-9 Underestimation – FNA or VACB versus reference standard

	Reference standard result	
	Benign	**Malignant**
FNA		
Non-diagnostic	27	2
Benign	83	5
Atypia/probably benign	0	0
Suspicious	1	1
Malignant	0	27
VACB		
Benign	111	1
Malignant	0	34

Abbreviations: FNA, fine needle aspiration; VACB, vacuum-assisted core biopsy.
Note: Lesions underestimated by FNA and VACB are shown in grey shading.

3.1.2 Clinical impact of diagnostic testing

3.1.2.1 Study characteristics

The literature search identified seven studies that provided the highest available evidence of the clinical impact of FNA versus core biopsy. The characteristics of the included studies are presented in Table 3-10.

To aid the analysis of data presented in the individual clinical studies, further information has been extracted from these studies, particularly relating to the additional considerations raised in the PICOS, as shown in Table 3-11. This data provides further detail on patient characteristics, biopsy methodology and study design to ensure there is adequate context to draw appropriate conclusions from the outcome data presented in these studies.

Table 3-10 Clinical impact of testing – key characteristics of included studies

Author and year *RoB*[18]	Study type NHMRC LoE	Country	Sample size and characteristics	Intervention vs. comparator	Clinical impact outcomes reported
Management outcomes					
Mak 2012 Low	Retrospective diagnostic cohort NHMRC Level III[19]	Hong Kong	N=208 Women who underwent both ultrasound-guided FNA and core biopsy of the same solid breast lesion from 1 January 2007 to 31 December 2009	FNA vs. core biopsy	Potential for repeat testing (via inadequate sampling)
Hukkinen 2008 High	Retrospective cohort NHMRC Level III-2(iii)	Finland	N=632 Patients with dubious, suspicious, or malignant lesions who underwent surgery at the Breast Surgery Unit of Helsinki University Central Hospital between February 1, 2005 and January 30, 2006	FNA vs. core biopsy	Diagnostic work-up, delay in cancer surgery, cost aspects
Lieske 2006 Low (for this outcome)	Prospective diagnostic cohort study NHMRC Level II[19]	England	N=763 Women attending mammographic screening who had both FNA and CB performed at the preoperative assessment and had malignancy confirmed by subsequent surgery	FNA vs. core biopsy	Potential for repeat testing (via inadequate sampling)
Bonifacino 2005 Low (for this outcome)	Prospective diagnostic cohort study NHMRC Level II[19]	Italy	N=135 (146 lesions) Women attending a single centre in Rome between September 1998 and June 2000 who in whom second-level examinations were absolutely necessary (i.e. controversial diagnosis with mammography and US or US only and < 1.5 cm, deep and non-palpable > 1.5 cm or difficult to sample by core biopsy)	FNA vs. VACB	Potential for repeat testing (via inadequate sampling)
Adverse event outcomes					
Wong 2009 Moderate	Retrospective cohort NHMRC Level III-2(iii)	Hong Kong	N=45 (29 relevant) Patients referred for ultrasonography-guided fine needle aspiration or biopsy	FNA vs. core biopsy	Adverse events: procedure-associated pain, procedural complications
Satchithananda 2005 High	Prospective cohort NHMRC Level III-2(ii)	United Kingdom	N=220 Patients undergoing a breast biopsy procedure	FNA vs. core biopsy	Adverse events: Pain level
Effectiveness outcomes					

[18] See Appendix C.
[19] Based on diagnostic accuracy hierarchy.

FNA and core biopsy: High level review of evidence and clinical practice guidelines

Author and year *RoB*[18]	Study type NHMRC LoE	Country	Sample size and characteristics	Intervention vs. comparator	Clinical impact outcomes reported
Liikanen 2016 Low for overall survival; high for other outcomes	Prospective cohort NHMRC Level III-2(ii)	Finland	N=1,525 Patients who had a tumour size below 2cm in diameter who had received surgery for breast cancer at the Breast Surgery Unit of the Helsinki University Central Hospital between February 2001 and August 2005	FNA vs. core biopsy	Effectiveness: Overall survival/death from any cause, breast cancer death, distant metastases, regional lymph node recurrence, local recurrence, local recurrence after breast conserving surgery, local recurrence after mastectomy

Abbreviations: FNA, fine needle aspiration; LoE, level of evidence; NHMRC, National Health and Medical Research Council; RoB, risk of bias; VACB, vacuum-assisted core biopsy.

Table 3-11 Additional study considerations – included studies

Author, year	Study type	N	Mean age (Y)	Lesion size (mm) or palpable vs. non-palpable	Index test guidance		Number of passes/samples per patient		Size of needle		Duration of follow up	Adjustment for potential confounding
					FNA	CB	FNA	CB	FNA	CB		
Management outcomes												
Mak, 2012	Retrospective diagnostic cohort	208	NR	Small palpable (< 1.5 cm) and non-palpable	UG	UG	NR	NR	21-gauge	14-gauge	NR	NA
Hukkinen, 2008	Retrospective cohort	632	59 (median)	Palpable and non-palpable	Free-hand, UG	Free-hand, UG	>5	≥3	20-gauge	14-gauge	NR	None
Lieske 2006	Prospective diagnostic cohort	763	NR	Identified via mammography screening	Free-hand, UG or SG	Free-hand, UG or SG	1-2 (free-hand or US); 5 (SG)	2-3 (free-hand of US); 5+ (SG)	21-gauge	14-gauge	Until surgery	NA
Bonifacino 2005	Prospective diagnostic cohort	135 (146 lesions)	50	Palpable and non-palpable	UG	VACB: UG	3+	VACB: 4-10	21 to 23-gauge	VACB: 11-gauge	Mean 39 months	NA
Adverse event outcomes												
Wong, 2009	Retrospective cohort	29	49	≤ 8 mm (FNA), > 8 mm (CB)	UG	UG	3-5	3-5	21-gauge	14-gauge	Up to 3 days	None
Satchithananda, 2005	Prospective cohort	220	NR	NR	UG	UG or SG (various subtypes)	NR	NR	21-gauge	14-gauge/11-gauge for VACB	Immediately following procedure	None
Effectiveness outcomes												
Liikanen, 2016	Prospective cohort	1,525	58 (median)	Palpable and non-palpable	UG or SG	UG or SG	NR	NR	NR	NR	Median 9.5 y	Multivariate analysis of overall survival only adjusting for age, tumour size, pN category, palpability, histological grade, ER status and MIB-1 expression

Abbreviations: CB, core biopsy; ER, oestrogen receptor; FNA, fine-needle aspiration; NR, not reported; pN, regional lymph node involvement; SG, stereotactic-guided; UG, ultrasound-guided, VACB, vacuum-assisted core biopsy; Y, years.

3.1.2.2 Study results

The results for the seven included studies are presented by outcome below. The following clinical impact-associated outcomes were specified for inclusion in the review: management outcomes (repeat testing, cascade ancillary testing, change in surgical technique or treatment decisions), adverse events (complications associated with biopsy techniques – including pain, bleeding and hospitalisation), quality of life (using validated instruments, measures of worry and anxiety related to breast cancer risk) and effectiveness (recurrence, progression-free survival, overall survival). Data was available for most outcomes with the exception of the following:

- Change in surgical technique or treatment decisions
- Unnecessary surgery
- Quality of life
- Worry or anxiety related to breast cancer risk
- Progression-free survival.

Management outcomes

Repeat tests

Inadequate/unsatisfactory sampling refers to the inability of an initial biopsy (FNA or core biopsy) to produce an appropriate sample to test, impacting on the follow-on management of those receiving a biopsy procedure. Mak 2012 notes that inadequate sampling often results in the repeat of a tissue biopsy, incurring additional costs and exposure to the risks associated with the procedure.[20]

No studies were identified that specifically addressed repeat testing. However, three studies examining inadequate/unsatisfactory sampling were identified (Mak 2012, Lieske 2006 and Bonifacino 2005; all low risk of bias for this outcome).

Mak 2012 performed a retrospective diagnostic cohort study in 208 women who received both FNA and core biopsy for their breast lesions from January 2007 to December 2009. Mak 2012 did not report lesion characteristics prior to FNA or core biopsy.

Data reported by Mak 2012 comparing inadequate/unsatisfactory sampling of FNA and core biopsy is presented in Table 3-12. Although statistical analysis was not conducted by the authors, the results indicate a substantially greater proportion of inadequate samples following FNA (16.3%) compared with core biopsy (1.4%). Mak 2012 conclude that the superiority of core biopsy in obtaining adequate tissue samples must be considered when choosing between these two biopsy techniques.

As noted previously, Lieske 2006 performed a prospective diagnostic cohort study in patients undergoing mammographic screening, which makes it more relevant to the current review. It should be noted that this data only includes those lesions tested that were ultimately found to be malignant via surgical biopsy. FNA was unsatisfactory (category C1) in 8% of patients compared with 5% non-diagnostic (unsatisfactory) for core biopsy (category B1). The results presented in Table 3-12 show that inadequate/unsatisfactory sampling occurred more frequently for FNA compared with core biopsy, with the biggest differences occurring for DCIS, lesions presenting as microcalcifications on mammogram, and when stereotactic guidance was used.

[20] "Inadequate" samples can be immediately rectified by repeat FNA or immediate core biopsy.

Bonifacino 2005 performed a prospective diagnostic cohort study comparing FNA with VACB. As shown in Table 3-12, 20% of FNA samples taken were considered to be inadequate; none were reported for VACB.

Table 3-12 Inadequate/unsatisfactory sampling

Author, year	Cases of inadequate/unsatisfactory sampling n/N (%)	
	FNA n/N (%)	CB n/N (%)
Mak 2012	34/208 (16.3)	3/208 (1.4)
Lieske 2006	C1[21] 61/763 (8.0)	B1[21] 36/763 (4.7)
Final histology		
Invasive ductal	NR (5.6)	NR (5.6)
Invasive lobular	NR (7.5)	NR (3.2)
Other invasive	NR (7.0)	NR (3.5)
DCIS	24/176 (13.6)	6/176 (3.4)
Mammographic presentation		
Microcalcification	29/231 (12.5)	6/231 (2.6)
Soft tissue lesion	32/532 (6.0)	30/532 (5.6)
Mode of biopsy		
Clinical	8/208 (3.8)	3/86 (3.5)
US	16/270 (5.9)	18/313 (5.8)
ST	37/283 (13.1)	14/359 (3.9)
Bonifacino 2005	29/146 (20.0)	VACB: 0/146 (0)

Source: Mak 2012: Figure 1, page 12; Lieske 2006: Table 5, page 64; Bonifacino 2005: Table IV and Table V, page 2467.
Abbreviations: B, biopsy; C, cytology; CB, core biopsy; DCIS, ductal carcinoma *in situ*; FNA, fine needle aspiration; ST, stereotactic; US, ultrasound; VACB, vacuum-assisted core biopsy.

Follow-on testing

Follow-on testing is the sequencing of tests to diagnose a disease. Follow-on testing is appropriate when a 'gold standard' method is technically demanding and/or costly and the diagnosis can usually be established by simpler or more cost-effective strategies. The sequence and number of additional tests may vary depending on the initial test performed and has a potential impact on the timing of diagnosis and management, cost and effectiveness.

Follow-on testing/diagnostic work-up is presented in Hukkinen 2008. Diagnostic workups in 572 patients with 580 breast lesions were retrospectively examined. Of these lesions, the initial test was FNA for 339 and core biopsy for 241.

This study was considered to be at a high risk of bias due to the fact that lesions were specifically selected for initial testing with FNA or core biopsy (i.e. both tests were not conducted in each lesion). Whilst this selection makes clinical sense, it may lead to selection bias from a methodological assessment perspective where the actual characteristics of the lesions initially tested with a certain modality may make them more or less susceptible to further testing down the track, rather than the further testing being related only to the type of testing method used. The characteristics of the lesions that received FNA or core biopsy are presented in Table 3-13. There were some differences in the types of lesions that received initial testing with FNA or core biopsy: lesions selected for initial FNA were more likely to be small (< 10 mm) and be invisible on

[21] See Table 3-4 for the definitions of C1 and B1.

mammography, while lesions selected for initial core biopsy were more likely to show mass plus calcification on mammography. The postoperative diagnoses following initial FNA or core biopsy were generally similar with the exception of invasive lobular carcinoma, which was identified more frequently following core biopsy than FNA.

Table 3-13 Characteristics of lesions initially tested with FNA or core biopsy

	FNA (n=339)	Core biopsy (n=241)	p-value
Clinical feature			
Symptomatic	60%	56%	0.05
Palpable	70%	69%	0.85
Median size on imaging	15 mm	20 mm	
Size < 10 mm	**22%**	**12%**	**0.005**
Range of size on imaging	3-80 mm	5-60 mm	
Finding in mammogram			
Mass	72%	70%	0.58
Mass + calcification	**5%**	**14%**	**<0.001**
Architectural distortion	6%	8%	0.31
Microcalcification	3%	4%	0.47
Invisible	**9%**	**2%**	**0.002**
Finding in ultrasound			
Benign in ultrasound	9%	8%	0.66
Intermediate on ultrasound	12%	8%	0.26
Malignant on ultrasound	78%	83%	0.20
Invisible on ultrasound	0.6%	0.4%	1
Suspicious lymph nodes	7%	4%	0.15
Postoperative diagnosis			
Benign	15%	11%	0.26
DCIS	2%	4%	0.45
IDC	57%	51%	0.18
ILC	**17%**	**27%**	**0.007**
Other invasive cancers	8%	7%	0.53

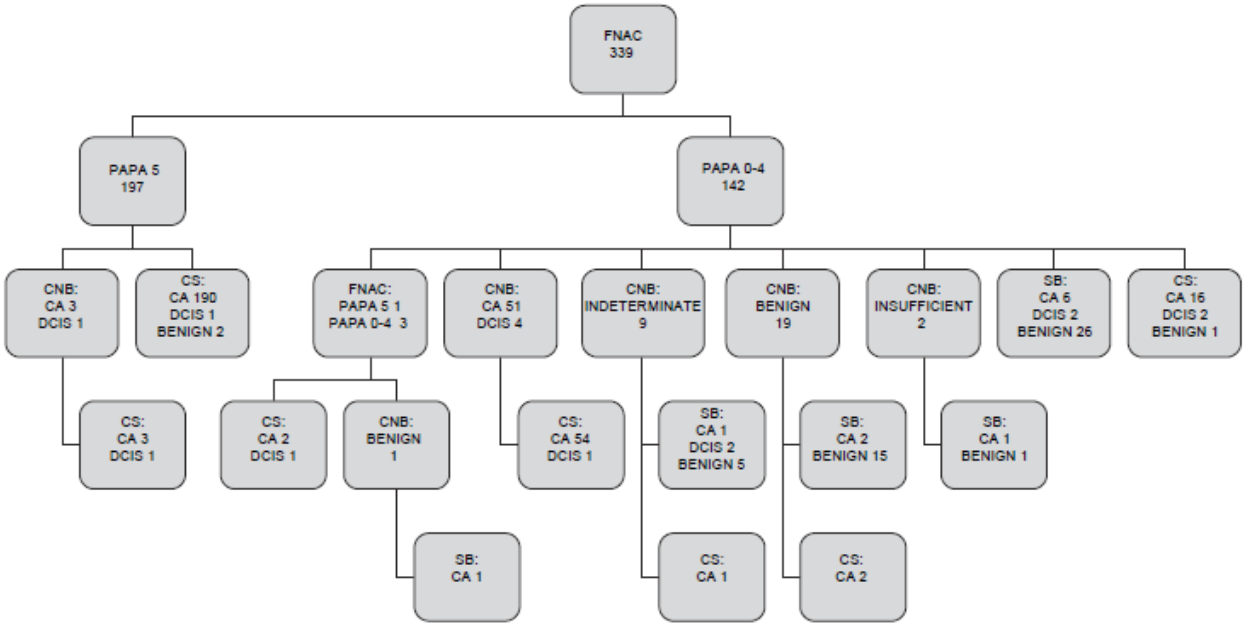
Source: Hukkinen 2008: Table I, page 1039.
Abbreviations: DCIS, ductal carcinoma *in situ*; FNA, fine needle aspiration; IDC, invasive ductal carcinoma; ILC. Invasive lobular carcinoma.

Flow diagrams showing the downstream testing following FNA or core biopsy are presented in Figure 3-7 and Figure 3-8, respectively. As can be seen from the diagrams, a greater amount of testing was required following FNA (0.46 additional tests per lesion initially tested with FNA), compared with core biopsy (0.15 additional tests per lesion initially tested with core biopsy). Therefore, approximately one in two lesions initially tested with FNA required additional testing compared with one in seven lesions initially tested with core biopsy. The authors note that "the frequent need for additional biopsies raised the total expenses of [FNA] over those of [core biopsy]."

Hukkinen 2008 also note that the requirement for additional biopsies may delay surgery. Following both initial FNA or core biopsy, the median time to definitive surgical treatment was 31days. Following initial FNA and additional testing, the median time to definitive surgical treatment was 42 days for those undergoing additional needle biopsy (N=69), 63 days for those requiring surgical biopsy (N=6) and 67 days for those requiring both additional needle biopsy and surgical biopsy (N=4). Following initial core biopsy, the median time to definitive surgical treatment was 13 days for the single lesion undergoing additional needle biopsy and 51 days for those undergoing surgical biopsy (N=4).

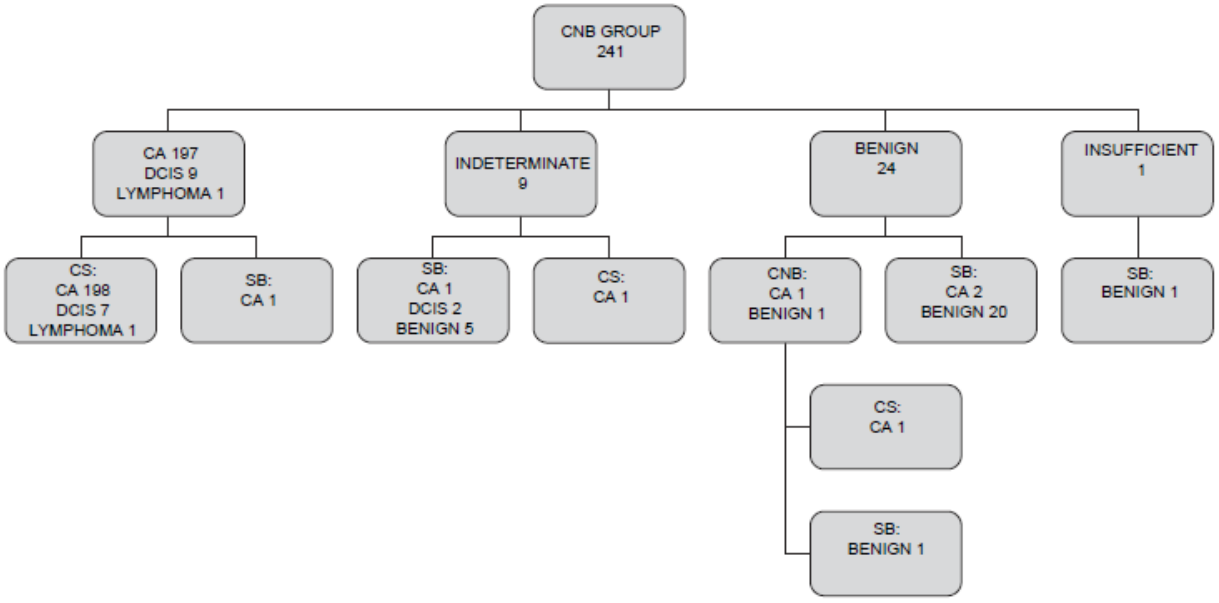
The findings from this study should be viewed with caution given the high risk of bias associated with this study.

Figure 3-7 Follow-on testing following FNA as first biopsy



Source: Hukkinen et al. (2008): Figure 1, page 1042.
Abbreviations: CA, invasive cancer; CNB, core needle biopsy; CS, cancer surgery; DCIS, ductal carcinoma *in situ*; FNAC, fine needle aspiration cytology; PAPA, Papanicolau grading of cytology;[22] SB, surgical biopsy.

Figure 3-8 Follow-on testing following core biopsy as first biopsy



Source: Hukkinen et al. (2008): Figure 1, page 1042.
Abbreviations: CA, invasive cancer; CNB, core needle biopsy; CS, cancer surgery; DCIS, ductal carcinoma *in situ*; FNAC, fine needle aspiration cytology; SB, surgical biopsy.

[22] 0 = insufficient; 1 = normal; 2 = probably benign; 3 = mildly suspicious; 4 = strongly suspicious of malignancy; 5 = malignant.

Adverse events outcomes

Complications associated with biopsy technique

Pain associated with both biopsy techniques was measured in Wong 2009 and Satchithananda 2005 and data from these studies is shown in Table 3-14. Satchithananda 2005 included data on both core biopsy and VACB.

The study by Satchithananda 2005 was a prospective cohort study in which all patients undergoing a breast biopsy over a 3-month period were asked to complete a questionnaire regarding their experience immediately following their procedure. The questionnaire included an adapted visual analogue scale (VAS) to rate the pain from the procedure from 0 (no pain) to 10 (worst pain imaginable). An additional question asked how the breast biopsy compared with other procedures (blood test, blood pressure measurement and cervical smear), with ratings of 'much better', 'better', 'same', 'worse' and 'much worse'. A total of 220 image-guided biopsies were undertaken during this period; the procedures included ultrasound (US)-FNA, stereotactic VACB, US core biopsy and 'other'. The results from the Satchithananda 2005 study are considered to be at a high risk of bias for two reasons: (i)different populations of patients (with likely different tumour characteristics that may result in more or less pain) underwent the different testing methods, and (ii) all patients undergoing core biopsy received anaesthesia, compared with only some patients undergoing FNA.

The mean pain score for US-FNA was approximately 4 compared with 2 for US-core biopsy (mean difference 2.0; 95% CI 0.8–3.3; p=0.001). The mean difference in pain score between US-FNA and stereotactic VACB was 1.54 (95% CI 0.3–2.8; p=0.01). While these findings suggest that FNA is associated with a significantly greater level of pain, it should be noted that in the centre where the study was conducted, local anaesthesia is always used for core biopsy and VACB but not always for FNA.

The study by Wong 2009 was a retrospective cohort study in which patients referred for US-guided FNA or core biopsy between November 2007 and March 2008 were divided into three groups based on lesion size: breast lesions ≤ 8 mm received FNA while breast lesions > 8 mm were randomised to either traditional core biopsy or coaxial biopsy. Only patients undergoing FNA or traditional core biopsy are considered here. The results from the Wong 2009 study are considered to be at a moderate risk of bias due to the different populations of patients undergoing different testing methods; however, unlike the Satchithananda 2005 study, all procedures were carried out using local anaesthesia.

Following the biopsy, patients were asked to assign a pain score from 0 (no pain) to 10 (most severe pain). The resulting mean pain scores were 3.8 for FNA and 3.7 for core biopsy. No procedural immediate or delayed complications were documented for either biopsy technique.

Table 3-14 Pain following FNA or core biopsy

Author, year	Pain scale	Local anaesthesia		Mean pain score	
		FNA	CB/VACB	FNA	CB/VACB
Satchithananda 2005	Fixed interval pain scale in the form of an adapted horizontal visual analogue scale (0: no pain, 10: the worst pain imaginable)	Maybe	Yes/Yes	~4	~2/~2.5
Wong 2009	Pain score (0: no pain, 10: most severe pain)	Yes	Yes/NA	3.8	3.7/NA

Source: Satchithananda 2005: estimated from Figure 4, page 400; Wong 2009: Table, page 247.
Abbreviations: CB, core biopsy; FNA, fine needle aspiration; NA, not applicable; VACB, vacuum-assisted core biopsy.

Effectiveness outcomes

Survival

Liikanen 2016 performed a prospective cohort study to examine the association between biopsy method and breast cancer outcome. A total of 1,851 patients who had surgery for breast cancer (tumour ≤ 2 cm in diameter) were considered for inclusion; of these 340 were excluded from the analysis.[23] Of the remaining included patients, 868 received FNA and 657 received core biopsy. 337 patients who had both procedures were analysed as having had core only. The median follow-up period was 9.5 years after breast surgery. During the follow-up 191 deaths occurred, of which 65 were considered breast cancer specific.

The main characteristics of the lesions undergoing FNA and core biopsy are summarised in Table 3-15. Lesions tested with FNA were slightly larger than those tested with core biopsy and were more likely to be palpable and have a ductal histology.

Table 3-15 Characteristics of lesions tested with FNA and core biopsy: Liikanen 2016

	FNA N=868	Core biopsy N=657	P value
Tumour size (median)	14 mm	12 mm	<0.001
Palpability			
Palpable	72%	57%	<0.001
Non-palpable	28%	43%	
Histology			
Ductal	69%	59%	<0.001
Lobular	17%	21%	
Other	14%	20%	
Multifocality			
Unifocal	83%	83%	0.87
Multifocal	17%	17%	

Source: Liikainen 2016: Table 1, page 66.
Abbreviations: FNA, fine needle aspiration.

Despite patients not being randomly allocated to FNA or core, the all-cause survival analysis for this study is considered to be at a moderate risk of bias. This is due to the use of multivariate analysis which adjusted for characteristics shown to be significant in univariate analyses, including age, tumour size, pN category, palpability, histological grade, oestrogen receptor (ER) status and MIB-1 gene expression. However, the analysis of breast cancer death was not adjusted for potential confounding and is considered to be at a high risk of bias.

The findings of the survival analyses are presented in Table 3-16. While the unadjusted analyses suggest death from any cause is statistically significantly greater following FNA compared with core biopsy, the multivariate-adjusted analysis suggests no difference in overall survival. The authors concluded that FNA did not have an adverse impact on survival (all-cause mortality or breast cancer death) compared with core biopsy. By contrast, we do not believe it is appropriate to make a conclusion of no difference based on data from a single study with insufficient precision to provide meaningful results.

[23] Reasons for exclusion: surgical biopsy or missing biopsy information (N=83), no axillary surgery (N=2), contralateral breast cancer (N=190), history of other malignancy (N=25), distant metastases at presentation (N=7), died of myocardial infarction immediately after surgery (N=1) and lost to follow-up (N=32).

Table 3-16 Survival

Author, year *RoB*	Outcome	FNA n/N (%)	CB n/N (%)	Univariate (unadjusted) analysis	Multivariate (adjusted) analysis
				RE (95% CI) or P value	RE (95% CI) or P value
Liikanen, 2016	Death from any cause/overall survival	129/868 (14.9)	62/657 (9.4)	NR P=0.003	HR 0.94 (0.69, 1.30)[24] P=0.72
	Breast cancer death	40/868 (4.6)	25/657 (3.8)	NR P=0.46	-

Source: Liikanen 2016: Table 2, page 67 and Table 3, page 68.
Abbreviations: CB, core biopsy; CI, confidence interval; FNA, fine needle aspiration; HR, hazard ratio; NR, not reported; RE, risk estimate.

Recurrence

Based on the prospective cohort study described above, Liikanen 2016 (high risk of bias for these outcomes) also reported the results of univariate (unadjusted) analyses of distant metastasis, regional lymph node recurrence and local recurrence in patients who had received FNA or core biopsy. Results for this outcome are presented in Table 3-17. The unadjusted analyses indicate that there is no significant difference in distant metastases or recurrence rates between FNA and core biopsy; however, it should be noted that this finding may be biased due to the lack of adjustment for potential confounding and are subject to a lack of precision.

Table 3-17 Recurrence

Author, year	Outcome	FNA n/N (%)	CB n/N (%)	Univariate (unadjusted) analysis	Multivariate (adjusted) analysis
				RE (95% CI) or P value	RE (95% CI) or P value
Liikanen, 2016	Distant metastasis	67/868 (7.7)	40/657 (6.0)	NR 0.18	-
	Regional lymph node recurrence	11/868 (1.3)	7/657 (1.0)	NR 0.68	-
	Local recurrence	34/868 (3.9)	25/657 (3.8)	NR 0.81	-
	Local recurrence after BCS	27/868 (3.1)	19/657 (2.9)	NR	-
	Local recurrence after mastectomy	7/868 (0.8)	6/657 (0.9)	NR	-

Source: Liikanen 2016: Table 2, page 67.
Abbreviations: BCS, breast cancer surgery; CB, core biopsy; CI, confidence interval; FNA, fine needle aspiration; NR, not reported; RE, risk estimate.

[24] Adjusted for age, tumour size, pN category, palpability, histological grade, ER status and MIB-1 expression.

3.2 Review of clinical guidance

The literature search identified 1688 unique records, of which eleven met the eligibility criteria for the review.

3.2.1 Australian guidance

The identified Australian clinical guidance was recently developed by BreastScreen Australia and Cancer Australia, as shown in Table 3-20. Only the guidance from BreastScreen Australia was specifically targeted to the screening setting. The relevant recommendations/guidance are presented in Table 3-19; both suggest that core biopsy is the preferred test for examining suspicious breast lesions.

Table 3-18 Included Australian guidance

Ref ID	Title	Developer/affiliation	Brief description of methodology
BSA 2017[25]	Use of Fine Needle Aspiration in BreastScreen Australia Services. Endorsed BreastScreen Australia Clinical Advisory Committee Response to: Guidance on the appropriate clinical use of Fine Needle Aspiration (FNA) for the biopsy evaluation of women in the BreastScreen Australia program	BreastScreen Australia	Informed by a review of the evidence including medical literature from PubMed and review of the NHS 2016 guidance
CA 2017[26]	The investigation of a new breast symptom: a guide for General Practitioners	Cancer Australia	Informed by a review of the evidence and expert consensus opinion

Abbreviations: BSA, BreastScreen Australia; CA, Cancer Australia; FNA, fine needle aspiration; NHS, National Health Service.
Note: Guidelines shown in shading are related to screening.

Table 3-19 Australian guidance: relevant recommendations/statements

Ref ID	Guidance type	Recommendation/guidance	Level of Evidence Grade	Evidence base (study ID/reference)		
				Clinical studies	Guidelines	Other
BSA 2017	Consensus based on review of evidence from PubMed and NHS 2016 guideline	"The use of FNA in the screening setting should be limited to cysts, lymph nodes and the rare situations where core biopsy is not possible."	NR	NR	NR	NR
		"Where possible core biopsy, including vacuum assisted biopsy, should be the procedure of choice."	NR	NR	NR	NR
		"For multifocal lesions, if the lesions are in more than one quadrant, at least two quadrants should have a core biopsy."	NR	NR	NR	NR
CA 2017	Consensus-based on a review of evidence	"Core biopsy is preferable for the investigation of suspicious lesions or when additional information, such as tumour type, histological grade and receptor status of cancer is required. This information is particularly important for patients who may be considered for neoadjuvant chemotherapy."	NR	NR	NR	NR
		"Core biopsy can differentiate between *in situ* and invasive cancer whereas FNA cytology cannot."	NR	NR	NR	NR
		"FNA cytology may be used to confirm the diagnosis of a cystic lesion or fibroadenoma identified by diagnostic imaging in centres with cytopathological expertise."	NR	NR	NR	NR

[25] Available at: http://cancerscreening.gov.au/internet/screening/publishing.nsf/Content/cac-Use-of-Fine-Needle-Aspiration.
[26] Available at: https://canceraustralia.gov.au/system/tdf/publications/investigation-new-breast-symptom-guide-general-practitioners/pdf/2017_inbs_gp_card.pdf?file=1&type=node&id=3439.

Ref ID	Guidance type	Recommendation/guidance	Level of Evidence Grade	Evidence base (study ID/reference)		
				Clinical studies	Guidelines	Other
		"Fluid aspirated from a cyst for diagnostic purposes should be sent for pathological examination."	NR	NR	NR	NR

Abbreviations: BSA, BreastScreen Australia; CA, Cancer Australia; FNA, fine needle aspiration; NHS, National Health Service; NR, not reported.
Note: Guidelines shown in shading are related to screening.

3.2.2 International guidance

The identified international clinical guidance consisted of eight clinical practice guidelines and one position paper, published from 2008 onwards, as shown in Table 3-20. Only one of the identified clinical practice guidelines (NHS 2016; United Kingdom), and the Position Statement (EUSOBI 2017; Europe) specifically related to breast cancer screening; however, there are also recommendations in the NCCN 2018 guideline that are related to use of core biopsy following mammographic evaluation (which encompasses both screening and non-screen diagnostic settings).

The relevant recommendations on the use of FNA and core biopsy in the diagnostic setting are presented in Table 3-21. From the nine identified international guidance documents, seven are evidence-based and two appear to be consensus-based but this is not specified. All guidance considered evidence from different sources. Apart from one source (MOHM 2013), the guidance is not based on articles included in the systematic review by Wang et al (2017) and does not appear to be based on comparative studies.

There is a mix of support presented in the guidance documents. Consistent with the BreastScreen recommendations, in both international guidance documents that relate to the use of testing in screening for breast cancer, one source recommends the use of core biopsy over FNA (NHS 2016), and one source presents a preference towards core biopsy (EUSOBI 2017). The NCCN 2018 recommendations following mammographic evaluation (which may refer to screening or the diagnostic setting) suggest only the use of core biopsy. The lack of cytology use in the UK in general (and concomitant lack of experienced cytopathologists) means that some international guidelines are of limited relevance for the Australian setting. The situation is similar in many parts of the USA.

In the other guidance identified (which relates to the use of testing for diagnostic purposes) two clearly recommend the use of core biopsy over FNA (ESMO 2015; Albert 2008), and two recommend core biopsy in the intervention pathway and make no mention of FNA (NCCN 2018; AGO 2013). Further, one of these sources states that FNA cannot be recommended as a standard method (Albert 2008). In contrast, three sources present no clear preference between core biopsy and FNA and advise that the choice depends on the level of expertise available (KCE 2013; JBCS 2013; MOHM 2010). Although not specifically stated, the Malaysian guidelines appear to preference the use of FNA over core biopsy as the initial method of pathological assessment for palpable breast lumps where facility and expertise are available. They also recommended that the use of core biopsy should be in combination with FNA.

Table 3-20 Included international clinical guidance

Ref ID	Title	Developer/affiliation	Brief description of methodology
Asia			
JBCS 2015	The Japanese Breast Cancer Society Clinical Practice Guideline for pathological diagnosis of breast cancer	Japanese Breast Cancer Society (JBSC)	Not reported
MOHM 2010	Management of breast cancer – 2nd edition	Ministry of Health Malaysia (MOHM)	Literature search using PUBMED/MEDLINE, Cochrane Database of Systematic Reviews, International Health Technology Assessment websites, Journal full text via OVID search engine and guideline databases as well as consultation with a committee of experts
Europe			
EUSOBI 2017	Position paper on screening for breast cancer by the European Society of Breast Imaging (EUSOBI) and 30 national breast radiology bodies from Austria, Belgium, Bosnia and Herzegovina, Bulgaria, Croatia, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Israel, Lithuania, Moldova, The Netherlands, Norway, Poland, Portugal, Romania, Serbia, Slovakia, Spain, Sweden, Switzerland and Turkey	European Society of Breast Imaging (EUSOBI)	Not reported
NHS 2016	NHS breast screening programme – clinical guidance for breast cancer screening assessment	National Health Service (NHS)	An expert group has made changes to guidance following consultation with national groups and organisations
Senkus 2015 (ESMO)	Primary breast cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up	European Society for Medical Oncology (ESMO)	Relevant literature has been selected by expert authors
Scharl 2013 (AGO)	AGO recommendations for diagnosis and treatment of patients with early breast cancer: update 2013	Arbeitsgemeinschaft Gynäkologische Onkologie (AGO)	Review and scoring of recent publications for their scientific validity
Wildiers 2013 (KCE)	Breast cancer in women: diagnosis, treatment and follow-up	Belgian Health Care Knowledge Centre (KCE)	Systematic review, consultation with experts and evidence grading using GRADE
Albert 2009	2008 update of the guideline: Early detection of breast cancer in Germany	Albert et al.	Literature review, consultation with experts and two formal consensus conferences
United States			
NCCN 2018	NCCN Guidelines Version 3.2018. Breast cancer screening and diagnosis	National Comprehensive Cancer Network (NCCN)	Review of the NCCN guidelines by cancer experts followed by a literature review using PubMed

Abbreviations: AGO, Arbeitsgemeinschaft Gynäkologische Onkologie (Austrian Gynaecological Oncology Group); CNB, core needle biopsy; DCIS, ductal carcinoma *in situ*; ESMO, European Society of Medical Oncology; EUSOBI, European Society of Breast Imaging; FNA, fine needle aspirations; FNAC, fine needle aspiration cytology; JBCS, Japanese Breast Cancer Society; KCE, Belgian Health Care Knowledge Centre; MOHM, Ministry of Health Malaysia; NCCN, National Comprehensive Cancer Network; NHS, National Health Service; VACB: vacuum-assisted core biopsy.
Note: Guidelines shown in shading are related (or partially related) to screening.

Table 3-21 International guidance: relevant recommendations/statements

Ref ID	Guidance type	Recommendation/evidence statement	Level of Evidence Grade	Evidence base (study ID/reference)		
				Clinical studies	Guidelines	Other
Asia						
Horii 2015 (JBCS)	Evidence-based	"FNAC is recommended as a diagnostic procedure for breast lesions"[27]	B[28]	Pisano 2001, Giard 1987, Yamaguchi 2012, Westenend 2001, Hatada 2000, Ballo 1996, Pijnappel 2004, Feoli 2008, Boland 2015		
		"CNB is recommended as a diagnostic procedure for breast lesions"[27]	B			
		"ST-VAB is recommended as a diagnostic procedure for non-palpable mammographically-visible lesions that are suspected to be breast cancer"	B	Gumus 2012, Penco 2010		
MOHM 2010	Evidence-based	"Patients presenting with a breast symptom should be evaluated with a full clinical examination, mammography and/or ultrasound followed by biopsy, either fine needle and/or core biopsy."	C[29]		KCE 2007, NICE 2009[30]	
		"FNAC may be considered as the initial method of pathological assessment for palpable breast lumps where facility and expertise are available."	C	Lieske 2006, Pilgrim 2005, Barra 2007, Garg 2007, Tharn 2009, Yong 1999		
		"Core biopsy may be used as a complement for pathological diagnosis if the fine needle aspiration cytology is equivocal."	C	Lieske 2006, Pilgrim 2005, Barra 2007, Garg 2007, Tharn 2009, Yong 1999		
		"Core biopsy in combination with Fine needle aspiration cytology may be used where facility and expertise are available."	C	Lieske 2006, Pilgrim 2005, Barra 2007, Garg 2007, Tharn 2009, Yong 1999		
Europe						
Sardanelli 2017 (EUSOBI)	Unclear[31]	"Preference should be given to needle sampling of breast lesions using core biopsy or vacuum-assisted biopsy instead of fine needle aspiration."	NA	Van Breest 2013		

[27] In the text it is noted that "FNAC and CNB are both recommended as a diagnostic procedure for breast lesions. In cases where physicians do not have adequate training, or the cytologist is inexperienced, CNB may be regarded as the first choice of diagnostic modality."

[28] No information on grading available.

[29] Grade - A = at least one meta-analysis, systematic review, or RCT, or evidence rated as good and directly applicable to target population; B = evidence from well conducted clinical trials, directly applicable to the target population and demonstrating overall consistency of results or evidence extrapolated from meta-analysis, systematic review or RCT; C = evidence from expert committee reports, or opinions and/or clinical experiences of respected authorities; indicates absence of directly applicable clinical studies of good quality.

[30] NICE guidelines were excluded from identified guidance as they do not make a recommendation regarding FNA and CNB, they only provide context: 'In most cases whether symptomatic or screen detected the diagnosis of breast cancer is made by triple assessment (clinical examination, mammography and/or ultrasonography imaging with core biopsy and/or fine needle aspiration cytology.'.

[31] Unclear as methods are not reported. This particular recommendation appears to be consensus-based as the reference provided in the evidence base describes the trends in FNA and CB use rather than being a clinical study.

Ref ID	Guidance type	Recommendation/evidence statement	Level of Evidence Grade	Evidence base (study ID/reference)		
				Clinical studies	Guidelines	Other
NHS 2016	Consensus-based	"Significant breast abnormalities should be assessed by core biopsy or VACB. If a service has access to high quality cytology with immediate reporting, then FNAC may be used in addition to core biopsy, but not instead of it. In exceptional cases FNAC may be used alone if core biopsy is not possible."	NA	Britton 1999, Liberman 2000, Bundred 2016	NHSBSP 2009	
Senkus 2015 (ESMO)	Evidence-based	Pathological diagnosis should be based on a core needle biopsy, obtained preferably by ultrasound or stereotactic guidance. A core needle biopsy (if this is not possible, at least a fine needle aspiration indicating carcinoma) must be obtained before any type of treatment is initiated.	IIIA[32]	Not reported		
Scharl 2013 (AGO)	Evidence- based	For DCIS, the pretherapeutic assessment of suspicious lesions (BIRADS IV), stereotactic core needle biopsy or vacuum assisted biopsy are recommended.	2bB[33] AGO++	Not reported		
Wildiers 2013 (KCE)	Evidence-based	The choice between core biopsy and/or FNAC depends on the clinician's, radiologist's and pathologist's experience. This forms the third step in the triple assessment approach, following clinical examination and imaging.	1C[34]	Feoli 2008, Kaur 2007, Klijanienko 1998		
Albert 2008	Evidence-based	The histological tissue diagnosis of ambiguous breast lesions has to be performed by imaging-guided core needle biopsy, vacuum-assisted biopsy or open biopsy.	A[35]	Schulz 2003	Albert 2004, NCCN 2007a, NICE 2006a, Perry 2006	
		Fine needle biopsy cannot be recommended as a standard method.	B	Schulz 2003	Albert 2004, NICE 2006a, NCCN 2007a	
United States						

[32] *Levels of evidence* - I: evidence from at least one large RCT, controlled trial of good methodological quality (low potential for bias) or meta-analysis of well-conducted RCTs without heterogeneity; II: small RCTs or large RCTs with a suspicion of bias (lower methodological quality) or meta-analysis of such trials or of trials with demonstrated heterogeneity; III = prospective cohort studies, IV = retrospective cohort studies or case-control studies, V = studies without control group, case reports, expert opinions. *Grade* - A = strong evidence for efficacy with a substantial clinical benefit, strongly recommended; B = strong or moderate evidence for efficacy but with a limited clinical benefit, generally recommended; C = insufficient evidence for efficacy or benefit does not outweigh the risk or disadvantages (adverse events, costs, …), optional; D = moderate evidence against efficacy or for adverse outcome, generally not recommended; E = strong evidence against efficacy or for adverse outcome, never recommended.

[33] Criteria for level of evidence scoring not provided. *Grade* – '++': This investigation or therapeutic intervention is highly beneficial for patients, can be recommended without restrictions, and should be performed; '+': This investigation or therapeutic intervention is of limited benefit for patients and can be performed; '+/-': This investigation or therapeutic intervention has not shown benefit for patients and may be performed only in individual cases. According to current knowledge a general recommendation cannot be giver; '-': This investigation or therapeutic intervention can be of disadvantage for patients and might not be performed; '-/-': This investigation or therapeutic intervention is of clear disadvantage for patients and should be avoided or omitted in any case.

[34] *Grade* - 1A = strong recommendation based on high level of evidence; 1B = strong recommendation based on moderate level of evidence; 1C = strong recommendation based on low or very low level of evidence; 2A = weak recommendation based on high level of evidence; 2B = weak recommendation based on moderate level of evidence; 2C = weak recommendation based on low or very low level of evidence.

[35] *Grade* - A = strong recommendation "must or should"; B = recommendation "shall or can"; 0 = recommendation open: "…" (open for action).

FNA and core biopsy: High level review of evidence and clinical practice guidelines

Ref ID	Guidance type	Recommendation/evidence statement	Level of Evidence Grade	Evidence base (study ID/reference)		
				Clinical studies	Guidelines	Other
NCCN 2018	Evidence-based	Within the management pathway following mammographic evaluation core biopsy is noted as the appropriate testing modality under the following circumstances: • BI-RADS category 3 assessment (probably benign finding) – perform diagnostic mammogram every 6 m for 1-2 y. o If return visit uncertain or strong patient preference may include biopsy. o If increased suspicion, perform core needle biopsy • BIRADS category 4 assessment (suspicious abnormality) or category 5 (highly suggestive of malignancy) – after complete imaging evaluation tissue sampling by ultrasound-guided core needle biopsy	2A[36]	NR	NR	NR

Abbreviations: AGO, Arbeitsgemeinschaft Gynäkologische Onkologie (German Gynaecological Oncology Group); CNB, core needle biopsy; DCIS, ductal carcinoma *in situ*; ESMO, European Society of Medical Oncology; EUSOBI, European Society of Breast Imaging; FNA, fine needle aspirations; FNAC, fine needle aspiration cytology; JBCS, Japanese Breast Cancer Society; KCE, Belgian Health Care Knowledge Centre; MOHM, Ministry of Health Malaysia; NCCN, National Comprehensive Cancer Network; NHS, National Health Service; NR, not reported; VACB: vacuum-assisted core biopsy.
Note: Guidelines shown in shading are related to screening.

[36] Category 2A: Based upon lower level evidence, there is uniform consensus that the intervention is appropriate.

3.3 Other considerations

Three comprehensive narrative reviews were identified during the literature search and have been used to provide further information on the use of FNA and core biopsy, particularly for areas not already covered in this report. These reviews have not been formally assessed but have been included to provide additional information that may be of interest.

The narrative studies that have been included are Lukasiewicz 2017, Mitra 2016 and Willems 2012. Lukasiewicz 2017 and Mitra 2016 are both recent publications and Willems 2012, while being a less recent publication, provides a comprehensive description of a large number of individual studies. All three reviews report information on diagnostic measures, procedural considerations and circumstances of the use of FNA and core biopsy. All three studies report no conflicts of interest for any authors.

Table 3-22 below predominantly outlines the findings of the Willems 2012 review, with the addition of one finding from Mitra 2016 (regarding anaesthesia). There were no inconsistencies in the findings across the three narrative reviews with the exception of the distinction between fibroadenoma and phyllodes tumour, which will be discussed below.

Procedural (dis)advantages

FNA and core biopsy differ in a number of procedural aspects, as outlined in Table 3-22. Compared to core biopsy, FNA is better at accessing deeper lesions sites, has a lower chance of complications and does not require anaesthesia, while core biopsy is not impacted as heavily by the level of operator experience and has a substantially higher success rate.

Limited information is provided by the included narrative reviews regarding deep lesion sites as a procedural consideration for both biopsy techniques. Lukasiewicz 2017 note that deep lesion location is a manual difficulty associated with core biopsy. Mitra 2016 states that deep lesion sites necessitate the use of FNA and Willems 2012 argues that FNA is preferred over core biopsy when sampling deep seated lesions. However, none of these narrative reviews note why deep lesion sites are difficult for core biopsy to sample.

As noted in this report, pain is typically a complication associated with both biopsy techniques. Additionally, Lukasiewicz 2017 note that haematoma and discomfort may occur following either biopsy procedure and are more common following core biopsy than FNA. Mitra 2016 and Willems 2012 also include infection and pneumothorax as possible complications associated with FNA and core biopsy. However, both reviews note that pneumothorax is very rare for either biopsy procedure.

The use of anaesthesia is only noted in Lukasiewicz 2017 and Mitra 2016. Both studies state that local anaesthesia is only required for core biopsy and not FNA. However, this is not expanded on in either review. Neither Lukasiewicz 2017 nor Mitra 2016 specify what circumstances would necessitate the use of anaesthesia for FNA. It should be noted that in the current evidence review, the findings of the Satchithananda 2005 study showed substantially greater pain associated with FNA compared with core biopsy (pain score ~4 versus ~2) when anaesthesia was not routinely used for FNA (see page 45). While the literature indicates that anaesthesia is used occasionally for FNA, mostly in cases when the patient is anxious about the procedure, and always for core biopsy procedures, it is not currently possible to comment on how the use of anaesthesia described in the literature compares with Australian clinical practice.

FNA requires highly experienced operators. It is emphasised in Willems 2012 that the quality of FNA and the interpretation of its results are predominantly determined by the skill and experience of the aspirator and cytopathologist respectively. This same point is argued in both Lukasiewicz 2017 and

Mitra 2016. This relates to the ability of the aspirator to be accurate and careful when sampling (see Section 3.1.2.2 for the data on inadequate sampling/repeat testing) and for the cytopathologist to be trained and experienced in distinguishing different types of breast lesions. Therefore, it seems that the success of FNA, and likely the diagnostic accuracy of the biopsy technique, is primarily dependent upon the operators. This may help explain the greater variability in the diagnostic results of FNA compared with core biopsy that were presented in Section 3.1.1.2.

Diagnostic performance

The dependence of diagnostic performance on *clinical/radiological features* for both tests differs. Core biopsy performs well for non-palpable lesions, palpable lesions and lesions smaller than 10mm and larger than 40mm, while FNA performs well for palpable lesions and sonographically targeted lesions.

The extent to which diagnostic performance is dependent on *microscopic features* also varied between FNA and core biopsy. FNA is not able to reliably distinguish *in situ* and invasive cancers and has poor performance when diagnosing pre-invasive lesions and borderline lesions such as papillary lesions. In contrast, core biopsy can distinguish between *in situ* and invasive cancers and performs relatively well when diagnosing pre-invasive lesions and borderline lesions. As noted above, there is discrepancy between the three narrative reviews regarding the ability of FNA and core biopsy to distinguish fibroadenoma and phyllodes tumours. As shown in the table, Willems 2012 concluded that FNA has a moderate ability and core biopsy a high ability to distinguish these tumour types. However, Lukasiewicz 2017 and Mitra 2016 concluded that both biopsy techniques are poor at distinguishing these types of tumours. Thus, core biopsy performs consistently over a wider range of lesion characteristics.

Assessment of prognostic and predictive biomarkers

The early assessment of prognostic and biomarkers (ancillary testing) can be used to inform treatment decisions and in particular guide neoadjuvant therapy. As shown in Table 3-22, FNA has a lower performance than core biopsy for all biomarkers listed. FNA is not generally suitable for tumour grading, assessment of proliferation using immunohistochemistry, oestrogen and progesterone receptor (ER/PR) assessment and human epidermal growth factor receptor 2 (HER2) assessment. Core biopsy has an acceptable ability to grade tumours and has a high ability to assess ER/PR, HER2 and proliferation. Therefore, core biopsy seems to be much more useful in the assessment of prognostic features and predictive biomarkers.

Time to perform

Based on clinical advice, the time taken to perform FNA is less than that for core biopsy, even with the use of rapid onsite evaluation (ROSE). However, it is noted in Willems 2012 and in a previous section in this report (see page 41) that due to the higher rate of inadequate sampling and lower diagnostic accuracy rates of FNA, more ancillary testing may be required following FNA compared with core biopsy. As FNA (performed without ROSE) often results in greater ancillary and repeat testing due to sample inadequacy, the time to reach a definitive diagnosis may be increased.

ROSE may reduce the impact of ancillary testing after FNA by reducing inadequate specimens by allowing immediate re-biopsy to obtain inadequate material. In a systematic review of the impact of ROSE on the adequacy rate of FNA cytology at various anatomic sites, Schmidt et al. (2013) found a significant improvement in adequacy rates overall (RD 0.12; 95% CI 0.08-0.16). However, in the three studies relating specifically to the use of ROSE of FNA in breast cancer, two studies found no difference in adequacy rates for ROSE compared with no ROSE (Hamill et al., 2002; Akalin et al., 2008) and one study found a significant difference (Dray et al., 2000; RD 0.14; 95% CI 0.08-0.20); the overall finding for breast cancer was not significant (RD 0.06; (% CI -0.05, 0.16). Schmidt et al. (2013) states the following "It is important to distinguish adequacy from diagnostic yield and

accuracy. *Diagnostic yield* refers to the rate at which a diagnosis is made (per slide or per case) and is distinct from adequacy. *Adequacy* measures whether a sample provides sufficient material for a diagnosis. *Accuracy* refers to the correspondence between cases for which a diagnosis was rendered (nondiagnostic cases are excluded) and a gold standard (histopathology or clinical follow-up). Adequacy is necessary but not sufficient for diagnosis. Diagnostic yield and accuracy are more directly related to patient outcomes than adequacy; however, these concepts are less directly related to sampling performance because they depend on the performance of both sampling (adequacy rate) and interpretation (rate of inconclusive samples, accuracy). Thus, adequacy is a more direct measure of sampling performance than diagnostic yield or accuracy".

Table 3-22 Summary of benefits and limitations of fine needle aspiration and core biopsy

	Fine needle aspiration	Core biopsy
Sensitivity and specificity		
Sensitivity (95% CI)[37]	74% (72-77)[38]	87% (84-88)[38]
Specificity (95% CI)[39]	96% (94-98)[38]	98% (96-99)[38]
Procedural advantages and disadvantages		
Ability to distinguish between *in situ* and invasive cancer	Low	High
Degree of invasiveness of technique	Low	Low to moderate
Success rate (rate of sufficient sampling)	Moderate	High
Complication rate	Very low	Low
Use of local anaesthesia	Optional	Required
Time taken to perform biopsy	Short duration (5-10 mins)	Moderate duration (15-20 mins)
Assessment of prognostic and predictive biomarkers		
Ability to assess tumour grade[40]	Low	Moderate to High
Ability to assess HER2 and ER/PR receptors	Receptor testing is recommended on core biopsies of the primary tumour[41]	

The conclusions of the three selected narrative reviews are presented in Table 3-23. While two of the reviews suggest core biopsy is preferable to FNA (Lukasiewicz 2017; Willems 2012), Mitra 2016 suggest that parameters such as palpability and cytopathology experience should be considered when choosing between the two methods.

Source: *Willems 2012: Table 1, page 290; supplemented by Mitra 2016: Table 1, page 2.

[37] The sensitivity of a diagnostic test quantifies its ability to correctly identify subjects with the disease. It is the proportion of true positives that are correctly identified by the test.

[38] Wang 2017 – the limitations of this review are discussed within the text of the Position Statement including that Ultrasound guidance was used in 5 of the 12 studies included in Wang, whereas ultrasound guidance for breast FNA has been standard practice in Australia since 1986.

[39] The specificity of a diagnostic test is the ability of a test to correctly identify subjects without the disease. It is the proportion of true negatives that are correctly identified by the test.

[40] Using Nottingham histological score (Elston, CW; Ellis, IO. Pathologic prognostic factors in breast cancer. I. The value of histological grades in breast cancer. Experience from a large study with long-term follow-up. Histopathology 19(5):403–10. Republished Histopathology 41:154–161. 2002).

[6]Royal College of Pathologists of Australasia 2018

Abbreviations: ADH, atypical ductal hyperplasia; ER, oestrogen receptor; HER2, human epidermal growth factor receptor 2; LN, lobular neoplasia; PR, progesterone receptor; ROSE, rapid on-site evaluation

Table 3-23 Conclusions from the selected narrative reviews

Citation	
Lukasiewicz 2017	"Microscopic verification of focal breast lesions is crucial for further therapeutic decisions. It has been proven that histopathological verification is more accurate and has more advantages than cytological assessment."
Mitra 2016	"Unlike the popular belief of an absolute superiority of [core biopsy] over [FNA], the literature review does not reveal a very distinct demarcation in many aspects. We recommend judicious use of these diagnostic modalities in resource-limited settings and screening programs considering parameters such as palpability and availability of an experienced cytopathologist."
Willems 2012	"Overall, [core biopsy] achieved better sensitivity and specificity especially in those lesions that were not definitively benign or malignant, non-palpable and/or calcified lesions. Although [FNA] is easier to perform, interpretation requires vast experience and even then, it is more often inconclusive requiring additional [core biopsy]. The authors conclude that overall [core biopsy] is to be preferred as a diagnostic technique."

Abbreviations: FNA, fine needle aspiration.

4 Discussion

The findings of the high-level review of published clinical evidence suggests that core biopsy is more sensitive than FNA at identifying breast cancer in a mammographic screening population, with similar specificity. This is based primarily on the findings of the prospective cohort study conducted in a mammographic screening population (Lieske et al., 2006) and the non-palpable lesion subgroup of a systematic review and meta-analysis that included only prospective cohort studies (Wang et al. 2017). Supporting this assertion are the recommendations from International guidance related to mammographic screening that core biopsy is the test of choice (NCCN 2018; Sardanelli et al, 2017; NHS 2016). However, there are two major limitations of the body of evidence examined for this high-level review that make interpretation of the findings and their extrapolation to the BreastScreen Australia program problematic: (i) the studies were conducted as far back as the 1990s and may not reflect current clinical practice and (ii) because practice has changed over time, the results based on non-current practice are likely to be biased against FNA. Each of these limitations will be explored in detail below.

The Lieske 2006 prospective cohort study was conducted in a mammographic screening population and as such is more applicable to the current setting than the Wang 2017 meta-analysis. Lieske 2006 conclude that "Core biopsy is better than [FNA] at preoperative diagnosis of screen-detected breast cancer as it missed fewer cancers." However, there are a number of factors that may reduce applicability to the BreastScreen Australia setting including the mode of guidance used and the size of the needle used for FNA. In the Lieske 2006 study, only 35% of FNAs and 41% of core biopsies were conducted using US-guidance, while the remainder were conducted using stereotactic or clinical guidance. Complete sensitivity of FNA compared with core biopsy was substantially lower when stereotactic guidance was used (71% versus 95%, respectively), than when US guidance was used (86% versus 94%). In addition, the needle size used for FNA was 21 gauge. Feedback from the EAG suggests that due to improvements in US technology, US is the preferred option for guidance for core biopsy, and the use of stereotactic guidance is known to result in a lower sensitivity for FNA. This is consistent with guidance from the NCCN 2018 and the stereotactic and US guidance findings described above. The EAG also noted that the use of a 21-gauge needle for FNA in Lieske 2006 is not consistent with current practice where 22-25 gauge needles are used. A meta-analysis by Yu et al. (2012) included an examination of the impact of needle size on the diagnostic accuracy of FNA and found no significant impact when comparing 21 gauge versus other sizes (relative diagnostic odds ratio 0.49 [95% CI 0.16, 1.53]). Both of these differences with current practice likely relate to the fact the population examined in the study underwent screening between 1999 to 2003; from 15 to 19 years ago.

Overall, the results of the Wang 2017 meta-analysis suggest that FNA has a lower sensitivity (74%; 95% CI 72-77) than core biopsy (87%; 95% CI 84-88) and a similar specificity (96%; 95% CI 94-98 and 98%; 95% CI 96-99, respectively). With regard to these findings, Wang 2017 concludes that "*From the perspective of screening, which prefers higher sensitivity, core biopsy will do a better job than FNAC. While from the perspective of specialists for diagnostic precision, which prefers higher specificity, core needle biopsy performs similarly to FNAC.*" However, a number of points should be noted about these findings. While the internal validity of this meta-analysis is high (due to the fact both tests were conducted in the same lesions for all included studies and the risk of bias is considered low), the external validity or applicability of the findings to the current context may be considered low due to the timeframe in which the studies included in the meta-analysis were conducted. Seven of the 12 included studies provided the study dates: one was conducted in the late 1980s, three were conducted largely in the 1990s, one in the early 2000s and two in the last decade (2007 to 2010, and 2009 to 2012). Differences between the included studies' characteristics and current Australian practice in the BreastScreen Australia program include that (i) the included

studies were not conducted in a mammographic screening setting, and (ii) freehand or stereotactic guidance were used in more than half of the included studies, which is not consistent with current Australian practice where US-guidance is largely used. Subgroup analysis based on non-palpable lesions may provide a more appropriate estimate of diagnostic accuracy in a mammographic screening program. This shows that FNA performs even more poorly than core biopsy for sensitivity [68% versus 84%, respectively]). However, further subgroup analyses where included studies were limited to those that used US-guidance and were more recent studies (published from 2007 onwards) showed the sensitivity of FNA to be closer to that of core biopsy (84% versus 88% for US-guidance and 81% versus 88% for more recent studies).

The fact the evidence comes from studies conducted up to 30 years ago, with no recent evidence, may bias the findings. For example, based on the results of these previous studies assessing FNA and core biopsy, it has been determined that FNA does not perform well in certain lesion types, such as those presenting with microcalcifications and if identified or suspected these should be tested using core biopsy. Thus, determining the sensitivity of FNA using the data from these studies may bias against FNA, (i) because FNA is known to perform poorly in these situations and (ii) FNA would not be used in current practice for these lesions.

The Lieske 2006 study includes only data on lesions that were found to be a malignancy at subsequent surgery. The final histology was as follows: 53% invasive ductal carcinoma, 12% invasive lobular carcinoma, 4% mixed ductal and lobular carcinoma, 4% tubular carcinoma (all with or without DCIS), 23% DCIS only and 3% other invasive carcinoma. In this study there was a larger difference between the complete sensitivities of FNA and core biopsy when lesions with microcalcifications were considered (73% versus 94%, respectively) than when soft tissue lesions were considered (85% versus 92%, respectively). Thus, the overall findings that include lesions with microcalcification could bias the results against FNA. Inadequate FNA sampling was also substantially greater for FNA than was non-diagnostic core biopsy in lesions presenting as microcalcification on mammography (12.5% versus 2.6%, respectively) and shown to be DCIS on final histology (13.6% versus 3.4%, respectively). The authors note that "The better preoperative diagnosis rate of [core biopsy] in our study was mainly owing to better diagnosis of DCIS, which mainly presented with mammographic microcalcification that often required stereotactic approach. It is known that the sensitivity of FNAC is the lowest when performed stereotactically (Britton, 1999), especially in the assessment of microcalcifications (Pisano et al, 1998), and our findings in this large study confirm this observation."

Lesion type was not considered or reported in the Wang 2017 meta-analysis. Thus, not knowing the types of lesions that the tests were performed in makes it difficult to interpret the findings of the meta-analysis, as the types of lesions may have included those that are known to be unsuitable for FNA, resulting in an underestimate of the diagnostic accuracy of FNA.

While the high-level review of clinical evidence identified data on follow-on testing, complications, breast cancer survival and recurrence, the evidence was considered to be at a high risk of bias due to the risk of potential confounding. Therefore, definitive conclusions around FNA and core biopsy based on these outcomes are not possible. While the analysis of overall survival in the Liikanen 2016 study was considered to be at a lower risk of bias due to the use of multivariate analysis of adjust for potential confounding variable, the analysis was underpowered, and thus the lack of difference in overall survival between FNA and core biopsy may be the result of a lack of precision rather than a true lack of effect.

The high-level review of existing clinical guidance largely suggests the use of core biopsy over FNA in a mammographic screening setting, with the NHS (2016) recommending the use of core biopsy over FNA, EUSOBI (2017) indicating a preference towards core biopsy, and the NCCN (2018)

suggesting only core biopsy following mammographic evaluation (which may refer to screening or a diagnostic setting), with no mention of FNA. The BreastScreen Australia Clinical Advisory Committee guidance (BSA 2017) provides more detailed guidance, stating "the use of FNA in the screening setting should be limited to cysts, lymph nodes and the rare situations where core biopsy is not possible", and "Where possible core biopsy, including VACB should be the procedure of choice." The BSA 2017 guidance was informed by a review of the evidence and the NHS 2016 guidance. The guidance around the use of FNA and core biopsy for diagnosis is more mixed: two sources clearly recommend the use of core biopsy over FNA (ESMO 2015; Albert 2008), two sources recommend core biopsy in the intervention pathway and make no mention of FNA (NCCN 2018; AGO 2013). Further, one of these sources states that FNA cannot be recommended as a standard method (Albert 2008). In contrast, three sources present no clear preference between core biopsy and FNA and advise that the choice depends on the level of expertise available (KCE 2013; JBCS 2013; MOHM 2010). Although not specifically stated, the Malaysian guidelines appear to preference the use of FNA over core biopsy as the initial method of pathological assessment for palpable breast lumps where facility and expertise are available. They also recommended that core biopsy should be used in combination with FNA.

While the evidence base used to inform these recommendations/evidence statements is not readily available for all identified guidance, it should be noted that Lieske 2006 is mentioned only in the MOHM 2010 guidance, as are Barra 2007 and Garg 2007 (which were included in the Wang 2017 meta-analysis). The majority of evidence cited is not that which was included in this high-level review, with most of it being pre-2010 and likely not comparing FNA with core biopsy.

The three narrative reviews outlined a number of additional considerations regarding the use of FNA and core biopsy. The issue of the appropriateness of FNA and core biopsy for different clinical or radiological features was noted, with core biopsy considered to perform well for both palpable and non-palpable, smaller or larger lesions (< 10 mm or > 40 mm) and lesions showing microcalcification on mammography.

FNA performs well for palpable lesions and larger lesions between 10 mm and 40 mm and is not useful where microcalcification is present. With regards to diagnostic performance based on histological features, FNA is not able to distinguish between *in situ* and invasive cancers, whereas core biopsy can (however these present with microcalcification in the majority of cases).

The three narrative reviews also state that FNA performs poorly compared with core biopsy in terms of the assessment of prognostic and predictive biomarkers. However, a 2013 review of FNA by Garbar and Curé includes discussion on liquid-based cytology (LBC), which standardises cell fixation, concentrates epithelial cells and discards blood cells and/or cell debris that can obscure the smear, and which allows tests such as immunocytochemistry, flow cytometry or molecular biology. They note that a study by Domanski et al. (2012) found concordance between FNA and core biopsy of 98% for oestrogen receptor (ER) status and 96% for progesterone receptor (PR) status. Another technique discussed is cell block cytology, which involves centrifuging collected cells into a pellet, and then embedding the pellet in a synthetic polymer gel that is processed in a paraffin block that can be cut at 4 μm, as per histological core biopsy slides. Using this technique, Ferguson et al. (2013) found a concordance rate of 95% for ER status, 90% for PR status and 88% for HER2. Garbar and Curé note that the concordance rate using *in situ* hybridisation methods (fluorescence, chromogenic and silver) is also good, and that these are useful where HER2 is uncertain. For further information regarding HER2 testing, refer to ASCO CAP 2018 HER2 Testing for Breast Cancer Guidelines - Recommendations for Practice in Australasia.

Finally, the narrative reviews all note that the success of FNA is operator-dependent. Lieske et al. (2006) note that FNA is more operator-dependent than core biopsy, and that an "adequately

trained and experienced cytopathologist skilled in the interpretation of FNA and an experienced aspirator is necessary to reduce the number of nondiagnostic and inadequate FNACs. The expertise required in the interpretation of FNA requires specific training and exposure, whereas core biopsies are in general easier to interpret without specific expertise required." As noted previously, this may help explain the greater variability in the diagnostic results of FNA compared with core biopsy across the studies that were included in the Wang 2017 meta-analysis in Section 3.1.1.2.

5 References

Ahmed, M. E., et al. (2010). "Ultrasound guided fine needle aspiration cytology versus core biopsy in the preoperative assessment of non-palpable breast lesions." Journal of Ayub Medical College, Abbottabad: JAMC 22(2): 138-142.

Akalin A, Lu D, Woda B, et al. (2008) Rapid cell blocks improve accuracy of breast FNAs beyond that provided by conventional cell blocks regardless of immediate adequacy evaluation. Diagn Cytopathol. 36:523-529.

Albert, U. S., et al. (2009). "2008 update of the guideline: Early detection of breast cancer in Germany." Journal of Cancer Research and Clinical Oncology 135(3): 339-354.

Altaaf, H. N. and F. Farooqui (2015). "A comparison of ultrasound guided fine needle aspiration cytology and core needle biopsy in evaluation of palpable breast lesions." Rawal Medical Journal 40(4): 392-395.

Barra, A. D. A., et al. (2008). "A comparision of aspiration cytology and core needle biopsy according to tumor size of suspicious breast lesions." Diagnostic Cytopathology 36(1): 26-31.

Bonifacino, A., et al. (2005) "Accuracy rates of US-guided vacuum-assisted breast biopsy." Anticancer Research 25: 2465-2470.

Cangiarella, J. F., et al. (2000). "The use of stereotaxic core biopsy and stereotaxic aspiration biopsy as diagnostic tools in the evaluation of mammary calcification." Breast Journal 6(6): 366-372.

Dennison, G., et al. (2003). "A Prospective Study of the Use of Fine-Needle Aspiration Cytology and Core Biopsy in the Diagnosis of Breast Cancer." Breast Journal 9(6): 491-493.

Domanski, A. M., Monsef, N., Domanski, H. A., Grabau, D., ferno, M. (2012) "Comparison of the oestrogen and progesterone receptors status in primary breast carcinomas as evaluated by immunohistochemistry and immunocytochemistry: a consecutive series of 267 patients." Cytopathology 12: 1365–2303.

Dowlatshahi K., et al. (1991). Nonpalpable breast lesions: findings of stereotaxic needle-core biopsy and fine-needle aspiration cytology. Radiology 181(3):745-750.

Dray M, Mayall F, Darlington A. (2000) Improved fine needle aspiration (FNA) cytology results with a near patient diagnosis service for breast lesions. Cytopathology 11:32-37.

Ferguson, J., Chamberlain, P., Cramer, H.M., Wu, H. (2013) "ER, PR, HER2 immunocytochemistry on cell-transferred cytologic smears of primary and metastatic breast carcinomas: a comparison study with formalin-fixed cell blocks and surgical biopsies." Diagnostic Cytopathology, 41(7): 575–581.

Garg, S., et al. (2007). "A comparative analysis of core needle biopsy and fine-needle aspiration cytology in the evaluation of palpable and mammographically detected suspicious breast lesions." Diagnostic Cytopathology 35(11): 681-689.

Hamill J, Campbell ID, Mayall F, et al. (20020 Improved breast cytology results with near patient FNA diagnosis. Acta Cytol 46:19-24.

Horii, R., et al. (2016). "The Japanese Breast Cancer Society clinical practice guidelines for pathological diagnosis of breast cancer, 2015 edition." Breast Cancer 23(3): 391-399.

Hukkinen, K., et al. (2008). "Unsuccessful preoperative biopsies, fine needle aspiration cytology or core needle biopsy, lead to increased costs in the diagnostic workup in breast cancer." Acta Oncologica 47(6): 1037-1045.

Leaver, A., et al. (2010). "Fine needle aspiration versus touch imprint cytology in ultrasound guided core of breast masses." Breast Cancer Research 12: S8.

Leifland, K., et al. (2003). "Comparison of stereotactic fine needle aspiration cytology and core needle biopsy in 522 non-palpable breast lesions." Acta radiologica (Stockholm, Sweden: 1987) 44(4): 387-391.

Lieske, B., et al. (2006). "Role of fine-needle aspiration cytology and core biopsy in the preoperative diagnosis of screen-detected breast carcinoma." British Journal of Cancer 95(1): 62-66.

Liikanen, J., et al. (2016). "Breast cancer prognosis and isolated tumor cell findings in axillary lymph nodes after core needle biopsy and fine needle aspiration cytology Biopsy method and breast cancer outcome." European Journal of Surgical Oncology 42(1): 64-70.

Lukasiewicz, E, Ziemiecka, A, Jakubowski, W, Vojinovic, J, Bogucevska, M, Dobruch-Sobczak, K. (2017). Fine-needle versus core-needle biopsy - which one to choose in preoperative assessment of focal lesions in the breasts? Literature review. J Ultrason. 17(71):267-74.

Mak, W. S., et al. (2012). "Ultrasound-guided biopsy of solid breast lesions: Should fine-needle aspiration be replaced by core biopsy?" Hong Kong Journal of Radiology 15(1): 10-14.

Malaysian Ministry of Health (2010). "management of Breast Cancer – 2nd Edition." Available at: http://www.acadmed.org.my/index.cfm?&menuid=67#Cancer.

Mitra, S, Dey, P. (2016). Fine-needle aspiration and core biopsy in the diagnosis of breast lesions: A comparison and review of the literature. CytoJournal. 13(1).

National Comprehensive Cancer Network (2018) Breast cancer screening and diagnosis. Version 3.2018. Available at: https://www.nccn.org/professionals/physician_gls/default.aspx.

National Health Service (2016) NHS Breast Screening Programme. Clinical guidance for breast cancer screening assessment. NHSBSP publication number 49. Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/567600/Clinical_guidance_for_breast_cancer_screening_assessment_Nov_2016.pdf.

Saha, A., et al. (2016). "FNAC versus core needle biopsy: A comparative study in evaluation of palpable breast lump." Journal of Clinical and Diagnostic Research 10(2): EC05-EC08.

Sardanelli, F., et al. (2017). "Position paper on screening for breast cancer by the European Society of Breast Imaging (EUSOBI) and 30 national breast radiology bodies from Austria, Belgium, Bosnia and Herzegovina, Bulgaria, Croatia, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Israel, Lithuania, Moldova, The Netherlands, Norway, Poland, Portugal, Romania, Serbia, Slovakia, Spain, Sweden, Switzerland and Turkey." European Radiology 27(7): 2737-2743.

Satchithananda, K., et al. (2005). "An audit of pain/discomfort experienced during image-guided breast biopsy procedures." Breast J 11(6): 398-402.

Scharl, A., et al. (2013). "AGO recommendations for diagnosis and treatment of patients with early breast cancer: Update 2013." Breast Care 8(3): 174-180.

Senkus, E., et al. (2015). "Primary breast cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up." Annals of Oncology 26: v8-v30.

Shea, B. J., et al. (2017). "AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both." BMJ 358: j4008.

The Royal College of Pathologists of Australasia. (2018). ASCO CAP 2018 HER2 Testing for Breast Cancer Guidelines - Recommendations for Practice in Australasia. https://www.rcpa.edu.au/Library/College-Policies/Guidelines/ASCO-CAP-2018-HER2-Testing-for-Breast-Cancer-Guide. Access date 21 March 2019.

Tikku, G. and P. Umap (2016). "Comparative study of core needle biopsy and fine needle aspiration cytology in palpable breast lumps: Scenario in developing nations." Turk Patoloji Dergisi 32(1): 1-7.

Wang, M., et al. (2017). "A sensitivity and specificity comparison of fine needle aspiration cytology and core needle biopsy in evaluation of suspicious breast lesions: A systematic review and meta-analysis." Breast 31: 157-166.

Whiting, P. F., et al. (2011). "QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies." Ann Intern Med 155(8): 529-536.

Wildiers H, Stordeur S, Vlayen J, Scholten R, van de Wetering F, Bourgain C, Carly B, Christiaens M-R, Cocquyt V, Lifrange E, Schobbens J-C, Van Goethem M, Villeirs G, Van Limbergen E, Neven P. Breast cancer in women: diagnosis, treatment and follow-up. Good Clinical Practice (GCP) Brussels: Belgian Health Care Knowledge Centre (KCE). (2013). KCE Reports 143 – 3rd EDITION. D/2013/10.273/38.

Willems, SM, Van Deurzen, CHM, Van Diest, PJ. (2012). Diagnosis of breast lesions: Fine-needle aspiration cytology or core needle biopsy? A review. Journal of Clinical Pathology. 65(4):287-92.

Wong, C. S., et al. (2009). "Is ultrasonography-guided modified coaxial core biopsy of the breast a better technique?" Hong Kong Medical Journal 15(4): 246-248.

Wu, Y. P., et al. (2004). "Clinical evaluation of three methods of fine-needle aspiration, large-core needle biopsy and frozen section biopsy with focus staining for non-palpable breast disease." Ai zheng = Aizheng = Chinese journal of cancer 23(3): 346-349.

Yu, Y-H., Wei, W., Liu, J-L. (2012) "Diagnostic value of fine-needle aspiration biopsy for breast mass: a systematic review and meta-analysis." BMC Cancer 12:41.

Appendix A Evidence hierarchy

The levels of evidence hierarchy for diagnostic accuracy and intervention questions, developed by the National Health and Medical Research Council (NHMRC), are shown in Table AppA.1.

Table AppA.1 Designations of levels of evidence for diagnostic accuracy and intervention questions

Level	Diagnostic accuracy	Intervention
I[a]	A systematic review of Level II studies	A systematic review of Level II studies
II	A study of test accuracy with an independent, blinded comparison with a valid reference standard[d], among consecutive persons with a defined clinical presentation[e]	A randomised controlled trial
III-1	A study of test accuracy with: an independent, blinded comparison with a valid reference standard[d], among non-consecutive persons with a defined clinical presentation[e]	A pseudo-randomised controlled trial (i.e. alternate allocation or some other method)
III-2	A comparison with reference standard that does not meet the criteria required for Level II and III-1 evidence	A comparative study with concurrent controls: • Non-randomised, experimental trial[b] • Cohort study • Case-control study • Interrupted time series with a control group
III-3	Diagnostic case-control study[e]	A comparative study without concurrent controls: • Historical control study • Two or more single-arm study[c] • Interrupted time series without a parallel control group
IV	Study of diagnostic yield (no reference standard)[f]	Case series with either post-test or pre-test/post-test outcomes

Source: National Health and Medical Research Council. NHMRC levels of evidence and grades for recommendations for developers of guidelines. Canberra: National Health and Medical Research Council, 2009.

a systematic review will only be assigned a level of evidence as high as the studies it contains, excepting where those studies are of Level II evidence. Systematic reviews of Level II evidence provide more data than the individual studies and any meta-analyses will increase the precision of the overall results, reducing the likelihood that the results are affected by chance. Systematic reviews of lower level evidence present results of likely poor internal validity and thus are rated on the likelihood that the results have been affected by bias, rather than whether the systematic review itself is of good quality. Systematic review quality should be assessed separately. A systematic review should consist of at least two studies. In systematic reviews that include different study designs, the overall level of evidence should relate to each individual outcome/result, as different studies (and study designs) might contribute to each different outcome.

b This also includes controlled before-and-after (pre-test/post-test) studies, as well as adjusted indirect comparisons (i.e. utilise A vs B and B vs C, to determine A vs C with statistical adjustment for B).

c Comparing single-arm studies i.e. case series from two studies. This would also include unadjusted indirect comparisons (i.e. utilise A vs B and B vs C, to determine A vs C but where there is no statistical adjustment for B).

d The validity of the reference standard should be determined in the context of the disease under review. Criteria for determining the validity of the reference standard should be pre-specified. This can include the choice of the reference standard(s) and its timing in relation to the index test. The validity of the reference standard can be determined through quality appraisal of the study (Whiting et al 2003).

e Well-designed population based case-control studies (eg. population based screening studies where test accuracy is assessed on all cases, with a random sample of controls) do capture a population with a representative spectrum of disease and thus fulfil the requirements for a valid assembly of patients. However, in some cases the population assembled is not representative of the use of the test in practice. In diagnostic case-control studies a selected sample of patients already known to have the disease are compared with a separate group of normal/healthy people known to be free of the disease. In this situation patients with borderline or mild expressions of the disease, and conditions mimicking the disease are excluded, which can lead to exaggeration of both sensitivity and specificity. This is called spectrum bias or spectrum effect because the spectrum of study participants will not be representative of patients seen in practice (Mulherin and Miller 2002).

f Studies of diagnostic yield provide the yield of diagnosed patients, as determined by an index test, without confirmation of the accuracy of this diagnosis by a reference standard. These may be the only alternative when there is no reliable reference standard.

Appendix B Search strategy

B.1 Clinical evidence review

B.1.1 Search strategy

Table AppB.1 Search strategy to identify clinical evidence

Source of information	Database/websites	Date limits and search terms
Electronic database	Embase® platform (which concurrently searches Embase and MEDLINE)[42] Cochrane Library (Cochrane Database of Systematic Reviews; Database of Abstracts of Reviews of Effect [DARE][43] ; Health Technology Assessment Database)	From 1 January 2000 to present. See Table AppB.2 Embase® search string for search strategy.
HTA websites	AHRQ (https://www.ahrq.gov/) Canadian Agency for Drugs and Technology in Health (CADTH) (https://www.cadth.ca/) CRD (https://www.crd.york.ac.uk/CRDWeb/) EuroScan (http://www.euroscan.bham.ac.uk/) HealthPACT (https://www.health.qld.gov.au/healthpact/html/tech-evaluated) HQO (http://www.hqontario.ca/) MSAC (http://www.msac.gov.au/) NHS Evidence (https://www.evidence.nhs.uk/) NICE (https://www.nice.org.uk/) SBU (https://www.sbu.se/en/) USPSTF (https://www.uspreventiveservicestaskforce.org/)	From 1 January 2000 to present. Search terms will depend on the complexity of the available search engine. In the first instance, 'breast' combined with each intervention.
Peak cancer authorities	AICR (http://www.aicr.org/) American Cancer Society (https://www.cancer.org/) ASCO (http://www.asco.org/) Canadian Cancer Society (http://www.cancer.ca/) Cancer Australia (https://canceraustralia.gov.au/) Cancer Care Ontario (https://www.cancercare.on.ca/) Cancer Research UK (https://www.cancerresearchuk.org/) EORTC (http://www.eortc.org/) ESMO (http://www.esmo.org/) IARC (http://www.iarc.fr/) NCCN (https://www.nccn.org/) NCI (https://www.cancer.gov/)	Search terms will depend on the complexity of the available search engine. In the first instance, 'breast' combined with each intervention.

AHRQ, Agency for Healthcare Research and Quality; AICR, American Institute for Cancer Research; ASCO, American Society of Clinical Oncology; CADTH, Canadian Agency for Drugs and Technologies in Health; CRD, Centre for Reviews and Dissemination at the University of York; EORTC, European Organization for Research and Treatment of Cancer; ESMO, European Society for Medical Oncology; EuroScan, European Information Network on New and Changing Health Technologies; FNA, fine needle aspiration; HQO, Health Quality Ontario; HTA, Health Technology Assessment; IARC, International Agency for Research on Cancer; MSAC, Medical Services Advisory Committee; NCCN, National Comprehensive Cancer Network; NCI, National Cancer Institute; NICE, National Institute for Health and Care Excellence; SBU, Swedish Agency for Health Technology Assessment and Assessment of Social Services; USPSTF, United States Preventive Services Task Force.

[42] Includes Ovid MEDLINE(R) Epub Ahead of Print, In-Process & Other Non-Indexed Citations, Ovid MEDLINE(R) Daily and Ovid MEDLINE(R) 1946 to Present

[43] DARE was produced by the Centre for Reviews and Dissemination at the University of York, UK, until April 2015. Funding for DARE ceased at the end of March 2015. The April 2015 updates contain bibliographic records from searches until the end of 2014.

Table AppB.2 Embase® search string for evidence review

String	Search terms
1	breast:ab,ti OR mamm*:ab,ti OR 'atypical ductal hyperplasia':ab,ti OR 'ductal carcinoma *in situ*':ab,ti
2	'fine needle'ab,ti OR 'aspiration cytology':ab,ti
3	'core needle'ab,ti OR 'core biopsy':ab,ti OR 'needle biopsy':ab,ti OR 'mechanical biopsy':ab,ti OR 'vacuum assisted':ab,ti OR 'vacuum biopsy':ab,ti
4	1 AND 2 AND 3
5	limit 5 to yr="2000 -Current"

B.2 Clinical guidance review

B.2.1 Search strategy

Table AppB.3 Search strategy to identify clinical guidance

Source of information	Database/website	Restrictions
Clinical practice guideline databases	Australian Clinical Practice Guidelines Portal (https://www.clinicalguidelines.gov.au/) AHRQ's National Guideline Clearinghouse (https://guideline.gov/) Guidelines International Network (http://www.g-i-n.net/library/international-guidelines-library) SIGN (http://sign.ac.uk/search.html) NICE (https://www.nice.org.uk/)	Evidence-informed clinical guidance published from 2008 onwards. Search terms depend on complexity of the search engine. In first instance try 'breast'.
Electronic databases	Embase® platform (which concurrently searches Embase and MEDLINE)[44]	Evidence-informed clinical guidance published from 2008 onwards.
Peak cancer authorities	AICR (http://www.aicr.org/) American Cancer Society (https://www.cancer.org/) ASCO (http://www.asco.org/) Canadian Cancer Society (http://www.cancer.ca/) Cancer Australia (https://canceraustralia.gov.au/) Cancer Care Ontario (https://www.cancercare.on.ca/) Cancer Research UK (https://www.cancerresearchuk.org/) EORTC (http://www.eortc.org/) ESMO (http://www.esmo.org/) IARC (http://www.iarc.fr/) NCCN (https://www.nccn.org/) NCI (https://www.cancer.gov/))	Evidence-informed clinical guidance published from 2008 onwards.

[44] Includes Ovid MEDLINE(R) Epub Ahead of Print, In-Process & Other Non-Indexed Citations, Ovid MEDLINE(R) Daily and Ovid MEDLINE(R) 1946 to Present

Table AppB.4 Embase® search string for guidance review

String	Search terms	Records
1	'breast tumor'/exp OR 'breast cancer':ti,ab OR ((breast NEAR/2 (tumo$r OR malignan*)):ti,ab) OR ((mammary NEAR/2 (tumo$r OR malignan*)):ti,ab) OR (('borderline lesion':ti,ab OR 'ultrasound lesion':ti,ab) AND (breast:ti,ab OR mammary:ti,ab))	530,070
2	'needle biopsy'/exp OR 'aspiration cytology'/exp OR 'core biopsy'/exp OR (('fine needle' NEAR/2 biopsy):ti,ab,de,kw) OR 'needle biopsy':de,kw,ti,ab OR 'aspiration cytology':de,kw,ti,ab OR 'core biopsy':de,kw,ti,ab OR 'population screening':ab,ti,de,kw OR 'screening program':ab,ti,kw,de OR 'breast screening':ab,ti,de,kw OR mammography:ab,ti,de,kw	163,521
3	'practice guideline'/exp OR 'guideline'/exp OR guideline$:ti,kw,de,ab OR (((practice OR treatment$ OR clinical) NEAR/2 guideline$):ab,de,kw,ti) OR ((((position:ti,ab,de,kw AND statement$:ti,ab,de,kw OR policy:ti,ab,de,kw) AND statement$:ti,ab,de,kw OR practice:ti,ab,de,kw) AND parameter$:ti,ab,de,kw OR best:ti,ab,de,kw) AND practice$:ti,ab,de,kw) OR (((critical OR clinical OR practice) NEAR/2 (path OR paths OR pathway OR pathways OR protocol*)):ti,ab,de,kw) OR ((care NEAR/2 (path$ OR pathway$ OR map$ OR plan$)):ti,ab,de,kw) OR ((algorithm$ NEAR/2 (screening OR examination$ OR test* OR assessment$ OR diagnos* OR therap* OR treatment$ OR intervention$)):ti,ab,de,kw) OR (((clinical OR practice) NEAR/2 (path$ OR pathway$ OR protocol$)):ti,ab,de,kw)	882,306
4	#1 AND #2 AND #3 AND [2008-2018]/py	2,747
5	#4 NOT ([conference abstract]/lim OR [conference review]/lim OR [editorial]/lim OR [erratum]/lim OR [letter]/lim OR [note]/lim OR [book]/lim) AND [english]/lim	1,684

Appendix C Quality assessment

Table AppC.1 Critical appraisal checklist – Wang 2017 (SR/AMSTAR 2)[45]

Question	Answer
1. Did the research questions and inclusion criteria for the review include the components of PICO?	Yes
2. Did the report of the review contain an explicit statement that the review methods were established prior to the conduct of the review and did the report justify any significant deviations from the protocol?	No
3. Did the review authors explain their selection of the study designs for inclusion in the review?	Yes
4. Did the review authors use a comprehensive literature search strategy?	Yes
5. Did the review authors perform study selection in duplicate?	Yes
6. Did the review authors perform data extraction in duplicate?	Yes
7. Did the review authors provide a list of excluded studies and justify the exclusions?	No
8. Did the review authors describe the included studies in adequate detail?	Yes
9. Did the review authors use a satisfactory technique for assessing the RoB in individual studies that were included in the review?	Yes
10. Did the review authors report on the sources of funding for the studies included in the review?	No
11. If meta-analysis was performed did the review authors use appropriate methods for statistical combination of results?	Yes
12. If meta-analysis was performed, did the review authors assess the potential impact of RoB in individual studies on the results of the meta-analysis or other evidence synthesis?	Yes
13. Did the review authors account for RoB in individual studies when interpreting/ discussing the results of the review?	Yes
14. Did the review authors provide a satisfactory explanation for, and discussion of, any heterogeneity observed in the results of the review?	Yes
15. If they performed quantitative synthesis did the review authors carry out an adequate investigation of publication bias (small study bias) and discuss its likely impact on the results of the review?	Yes
16. Did the review authors report any potential sources of conflict of interest, including any funding they received for conducting the review?	Yes
Systematic review risk of bias: low	

Abbreviations: AMSTAR 2, A MeaSurement Tool to Assess systematic Reviews 2; PICO, population/intervention/comparator/outcome; RoB, risk of bias; SR, systematic review.

[45] https://www.bmj.com/content/358/bmj.j4008?gclid=EAIaIQobChMI8fOmrtXr3QIVh1p-Ch1hQwYJEAAYASAAEgKMBPD_BwE.

Table AppC.2 Critical appraisal checklist – Lieske 2006 (prospective diagnostic cohort study/QUADAS 2)[46]

Question			Answer	RoB tating
Patient selection				
A.	Risk of bias	Was a consecutive or random sample of patients enrolled?	Yes	Low
		Was a case-control design avoided?	Yes	
		Did the study avoid inappropriate exclusions?	Yes	
		Could the selection of patients have introduced bias?	No	
B.	Concerns regarding applicability	Is there a concern that the included patients do not match the review question?	No	Low
Patient selection				
A.	Risk of bias	Were the index tests interpreted without knowledge of the results of the reference standard?	Yes	Low
		Is a threshold was used, was it prespecified?	Yes	
		Could the conduct or interpretation of the index test have introduced bias?	No	
B.	Concerns regarding applicability	Is there concern that the index test, its conduct, or interpretation differ from the review question?	No	Low
Reference standard				
A.	Risk of bias	Is the reference standard likely to correctly classify the target condition?	Yes	Low
		Were the reference standard results interpreted without the knowledge of the results of the index test?	No	
		Could the reference standard, its conduct, or its interpretation have introduced bias?	No	
B.	Concerns regarding applicability	Is there concern that the target condition as defined by the reference standard does not match the review question?	No	Low
Flow and timing				
A.	Risk of bias	Was there an appropriate interval between index test(s) and reference standard?	Yes	Moderate
		Did all patients receive a reference standard?	No – surgical excision not carried out on those with radiologically nonsuspicious lesions with benign FNA and CB findings, thus only those with malignancy on surgical excision included in analysis (FN may be underestimated). Likely to be small risk.	
		Did patients receive the same reference standard?	Yes	
		Were all patients included in the analysis?	Yes	
		Could the patient flow have introduced bias?	Yes	
Overall study risk of bias:		Moderate for sensitivity outcomes; low for inadequate sampling outcome		

Abbreviations: CB, core biopsy; FNA, fine needle aspiration; QUADAS 2, Quality Assessment of Diagnostic Accuracy Studies 2; RoB, risk of bias.

[46] http://annals.org/aim/fullarticle/474994/quadas-2-revised-tool-quality-assessment-diagnostic-accuracy-studies.

Table AppC.3 Critical appraisal checklist – Bonifacino 2005 (prospective diagnostic cohort study/QUADAS 2)[47]

Question			Answer	RoB tating
Patient selection				
A.	Risk of bias	Was a consecutive or random sample of patients enrolled?	Yes	Low
		Was a case-control design avoided?	Yes	
		Did the study avoid inappropriate exclusions?	Yes	
		Could the selection of patients have introduced bias?	No	
B.	Concerns regarding applicability	Is there a concern that the included patients do not match the review question?	No	Low
Patient selection				
C.	Risk of bias	Were the index tests interpreted without knowledge of the results of the reference standard?	Yes	Low
		Is a threshold was used, was it prespecified?	Yes for FNA, unclear for VACB	
		Could the conduct or interpretation of the index test have introduced bias?	No	
D.	Concerns regarding applicability	Is there concern that the index test, its conduct, or interpretation differ from the review question?	No	Low
Reference standard				
C.	Risk of bias	Is the reference standard likely to correctly classify the target condition?	Yes for those who underwent surgical biopsy but possibly not for those who didn't and underwent clinical follow-up – period may be inadequate (mean 39 months)	Moderate
		Were the reference standard results interpreted without the knowledge of the results of the index test?	No	
		Could the reference standard, its conduct, or its interpretation have introduced bias?	No	
D.	Concerns regarding applicability	Is there concern that the target condition as defined by the reference standard does not match the review question?	No	Low
Flow and timing				
B.	Risk of bias	Was there an appropriate interval between index test(s) and reference standard?	Yes	Low
		Did all patients receive a reference standard?	Yes	
		Did patients receive the same reference standard?	No – varied between patients; those who didn't undergo surgical biopsy were followed up clinically	
		Were all patients included in the analysis?	Yes	
		Could the patient flow have introduced bias?	Yes	
Overall study risk of bias:		Moderate for diagnostic outcomes due to possible inadequacy of reference standard in some patient; low for inadequate sampling outcome		

Abbreviations: CB, core biopsy; FNA, fine needle aspiration; QUADAS 2, Quality Assessment of Diagnostic Accuracy Studies 2; RoB, risk of bias.

Table AppC.4 Critical appraisal checklist – Mak 2012 (retrospective diagnostic cohort study/QUADAS 2)

Question			Answer	RoB rating
Patient selection				

[47] http://annals.org/aim/fullarticle/474994/quadas-2-revised-tool-quality-assessment-diagnostic-accuracy-studies.

Question			Answer	RoB rating
A.	Risk of bias	Was a consecutive or random sample of patients enrolled?	Yes – retrospective but all women who received both FNA and CB included.	Low
		Was a case-control design avoided?	Yes	
		Did the study avoid inappropriate exclusions?	Yes	
		Could the selection of patients have introduced bias?	No	
B.	Concerns regarding applicability	Is there a concern that the included patients do not match the review question?	No. Not specifically in a mammographic screening programme but all lesions non-palpable or small palpable ones (<1.5 cm)	Low
Patient selection				
A.	Risk of bias	Were the index tests interpreted without knowledge of the results of the reference standard?	Yes	Low
		Is a threshold was used, was it prespecified?	NR	
		Could the conduct or interpretation of the index test have introduced bias?	No	
B.	Concerns regarding applicability	Is there concern that the index test, its conduct, or interpretation differ from the review question?	No	Low
Reference standard				
A.	Risk of bias	Is the reference standard likely to correctly classify the target condition?	Unclear	Unclear
		Were the reference standard results interpreted without the knowledge of the results of the index test?	Unclear	
		Could the reference standard, its conduct, or its interpretation have introduced bias?	Unclear	
B.	Concerns regarding applicability	Is there concern that the target condition as defined by the reference standard does not match the review question?	Unclear	
Flow and timing				
A.	Risk of bias	Was there an appropriate interval between index test(s) and reference standard?	Unclear	Unclear
		Did all patients receive a reference standard?	Yes	
		Did patients receive the same reference standard?	Unclear	
		Were all patients included in the analysis?	Yes	
		Could the patient flow have introduced bias?	Unclear	
Overall study risk of bias:		Defined as low for outcome of interest to this review – inadequate sampling		

Abbreviations: CB, core biopsy; FNA, fine needle aspiration; NR, not reported; QUADAS 2, Quality Assessment of Diagnostic Accuracy Studies 2; RoB, risk of bias.

Table AppC.5 Critical appraisal checklist – Liikanen 2016 (prospective cohort study/JBI cohort)[48]

	Appraisal questions for cohort studies	Yes/No/Unclear/ Not applicable
1.	**Were the groups similar and recruited from the same population?**	**Partial yes**
	Comment: All patients were recruited from the Breast Surgery Unit of the Helsinki University Central Hospital, had not received neoadjuvant chemotherapy for a breast tumour and had a breast tumour less than or equal to 2cm in diameter. There was some variability in the tumour characteristics between the FNA group and core biopsy group.	
2.	**Were the exposures measured similarly to assign people to both exposed and unexposed groups?**	**Not applicable**
	Comment: Exposure was the application of FNA or core biopsy.	
3.	**Was the exposure measured in a valid and reliable way?**	**Not applicable**
	Comment: Exposure was the application of FNA or core biopsy.	
4.	**Were confounding factors identified?**	**Yes**
	Comment: The following potential confounding variables were tested in univariate analyses of survival: age, tumour size, pN category, breast tumour site, palpability, histological grade, tumour histology, tumour multifocality, ER status. PR status, MIB-1 expression, HER-2 status and chemotherapy.	
5.	**Were strategies to deal with confounding factors stated?**	**Yes**
	Comment: Multivariate Cox proportional hazards regression survival analysis was carried out to determine the impact of confounding factors on outcomes of interest. Age, tumour size, pN category, palpability, histological grade, ER status and MIB-1 expression included in multivariate analysis.	
6.	**Were the groups/participants free of the outcome at the start of the study (or at the moment of exposure)?**	**Partial yes**
	Comment: Yes for survival and local and regional lymph node recurrence. Unclear for distant metastasis.	
7.	**Were the outcomes measured in a valid and reliable way?**	**Yes**
	Comment: Measured by frequent follow up and follow up diagnostic testing if recurrence was suspected. Also, hospital case records, and the Finnish Cancer registry for recurrence, breast cancer-specific survival and overall survival.	
8.	**Was the follow up time reported and sufficient to be long enough for outcomes to occur?**	**Yes**
	Comment: Mean follow up for FNA and core biopsy groups were 116 months and 113 months, respectively. This is an adequate length of time to determine rates of overall survival and recurrence.	
9.	**Was follow-up complete, and if not, were the reasons to loss to follow-up described and explored?**	**Unclear**
	Comment: Dropout rates were not reported by the authors. It was noted that 191 deaths occurred during the follow up period and the time point of each death was presented graphically.	
10.	**Were strategies to address incomplete follow-up utilised?**	**Unclear/not applicable**
	Comment: It is unclear if there was incomplete follow up	
11.	**Was appropriate statistical analysis used?**	**Partial yes**
	Comment: Yes for overall survival (adjusted multivariate analysis); no for breast cancer death and recurrence/metastasis (unadjusted analysis only).	
	Overall risk of bias: low for overall survival; high for other outcomes.	

Abbreviations: ER, oestrogen receptor; FNA, fine needle aspiration; HER-2, human epidermal growth factor receptor-2; JBI, Joanna Briggs Institute; MIB-1, methylation-inhibited binding protein 1; pN category. Regional lymph nodes.

[48] http://joannabriggs.org/assets/docs/critical-appraisal-tools/JBI_Critical_Appraisal-Checklist_for_Cohort_Studies2017.pdf.

Table AppC.6 Critical appraisal checklist – Satchithananda 2005 (prospective cohort study/JBI cohort)

	Appraisal questions for cohort studies	Yes/No/Unclear/ Not applicable
1.	**Were the groups similar and recruited from the same population?**	**Unclear**
	Comment: No description of patient group characteristics were reported. Authors noted that all patients included were undergoing a breast biopsy procedure.	
2.	**Were the exposures measured similarly to assign people to both exposed and unexposed groups?**	**Not applicable**
	Comment: Exposure was the application of FNA or core biopsy.	
3.	**Was the exposure measured in a valid and reliable way?**	**Not applicable**
	Comment: Exposure was the application of FNA or core biopsy.	
4.	**Were confounding factors identified?**	**No**
	Comment: There was no explicit mention of confounding factors. Local anaesthesia was given to all patients undergoing core biopsy, but not all patients undergoing FNA, which is a major confounder for a pain outcome. The authors did discuss and measure the association of pain with multiple variables (radiologist, procedure type, number of cores and needle size) but results by procedure type (of interest to this review) do not appear to have been adjusted for needle size which was the other significant variable.	
5.	**Were strategies to deal with confounding factors stated?**	**No**
	Comment: Were stated but insufficient to deal with impact of anaesthesia	
6.	**Were the groups/participants free of the outcome at the start of the study (or at the moment of exposure)?**	**Yes**
	Comment: As the outcome was pain associated with biopsy procedure it can be assumed that this was not present prior to the biopsy. However, it is possible that patients were already experiencing pain before the procedure was performed.	
7.	**Were the outcomes measured in a valid and reliable way?**	**Yes**
	Comment: Pain levels were measured with a validated pain measurement tool. The tool was an adapted horizontal visual analog scale (VAS).	
8.	**Was the follow up time reported and sufficient to be long enough for outcomes to occur?**	**Not applicable**
	Comment: No follow up time reported as outcome was recorded immediately after the biopsy procedure.	
9.	**Was follow-up complete, and if not, were the reasons to loss to follow-up described and explored?**	**Not applicable**
	Comment: No reporting on follow-up completion as outcome was recorded immediately after the biopsy procedure.	
10.	**Were strategies to address incomplete follow-up utilized?**	**Not applicable**
	Comment: Strategies were not required to address incomplete follow-up as outcome data was recorded immediately after the biopsy procedure.	
11.	**Was appropriate statistical analysis used?**	**Yes**
	Comment: pain scores for FNA and core biopsy were not compared, and not adjusted for potential confounder local anaesthesia	
	Overall risk of bias: High due to confounding by pain relief	

Abbreviations: FNA, fine needle aspiration; JBI, Joanna Briggs Institute.

Table AppC.7 Critical appraisal checklist – Wong 2009 (retrospective cohort/JBI cohort)

	Appraisal questions for cohort studies	Yes/No/Unclear/ Not applicable
1.	**Were the groups similar and recruited from the same population?**	**Partial yes**
	Comment: Patients were from a single regional hospital in Hong Kong. The only patient characteristics provided were the mean age and the size of the lesion. The mean ages were similar. All patients in the FNA group had a lesion size 8mm or less and all patients in the core biopsy group had lesions larger than 8mm. This is a considerable difference in the patient groups. It was also stated by the authors that patients "…having a repeat procedure due to prior failure to obtain a pathological diagnosis were excluded."	
2.	**Were the exposures measured similarly to assign people to both exposed and unexposed groups?**	**Not applicable**
	Comment: Exposure was the application of FNA or core biopsy.	
3.	**Was the exposure measured in a valid and reliable way?**	**Not applicable**
	Comment: Exposure was the application of FNA or core biopsy.	
4.	**Were confounding factors identified?**	**Yes**
	Comment: The authors identified that a lack of patient background information and lack of randomisation were possibly confounding factors. As noted above, tumour size was a potential confounding factor. Local anaesthesia was given to all patients, regardless of biopsy method used.	
5.	**Were strategies to deal with confounding factors stated?**	**No**
	Comment: However, all patients received pain relief, which negates a major potential confounder for a pain outcome.	
6.	**Were the groups/participants free of the outcome at the start of the study (or at the moment of exposure)?**	**Yes**
	Comment: As the outcome was pain associated with biopsy procedure it can be assumed that this was not present prior to the biopsy. However, it is possible that patients were already experiencing pain before the procedure was performed.	
7.	**Were the outcomes measured in a valid and reliable way?**	**Unclear**
	Comment: Patients self-reported pain score on a scale of 0 (no pain) to 10 (most severe pain). There is no more information on how this data was collected.	
8.	**Was the follow up time reported and sufficient to be long enough for outcomes to occur?**	**Not applicable**
	Comment: Outcome data was collected immediately following the procedure.	
9.	**Was follow-up complete, and if not, were the reasons to loss to follow-up described and explored?**	**Not applicable**
	Comment: Outcome data was collected immediately following the procedure.	
10.	**Were strategies to address incomplete follow-up utilized?**	**Not applicable**
	Comment: Outcome data was collected immediately following the procedure.	
11.	**Was appropriate statistical analysis used?**	**Unclear**
	Comment: Statistical methods were not described by the authors.	
	Overall risk of bias: Moderate for pain outcome. No confounding due to differential pain relief. Possible confounding due to tumour size but not considered to confer a high risk of bias for pain outcome.	

Table AppC.8 Critical appraisal checklist – Hukkinen 2008 (retrospective cohort study/JBI cohort)

	Appraisal questions for cohort studies	Yes/No/Unclear/ Not applicable
1.	**Were the groups similar and recruited from the same population?**	**Partial yes**
	Comment: The groups were recruited from the Breast Surgery Unit of Helsinki University Central Hospital. Patient and lesion characteristics were also recorded indicating that the groups were mostly similar between those having initial FNA or CB with the exception of lesion size, mammographic finding of mass plus microcalcification, mammographic finding of invisible, and postoperative diagnosis of invasive lobular carcinoma.	
2.	**Were the exposures measured similarly to assign people to both exposed and unexposed groups?**	**Not applicable**
	Comment: Exposure was the application of FNA or core biopsy.	
3.	**Was the exposure measured in a valid and reliable way?**	**Not applicable**
	Comment: Exposure was the application of FNA or core biopsy.	
4.	**Were confounding factors identified?**	**Unclear**
	Comment: The authors made no explicit comments regarding confounding factors. However, data for a variety of variables associated with the patient and the tumour were collected.	
5.	**Were strategies to deal with confounding factors stated?**	**No**
	Comment:	
6.	**Were the groups/participants free of the outcome at the start of the study (or at the moment of exposure)?**	**Not applicable**
	Comment: The outcome related to follow-up testing and thus could not have been present at the time of exposure or before the exposure.	
7.	**Were the outcomes measured in a valid and reliable way?**	**Yes**
	Comment: As the outcome related to follow-up, access to patient records enabled the authors to measure this outcome reliably.	
8.	**Was the follow up time reported and sufficient to be long enough for outcomes to occur?**	**Yes**
	Comment: The outcome of interest was cascade testing. As patient data was collected retrospectively from the point of surgery, the follow up time was appropriate to capture this outcome.	
9.	**Was follow-up complete, and if not, were the reasons to loss to follow-up described and explored?**	**Yes**
	Comment: Cascade testing was captured for all patients that underwent surgery.	
10.	**Were strategies to address incomplete follow-up utilized?**	**Not applicable**
	Comment: See above comment.	
11.	**Was appropriate statistical analysis used?**	**Unclear**
	Comment: No statistical analysis has been conducted by the authors for this outcome. However, this is likely to be because no direct comparisons between the FNA group and core biopsy group for this outcome have been made by the authors.	
	Overall risk of bias: High. There were some differences in tumour characteristics between groups and these may have resulted in differences in additional testing, rather than the additional testing being related to the initial test.	

Abbreviations: CB, core biopsy; FNA, fine needle aspiration; JBI, Joanna Briggs Institute.

Appendix D Additional results

Table AppD.1 Wang 2017 individual study additional diagnostic accuracy results[49]

	PLR		NLR		PPV		NPV		Accuracy	
Study ID	FNA% (95% CI)	CB% (95% CI)	FNA% (95% CI)	CB% (95% CI)	FNA% (95% CI)	CB% (95% CI)	FNA% (95% CI)	CB% (95% CI)	FNA% (95% CI)	CB% (95% CI)
Saha 2016	-	-	0.31 (0.20–0.31)	0.17 (0.08–0.33)	100	100	38 (28–49)	53 (37–69)	74 (60–85)	86 (73–94)
Tikku 2015	-	-	0.35 (0.24–0.52)	0.04 (0.01–0.16)	100	100	69 (60–76)	95 (83–99)	80 (70–88)	98 (92–100)
Altaf 2015	-	10 (3.2–32)	0.06 (0.01–0.42)	0.34 (0.16–0.70)	100	79 (54–92)	98 (87–100)	89 (80–94)	98 (91–100)	87 (75–94)
Ahmed 2010	-	11 (2.8–40)	0.07 (0.03–0.19)	0.08 (0.03–0.20)	98 (89, 100)	96 (88–99)	85 (68–93)	84 (67–93)	94 (86–98)	93 (84–97)
Leaver 2010	-	-	0.39 (0.28–0.54)	0.17 (0.09–0.30)	100	100	65 (57–72)	81 (70–89)	77 (68–85)	90 (82–95)
Barra 2008	-	-	0.14 (0.10–0.20)	0.12 (0.08–0.17)	100	100	57 (49–64)	62 (53–70)	88 (83–92)	90 (86–93)
Garg 2007	-	-	0.06 (0.02–0.24)	0.09 (0.03–0.28)	100	100	90 (70–97)	86 (67–95)	96 (86–100)	94 (83–99)
Wu 2004	4.2 (1.0–17)	13 (2.0–90)	0.51 (0.24–1.09)	0.12 (0.02–0.76)	71 (38–91)	89 (54–98)	76 (60–87)	93 (69–99)	75 (53–90)	92 (73–99)
Dennison 2003	3.4 (2.0–5.9)	-	0.13 (0.07–0.24)	0.05 (0.02–0.11)	90 (85–94)	100	74 (60–84)	88 (76–95)	86 (79–91)	97 (92–99)
Leifland 2003	50 (7.1–350)	17 (6.5–44)	0.33 (0.29–0.38)	0.10 (0.07–0.13)	100 (98–100)	99 (98–100)	33 (30–36)	63 (55–69)	72 (68–76)	91 (88–93)
Cangiarella 2000	54 (14–218)	31 (9.8–97)	0.25 (0.14–0.45)	0.37 (0.24–0.57)	93 (77–98)	88 (71–96)	94 (90–97)	92 (88–94)	94 (90–97)	91 (86–95)
Dowlatshahi 1991	10 (5.2–20)	45 (11–180)	0.50 (0.39–0.63)	0.49 (0.39–0.62)	82 (69–90)	95 (83–99)	82 (78–85)	82 (79–85)	82 (77–87)	84 (79–89)

Abbreviations: CB, core biopsy; CI, confidence interval; FNA, fine needle aspiration; NLR, negative likelihood ratio; NPV, negative predictive value; PLR, positive likelihood ratio; PPV, positive predictive value.

[49] Calculated from data available in Wang 2017 using MedCalc (https://www.medcalc.org/calc/diagnostic_test.php).