National Institute for Health and Care Excellence

FINAL

Thyroid disease: assessment and management

[O] Ultrasound guidance for fine needle aspiration

NICE guideline NG145

Diagnostic evidence review underpinning recommendations 1.9.1 to 1.9.6 in the guideline. See also evidence review N

2019

FINAL

Developed by the National Guideline Centre, hosted by the Royal College of Physicians

Disclaimer

The recommendations in this guideline represent the view of NICE, arrived at after careful consideration of the evidence available. When exercising their judgement, professionals are expected to take this guideline fully into account, alongside the individual needs, preferences and values of their patients or service users. The recommendations in this guideline are not mandatory and the guideline does not override the responsibility of healthcare professionals to make decisions appropriate to the circumstances of the individual patient, in consultation with the patient and, where appropriate, their carer or guardian.

Local commissioners and providers have a responsibility to enable the guideline to be applied when individual health professionals and their patients or service users wish to use it. They should do so in the context of local and national priorities for funding and developing services, and in light of their duties to have due regard to the need to eliminate unlawful discrimination, to advance equality of opportunity and to reduce health inequalities. Nothing in this guideline should be interpreted in a way that would be inconsistent with compliance with those duties.

NICE guidelines cover health and care in England. Decisions on how they apply in other UK countries are made by ministers in the Welsh Government, Scottish Government, and Northern Ireland Executive. All NICE guidance is subject to regular review and may be updated or withdrawn.

ISBN: 978-1-4731-3595-6

Contents

1

1 Ultrasound guidance for fine needle aspiration

1.1 Review question: Should a fine-needle aspiration be under ultrasound guidance?

1.2 Introduction

Fine Needle Aspiration (FNA) of the thyroid is a minimally invasive method to obtain tissue for cytological assessment and classification of malignancy risk, commonly using the Royal College of Pathologists grading system (which is similar to the US Bethesda system). The FNA has historically been performed through palpation guidance although in recent years, common practice has seen this become more routinely performed under ultrasound guidance. This latter change in practice has been largely driven in an attempt to reduce the rate of inadequate samples that occur in tissue sampling. It is recognised that while there are specialty society guidelines for practice there are no formal guidelines that demand imaging guided over palpation guided FNA, or vice-versa.

This review seeks to assess both the evidence base and the cost effectiveness of these two methods of FNA to identify if there is a clinical and/or financial benefit to one over the other.

1.3 PICO table

For full details see the review protocol in Appendix A:.

Table 1: PICO characteristics of review question

Population	People with thyroid nodules
Target condition	Malignancy
Index test **Comparator**	Ultrasound guided fine needle aspiration cytology (UGFNAC) Palpation guided fine needle aspiration cytology (PGFNAC)
Reference standard	Diagnosis of malignancy on core biopsy or later surgery
Statistical measures **Outcomes**	Sensitivity Specificity Inadequate sample (dichotomous)
Study design	Diagnostic accuracy studies

As per the full protocol, evidence was extracted preferentially from studies in which at least some of the participants had both UGFNAC and PGFNAC in order to provide the most direct comparative evidence. The committee agreed this evidence was sufficient for decision making.

The committee noted that while this review was focused on accuracy type data, studies also reported the rates that each testing strategy returned inadequate samples. The committee agreed that the most appropriate way to handle this important information was to extract the ratio of inadequate sampling of each strategy as per an intervention review.

1.4 Clinical evidence

1.4.1 Included studies

Five studies were included in the review[11, 18, 21, 46, 50]; these are summarised in Table 2 below. Evidence from these studies is summarised in the clinical evidence summary below (Table 3).

All studies assessed the diagnostic accuracy of UGFNAC compared to PGFNAC using histopathological findings (surgery) as the reference standard. 100% of participants underwent both tests in two studies, with the majority of patients undergoing PGFNAC in two studies while all patients underwent UGFNAC with the minority undergoing both tests in one study. None of the included studies were conducted in Europe. Diagnostic accuracy outcome measures were calculated based on the number of participants for which histopathological data was available in each study.

See also the study selection flow chart in Appendix C:, sensitivity and specificity forest plots in Appendix E:, and study evidence tables in Appendix D:.

1.4.2 Excluded studies

See the excluded studies list in Appendix I:.

1.4.3 Summary of clinical studies included in the evidence review

Table 2: Summary of studies included in the evidence review

Study	Population	Target condition	Index tests	Reference standard	Comments
Cesur 2006 [11]	Adults: n=215, mean age (SD) 48.7 (13.5) with 1-3 palpable nodules (diameter: 1 - 2.5 cm) Turkey	Thyroid cancer	UGFNAC PGFNAC 100 % of patients underwent both tests	Histopathology	Surgery performed in 13 patients (26 nodules)
Jalan 2017 [18]	Patients: n=84, age range 8-71; UG findings available in n=36 India	Thyroid cancer	UGFNAC PGFNAC 43% (n=36) gave consent for both tests, 57% (n=48) underwent PGFNAC only	Histopathology	Histopathology was available in 40 cases (18 from PGFNA, 22 from combined PG-& UG-FNA)
Krishnappa 2013 [21]	Patients: n=91, mean age (range) 38.5 (8-80); 83.5% euthyroid, 16.5% signs and symptoms of hyperthyroidism, 3.3% previous thyroid surgery India	Thyroid cancer	UGFNAC PGFNAC 100 % of patients underwent both tests	Histopathology	Surgery performed in 25 patients
Takashima 1994 [46]	Patients: n=210, mean age (range) 53 (12-88); 268 aspirated nodules	Thyroid cancer	UGFNAC PGFNAC	Histopathology	Histopathologic confirmation in 34 patients (62 nodules) Thyroid disease n=72, neck radiation therapy or surgery

Study	Population	Target condition	Index tests	Reference standard	Comments
	Japan		27% (n=57) of patients underwent both tests, all had UGFNAC		or both n=15, history of cancer at other site n=22
Zawawi 2016 [50]	Patients: n=150, mean age 41.6; 183 FNAs Saudi Arabia	Thyroid cancer	UGFNAC PGFNAC 77 UGFNACs, 151 PGFNACs; unclear number of patients undergoing both tests.	Histopathology	Unclear availability of histopathological confirmation

See appendix D for full evidence tables.

1.4.4 Quality assessment of clinical studies included in the evidence review

Table 3: Clinical evidence summary: UGFNAC vs PGFNAC, diagnostic accuracy

Index Test (Threshold)	Number of studies	n	Quality	Specificity % (95% CI)	Sensitivity % (95% CI)
UGFNAC	5	750	LOW[a,c] due to risk of bias, imprecision	86 (72 to 96)	90 (76 to 98)
PGFNAC	5	750	VERY LOW[a,b,c] due to risk of bias, imprecision, inconsistency	82 (59 to 96)	71 (48 to 87)

The assessment of the evidence quality was conducted with emphasis on sensitivity as this was identified by the committee as the primary measure in guiding decision-making.

(a) Risk of bias was assessed using the QUADAS-2 checklist. The evidence was downgraded by 1 increment if the majority of studies were rated at high risk of bias, and downgraded by 2 increments if the majority of studies were rated at very high risk of bias.

(b) Inconsistency was assessed by inspection of the sensitivity and specificity plots. Particular attention was placed on the sensitivity threshold set by the committee as an acceptable level to recommend a test. The evidence was

- *downgraded by 1 increment if the individual study values varied across 2 areas: where values of individual studies are both above and below 50%, or both above and below the acceptable threshold 90%*
- *downgraded by 2 increments if the individual study values varied across 3 areas, where values of individual studies are above and below 50%, and also above and below the acceptable threshold 90%*

(c) Imprecision was assessed based on inspection of the confidence region of sensitivity in the diagnostic meta-analysis

Table 4: Clinical evidence summary: UGFNAC vs PGFNAC, inadequate sample

Outcomes	No of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with PGFNAC	Risk difference with UGFNAC (95% CI)
Inadequate sample	750 (5 studies)	VERY LOW[1] due to risk of bias	0.56 (0.44 to 0.72)	172 per 1000	76 fewer per 1000 (from 96 fewer to 48 more)

[1] Downgraded by 1 or 2 increments because the majority of the evidence was at high risk of bias or very high risk of bias

1.5 Economic evidence

1.5.1 Included studies

One health economic study with the relevant comparison has been included in this review. [11] This is summarised in the health economic evidence profile below (Table 5) and the health economic evidence table in Appendix G:.

1.5.2 Excluded studies

One economic study relating to this review question was identified but was excluded due to limited applicability. [8]This is listed in appendix I, with reasons for exclusion given.

See also the health economic study selection flow chart in Appendix F:.

1.5.3 Summary of studies included in the economic evidence review

Table 5: Health economic evidence profile: Palpation-guided fine-needle aspiration cytology (PGFNAC) versus Ultrasound-guided fine-needle aspiration cytology (UGFNAC)

Study	Applicability	Limitations	Other comments	Incremental cost	Incremental effects	Cost effectiveness	Uncertainty
Cesur 2006 [4, 114, 114, 11] (Turkey)	Partially applicable(a)	Potentially serious limitations(b)	Diagnostic cohort study	+£13(c)	+0.009 extra cancers detected (d)	£1,361 per extra cancer detected	No sensitivity analysis was conducted.
National Guideline Centre – original guideline model (UK)	Partially applicable(e)	Minor limitations	Decision tree model	With repeat test(f): -£21.00 Without repeat test -£58.00	With repeat test: +0.016 cancers detected Without repeat test +0.006 cancers detected	With repeat test: UGFNAC dominant Without repeat test UGFNAC dominant	Various one-way sensitivity analyses With repeat test: UGFNAC dominant except when the price of UGFNAC was increased by 50% - £5,203 per cancer or surgery cost high- £183 per cancer detected Without repeat test: UGFNAC dominant except when the price of UGFNAC was increased by 50% - £15,162 per cancer

(a) Turkish health service perspective; outcomes were not valued using QALYs.
(b) Data taken from single study of 215 patients; currency and cost year not stated, costs taken from Turkish hospitals (private and state hospitals); no sensitivity analysis undertaken.
(c) 2006 US Dollars, presented as UK pound. US dollars converted using 2006 purchasing power parities [40]. Costs incorporated are: prices of thyroid ultrasonography, PGFNAC, UGFNAC and cytologic examinations.
(d) Two extra cancers detected in the whole population
(e) Quality-adjusted life-years were not calculated and only costs for the diagnostic pathway were used.
(f) Test repeated after an initial benign test result.

1.5.4 Health economic modelling

This area was prioritised for new cost-effectiveness analysis. The economic analysis was to determine the most cost-effective diagnostic strategy when testing with Fine-Needle Aspiration Cytology (FNAC) to detect thyroid malignancy and treat patients. This will compare the different diagnostic strategies for ultrasound guided FNAC (UGFNAC) and palpation guided FNAC (PGFNAC) with and without repeating tests after a benign diagnosis.

Thyroid nodules are common, and 4-7% of all thyroid nodules are found to be malignant. After preliminary investigation using clinical evaluation and ultrasound, people presenting with thyroid enlargement receive FNAC where there is a suspicion of thyroid cancer. FNAC is the most accurate and reliable tool for diagnosing thyroid malignancy and it can be performed under palpation guidance (PG) or ultrasound guidance (UG). UG is the more accurate approach but has a higher unit cost.

Therefore, original cost-effectiveness modelling was undertaken for this question. A summary is included here. Evidence statements summarising the results of the analysis can be found below. The full analysis can be found in Supplement 2.

1.5.4.1 Methods

A cost-consequence analysis was conducted comparing different diagnostic strategies for UGFNAC and PGFNAC. A decision tree was used to estimate short-term benefits and costs from a current UK NHS and personal social services perspective (PSS). In addition, the committee wished to explore the impact of different estimates of prevalence, costs of FNAC for both UG and PG, the cost of surgery and the diagnostic accuracies of the different tests.

The modelled population was people with an enlarged but normally functioning thyroid gland being investigated for possible malignancy after a positive ultrasound scan (USS).

The committee agreed that an USS should be the preliminary investigation method to aid decision-making about which nodules to perform FNAC and it is current practice in the UK. The committee noted that only those with U3-U5 grade on USS (U3 indeterminate, U4 suspicious for malignancy, and U5 likely malignant) would be referred for a FNAC and it is these people specifically who are the subject of the model.

There are different pathways that can be followed when carrying out PGFNAC or UGFNAC tests.

The following diagnostic strategies were chosen as comparators:

- UGFNAC without repeat after an initial benign diagnosis (‘UGFNAC without benign repeat’);
- UGFNAC with repeat after an initial benign diagnosis (‘UGFNAC with benign repeat’);
- PGFNAC without repeat after an initial benign diagnosis (‘PGFNAC without benign repeat’);
- PGFNAC with repeat after an initial benign diagnosis (‘PGFNAC with benign repeat’).

A decision tree was used to calculate the proportion of the population that fall into one of a number of cohorts according to their test result. The decision tree calculates the proportion of patients who will receive a false negative (FN), false positive (FP), true negative (TN), true positive (TP) diagnosis according to the sensitivity, specificity and prevalence data.

The committee considered that the after FNAC the most likely procedure would be surgery to remove part of the thyroid (hemithyroidectomy) as it can be used as both a diagnostic tool and a treatment. The surgery would identify the true condition.

Therefore, the outcomes for the FNAC test included in the model to make sure the model reflects the clinical pathway are as follows;

- malignant; Thy5(diagnostic of malignancy) and Thy3F (follicular neoplasm)
- benign; Thy2(non-neoplastic)
- indeterminate; Thy3A (neoplasm possible with atypical features) and Thy4 (suspicious)
- inadequate; Thy1 (non-diagnostic)

Patients identified as malignant after a single FNAC are referred directly to surgery. Patients identified as benign are either discharged or referred to a repeat FNAC and this forms part of the variation in the comparators.

After repeating the FNAC, those patients identified as malignant, indeterminate, and inadequate are referred to surgery. Only those patients identified as benign are discharged.

In patients with thyroid cancer, the probability that the PG or UG FNAC test is positive (malignancy detected) is determined by the test sensitivity. Therefore, the probability that the test is negative, which means the test failed to detect the malignancy, is 1 – sensitivity.

To determine the proportion of patients that received a benign, indeterminate, or inadequate test result, a weighted average was calculated using a study that was identified that was included in both the clinical and economic evidence review (Cesur et al 2006). [2]

For patients with cancer, a TP result is assigned if they are identified as malignant, indeterminate, or inadequate after their final FNAC. FN results are only assigned to those patients exiting the model as benign.

In patients who do not have cancer, the probability that FNAC test is negative is determined by the test specificity. For these patients, the probability that the FNAC test is positive is 1 – specificity.

For patients without cancer, they are assigned as TN status if they receive a benign result for their final FNAC, and therefore are discharged without surgery. FP test results are those that received surgery for thyroid cancer i.e. those patients identified as malignant, indeterminate, or inadequate after their final FNAC.

For more detailed explanation of the model structure, please refer to the technical report in Supplement 2.

A number of assumptions were made when developing the model and a sensitivity analyses were undertaken in areas of uncertainty to see how robust the model results are. The sensitivity analyses are outlined below but are also discussed in more detail in Supplement 2:

- cancer prevalence
- cost of UGFNAC and PGFNAC
- cost of surgery
- cost of FN (delayed diagnosis)
- ultrasound sensitivity and specificity
- UGFNAC sensitivity and specificity
- PGFNAC sensitivity and specificity

Model inputs were based on clinical evidence identified in the systematic review undertaken for the guideline, supplemented by additional data sources as required. These are described in full in the technical report in Supplement 2. All model inputs and assumptions were validated by the guideline committee, see Table 6 for a summary of the base case model inputs used in the model.

Table 6: Summary of the base case model inputs used in the model

Parameter description	Point estimate	Source	Distribution
Diagnosis parameters			
Prevalence of cancer among patients with a normally functioning but enlarged thyroid	0.05	Borget, et al 2018[1]	Beta
Positive predictive value (PPV) of US	0.115	Calculation	Function of the prevalence of cancer above and the Sensitivity and Specificity of ultrasound
Sensitivity of US	0.904	Persichetti 2018 [7]	Function of the prevalence and DOR of US
Specificity of US	0.634	Persichetti 2018 [7]	Beta
Diagnostic odds ratio (DOR) of US	16.295	Function of sensitivity and specificity	Log Normal
Sensitivity of UGFNAC	0.900	Pooled estimate	Sampled from the joint distribution from WinBUGS
Specificity of UGFNAC	0.865	Pooled estimate	Sampled from the joint distribution from WinBUGS
Sensitivity of PGFNAC	0.71	Pooled estimate	Sampled from the joint distribution from WinBUGS
Specificity of PGFNAC	0.82	Pooled estimate	Sampled from the joint distribution from WinBUGS
Cost (£)			
UGFNAC	£295	Committee member	Gamma
PGFNAC	£242	Committee member	Gamma
Surgery	£3,689	NHS reference costs 2016/17	Gamma
FN cost (delayed diagnosis)	£4,197	NHS reference costs 2016/17	Gamma

Abbreviations: US: ultrasound; UGFNAC: Ultrasound guided fine-needle aspiration cytology; PGFNAC: Palpation guided fine-needle aspiration cytology; FN: false negatives

1.5.4.2 Results

The base-case results are presented below. For a full write up of the model results and sensitivity analyses see Supplement 2.

UGFNAC without benign repeat was found to be the lowest cost option and had the least false positive results. It was dominant compared to PGFNAC without benign repeat because it detected more cancers at a cheaper cost.

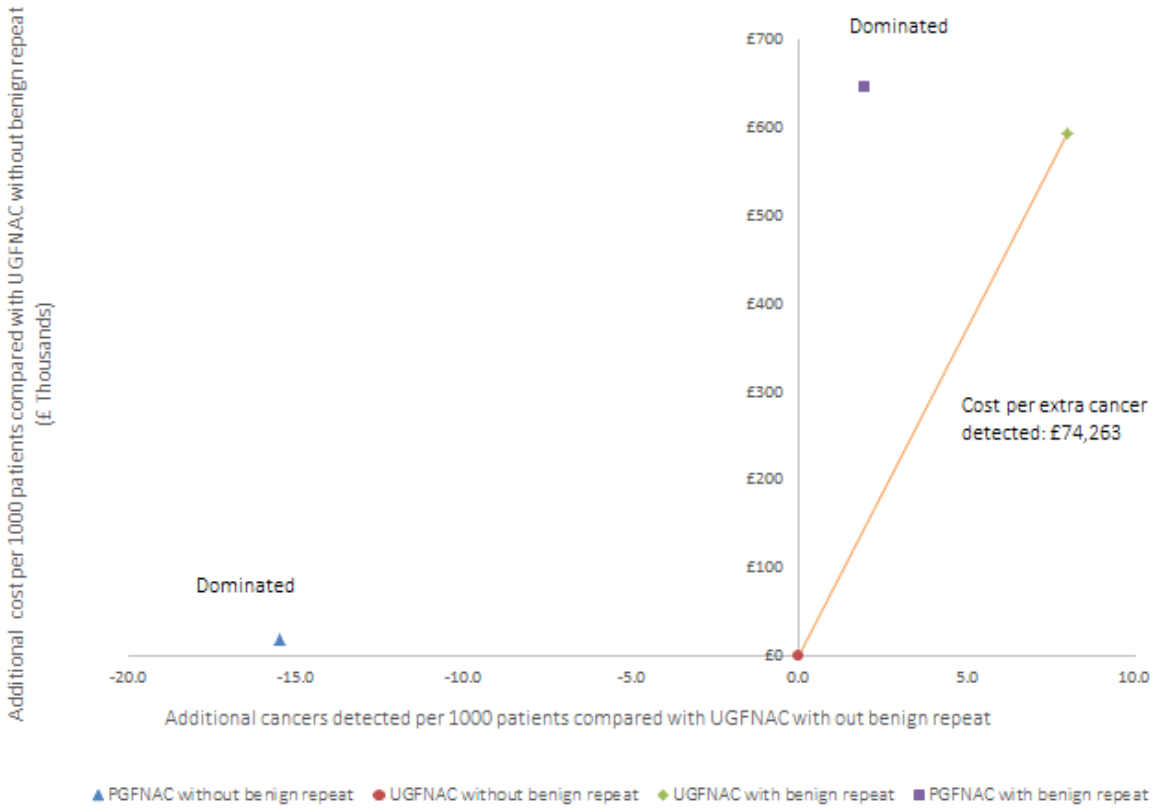
UGFNAC with benign repeat was more effective at detecting cancers and more costly compared to UGFNAC without benign repeat with a cost per extra cancer detected of £74,263.

UGFNAC with benign repeat was dominant compared to PGFNAC with benign repeat as PGFNAC with benign repeat was more costly and less effective in detecting cancer. Results are summarised below in Table 7. The incremental costs and true positives from the probabilistic analysis have also been presented graphically on the cost-effectiveness plane, Figure 1.

Table 7: Base case analysis results per 1000 patients in order of cost (probabilistic analysis)

Strategy	Costs	Cancers detected (True Positives)	Additional Cost (compared with row above)	Additional cancers detected (compared with row above)	Additional cost per extra cancer detected
UGFNAC without benign repeat	£858,932	105	-	-	-
PGFNAC without benign repeat	£879,936	90	£21,004	-15.7	Dominated
UGFNAC with benign repeat	£1,451,488	113	£571,551	23.7	£74,263 (vs UGFNAC without benign repeat)
PGFNAC with benign repeat	£1,509,489	107	£58,002	-6.2	Dominated

Figure 1: Base case cost-effectiveness plane showing the different diagnostic strategies (probabilistic)



Several analyses were run in order to see what effect they had on the cost per cancer detected. This includes prevalence, costs, and the sensitivity and specificity of the different tests.

One- way sensitivity analyses were run deterministically, and the results are summarised below. These showed that in general, changes in the cost of test or treatment do not result in very different estimates of the cost per cancer detected.

The PGFNAC without benign repeat versus the UGFNAC without benign repeat, the four analyses that resulted in a change in cost effectiveness were:

- a drop in the cost of PGFNAC;
- an increase in the costs of UGFNAC;
- increase in the surgery cost; and
- a drop in the FN cost.

In each case, PGFNAC was no longer dominated but for UGFNAC the additional cost per cancer detected was low.

The cost per cancer detected for UGFNAC with benign repeat versus UGFNAC without benign repeat was stable with respect to changes of the prevalence and costs.

In most of the analyses, the PGFNAC with benign repeat was dominated (higher costs and lower true positives) by UGFNAC with benign repeat, except in two analyses where they become less costly and but also detected fewer cancers (true positives). This occurred when

- the cost of UGFNAC increased and
- the cost of PGFNAC was reduced.

1.5.4.3 Limitations and interpretation

This analysis suggests that UGFNAC without benign repeat had a relatively low cost per extra cancer for diagnosing thyroid cancer in patients with positive US scan results. Many uncertainties in the model structure and assumptions were explored in sensitivity analyses.

The primary limitation is the uncertainty around the cost and health consequences of missing a cancer. For simplicity of the model, it was assumed that all FN will re-present later and would be correctly diagnosed as the number of FN that do not re-present or may re-present years later was difficult to model. The committee noted that patients who are US positive and have cancer are more than likely re-present, but the small proportion that might not was difficult to quantify and was not believed to have a substantial effect on the results. However, as the FN costs were consensus based, it was tested in the sensitivity analysis.

The second limitation of this model is that the diagnostic accuracy data for the US scan was taken from one diagnostic accuracy study. A meta-analysis was discussed but it was decided that for a meaningful meta-analysis, five or more studies were needed. The committee agreed on choosing one study to represent best available evidence, study by Persichetti 201842 that was more representative of UK current practice.

A third limitation is that it's unlikely that initial and subsequent tests would be fully independent of one another - for example, sensitivity of UGFNAC is probably less than 90% after an initial negative test result. This means that the cost effectiveness of UGFNAC+ benign repeat vs UGFNAC without benign repeat is likely to be even worse than seen in this analysis.

A fourth limitation of this model is that some structural assumptions were required with little clinical evidence to allow direct estimates to be made. In particular, it is difficult to test the assumptions made about the suspicious results that were grouped together with the indeterminate (Thy3A) results. The committee had a lengthy discussion to split the group into indeterminate and suspicious but there was no consensus and the clinical evidence did not help quantify this issue. It was therefore agreed that for simplicity of the model, they are to be grouped together.

1.6 Evidence statements

1.6.1 Clinical evidence statements

Five studies that evaluated the two diagnostic tests were included in the review. Of these, the committee noted that. The evidence was of low to very low quality.

- **UGFNAC**: Low quality evidence from 5 studies with 750 participants showed that UGFNAC has a specificity of 86% and a sensitivity of 90%.
- **PGFNAC**: Very low quality evidence from 5 studies with 750 participants showed that UGFNAC has a specificity of 82% and a sensitivity of 71%.

Five studies reported inadequate sample rates. There was no clinically important difference in inadequate sample rates (very low quality).

1.6.2 Health economic evidence statements

- One cost-effectiveness analysis found that in adults with nodular goitre, UGFNAC was more costly and more effective than PGFNAC for detecting malignancy (ICER: £1,361 per extra cancer detected). This analysis was assessed as partially applicable with potentially serious limitations.
- An original cost-consequence analysis found that
  - PGFNAC with a repeat test* was dominated by UGFNAC with a repeat test*

- PGFNAC without a repeat test was dominated by UGFNAC without a repeat test
- UGFNAC with a repeat test* cost an extra £74,263 per extra cancer detected compared to UGFNAC without a repeat test
- *FNAC was repeated after an initial benign test result.
- This was rated as partially applicable with minor limitations.

1.7 The committee's discussion of the evidence

1.7.1 Interpreting the evidence

1.7.1.1 The diagnostic measures that matter most

The committee considered the diagnostic measures of sensitivity, specificity, positive and negative predictive value of the index tests for diagnosing thyroid cancer. The rate of inadequate sample that each test returned that was reported in the evidence was also considered important by the committee and was therefore taken into account. The sensitivity of tests was deemed the most important measure in this review. There was agreement on the importance of identifying all patients with thyroid cancer and the serious consequences associated with a missed diagnosis of the condition. Thus, sensitivity was prioritised for decision making.

1.7.1.2 The quality of the evidence

Clinical evidence for the diagnostic accuracy of UGFNAC and PGFNAC was available from five two gate diagnostic accuracy studies. Evidence for sensitivity and specificity was of low and very low quality for those tests respectively. The evidence for both tests was downgraded due to risk of bias and imprecision. Evidence for the PGFNAC was furthermore downgraded for inconsistency. Clinical evidence for inadequate sample rates was also available from five studies. This was of very low quality due to risk of bias partly because the studies were non-randomised. Overall, the clinical evidence was derived from studies including a total of 750 participants, not all of which had undergone both index tests. In addition, the diagnostic accuracy evidence was based on a limited number of patients for which histopathological confirmation was available.

The committee noted that the diagnostic accuracy evidence was in regard to palpable nodules that were investigated in the studies included in the present review. The size of nodules was also raised as an important factor that could influence diagnostic accuracy. Specifically, the committee agreed that decision making should ideally be based on the sensitivity and specificity of the tests for small size nodules as well, which was not currently available.

1.7.1.3 Benefits and harms

Evidence for the diagnostic accuracy of UGFNAC compared to PGFNAC suggested that for the former index test both measures of sensitivity and specificity were higher. Considering that sensitivity was prioritised for decision making, the considerable discrepancy of almost 20% in sensitivity that was identified between the two tests was noted by the committee.

Based on the diagnostic accuracy evidence and the inadequate sample results of the index tests and their clinical experience, the committee agreed on offering UGFNAC when performing FNAC for thyroid nodules.

The committee emphasised an additional benefit associated with ultrasound guidance, in that it can provide information about the sonographic characteristics of a nodule and its malignancy status prior to the use of a needle.

Evidence suggested no clinically important difference of UGFNAC compared to PGFNAC in terms of inadequate sample. The lower rate of inadequate sample that UGFNAC returned, despite being deemed not clinically important based on the pre-specified cut off (100 per 1000) employed, was noted by the committee and taken into account in decision making. The committee noted that any increase in inadequate samples would lead to an increase in the need for repeat sampling.

1.7.2 Cost effectiveness and resource use

One economic analysis was included in the economic literature review that assessed cost effectiveness in terms of cost per cancer avoided from a Turkish perspective. It compared palpation-guided fine-needle aspiration cytology (PGFNAC) with ultrasound-guided fine-needle aspiration cytology (UGFNAC) for the diagnosis of malignancy of thyroid nodules. In addition, original economic analysis was undertaken for this question. This assessed the short-term benefits and costs in terms of cost per cancer avoided from a current UK NHS and personal social services perspective. It compared four different diagnostic strategies for Ultra-sound guided fine-needle aspiration cytology (UGFNAC) and palpation guided fine-needle aspiration cytology (PGFNAC) with and without repeat after a benign diagnosis, which can be followed when carrying out FNAC.

In the published Turkish analysis, PGFNAC had a slightly lower mean cost per patient (£51) than UGFNAC (£64). The costs included the costs of the thyroid ultrasonography, PGFNAC, UGFNAC and cytologic examinations. It was also less effective with a true positive rate of 1.89% compared to 2.79%. The incremental cost effectiveness ratio for UGFNAC compared to PGNAC was £1,361 per extra cancer detected. The study was assessed as partially applicable as it did not utilise an NHS perspective and used unit costs from a Turkish health service (state and private hospital) perspective in 2006. The study also did not report outcomes in terms of QALYs. It was also assessed to have potentially serious limitations as the estimates of relative treatment effects are based on the single study of 215 patients and not based on meta-analysis of all the available evidence identified in the clinical review for the guideline. Some costs were taken from private hospitals and may be overestimated. Additionally, no sensitivity analysis was undertaken to adequately assess parameter uncertainty.

Original modelling was done for this review because of the potentially serious limitations and partial applicability of the Turkish analysis, and because UGFNAC appeared more costly and more effective than PGFNAC. An original cost-consequence analysis found that UGFNAC without benign repeat was the cheapest option and was dominant compared to the PGFNAC without benign repeat (less costly and more effective in detecting cancer). PGFNAC with benign repeat was dominated by UGFNAC with benign repeat, as it is less costly and more effective at detecting cancer. The committee noted that the UGFNAC with benign repeat is unlikely to be cost effective compared to UGFNAC without benign repeat as the cost per extra cancer detected £74,263, was considered relatively high. The committee concluded that UGFNAC without benign repeat is also better then UGFNAC with benign repeat, because it results in less false negatives. This will reduce costs but also improve patient's quality of life.

Furthermore, the committee was aware of the issues associated with late versus early detection of cancer (malignancy). They noted that earlier detection has a higher chance of survival compared to late detection or undetected cancers, which could mean a lost chance of treatment to the patient, increased risk of complications and mortality. Late detection will incur additional costs and reduce quality of life.

This supported a strong recommendation to offer UGFNAC for the diagnosis of malignancy in thyroid nodules. The committee noted that the results of the economic evidence and the original cost-analysis were in line with current practice and were not likely to have a substantial cost impact.

1.7.3 Other factors the committee took into account

The committee noted that while they would generally recommend ultrasound guidance for FNAC, there may be the occasional scenario in which clinical features are highly suggestive of malignancy and the potential delay in obtaining an ultrasound guided FNAC (as opposed to a palpation guided FNAC which could be done in the initial assessment appointment) may not be warranted as the key issue would be to begin management as soon as possible. However, they agreed that ideally an urgent UG FNAC would be available and avoid the need for PG FNAC at any point.

References

1. Aksu O, Koroglu BK, Ersoy IH, Koroglu M, Ciris M, Ersoy S et al. Thyroid fine needle aspiration biopsy: Which method should be preferred in an endemic goiter region? Acta Medica Mediterranea. 2014; 30(1):297-301

2. Al Maqbali T, Tedla M, Weickert MO, Mehanna H. Malignancy risk analysis in patients with inadequate fine needle aspiration cytology (FNAC) of the thyroid. PloS One. 2012; 7(11):e49078

3. Bohacek L, Milas M, Mitchell J, Siperstein A, Berber E. Diagnostic accuracy of surgeon-performed ultrasound-guided fine-needle aspiration of thyroid nodules. Annals of Surgical Oncology. 2012; 19(1):45-51

4. Borget I, Vielh P, Leboulleux S, Allyn M, Iacobelli S, Schlumberger M et al. Assessment of the cost of fine-needle aspiration cytology as a diagnostic tool in patients with thyroid nodules. American Journal of Clinical Pathology. 2008; 129(5):763-71

5. Braga M, Cavalcanti TC, Collaco LM, Graf H. Efficacy of ultrasound-guided fine-needle aspiration biopsy in the diagnosis of complex thyroid nodules. Journal of Clinical Endocrinology and Metabolism. 2001; 86(9):4089-91

6. Cai XJ, Valiyaparambath N, Nixon P, Waghorn A, Giles T, Helliwell T. Ultrasound-guided fine needle aspiration cytology in the diagnosis and management of thyroid nodules. Cytopathology. 2006; 17(5):251-6

7. Cam OH, Tekin M, Acar GO, Kilicaslan A. What is the Role of Diffusion Weigh Magnetic Resonance Imaging in Evaluation of Thyroid Nodules? Indian Journal of Otolaryngology & Head & Neck Surgery. 2014; 66(3):336-40

8. Can AS. Cost-effectiveness comparison between palpation- and ultrasound-guided thyroid fine-needle aspiration biopsies. BMC Endocrine Disorders. 2009; 9(14)

9. Can AS, Peker K. Comparison of palpation-versus ultrasound-guided fine-needle aspiration biopsies in the evaluation of thyroid nodules. BMC Research Notes. 2008; 1:12

10. Carmeci C, Jeffrey RB, McDougall IR, Nowels KW, Weigel RJ. Ultrasound-guided fine-needle aspiration biopsy of thyroid masses. Thyroid. 1998; 8(4):283-9

11. Cesur M, Corapcioglu D, Bulut S, Gursoy A, Yilmaz AE, Erdogan N et al. Comparison of palpation-guided fine-needle aspiration biopsy to ultrasound-guided fine-needle aspiration biopsy in the evaluation of thyroid nodules. Thyroid. 2006; 16(6):555-61

12. Danese D, Sciacchitano S, Farsetti A, Andreoli M, Pontecorvi A. Diagnostic accuracy of conventional versus sonography-guided fine-needle aspiration biopsy of thyroid nodules. Thyroid. 1998; 8(1):15-21

13. de Meer SG, Schreinemakers JM, Zelissen PM, Stapper G, Sie-Go DM, Rinkes IH et al. Fine-needle aspiration of thyroid tumors: identifying factors associated with adequacy rate in a large academic center in the Netherlands. Diagnostic Cytopathology. 2012; 40 (Suppl 1):E21-6

14. Deandrea M, Mormile A, Veglio M, Motta M, Pellerito R, Gallone G et al. Fine-needle aspiration biopsy of the thyroid: Comparison between thyroid palpation and ultrasonography. Endocrine Practice. 2002; 8(4):282-286

15. Esfahanian F, Aryan A, Ghajarzadeh M, Yazdi MH, Nobakht N, Burchi M. Application of sonoelastography in differential diagnosis of benign and malignant thyroid nodules. International Journal of Preventive Medicine. 2016; 7(1):55

16. Hatada T, Okada K, Ishii H, Ichii S, Utsunomiya J. Evaluation of ultrasound-guided fine-needle aspiration biopsy for thyroid nodules. American Journal of Surgery. 1998; 175(2):133-6

17. Izquierdo R, Arekat MR, Knudson PE, Kartun KF, Khurana K, Kort K et al. Comparison of palpation-guided versus ultrasound-guided fine-needle aspiration biopsies of thyroid nodules in an outpatient endocrinology practice. Endocrine Practice. 2006; 12(6):609-14

18. Jalan S, Sengupta S, Ray R, Mondal R, Phukan J, Bardhan J et al. A comparative evaluation of USG-guided FNAC with conventional FNAC in the preoperative assessment of thyroid lesions: A particular reference to cyto-histologically discordant cases. Bangladesh Journal of Medical Science. 2017; 16(2):274-281

19. Kawai T, Nishihara E, Kudo T, Ota H, Morita S, Kobayashi K et al. Histopathological diagnoses of "accessory" thyroid nodules diagnosed as benign by fine-needle aspiration cytology and ultrasonography. Thyroid. 2012; 22(3):299-303

20. Kimoto T, Suemitsu K, Eda I, Shimizu T, Ohtani M, Nabika T. The efficiency of performing ultrasound-guided fine-needle aspiration biopsy following mass screening for thyroid tumors to avoid unnecessary surgery. Surgery Today. 1999; 29(9):880-3

21. Krishnappa P, Ramakrishnappa S, Kulkarni MH. Comparison of free hand versus ultrasound-guided fine needle aspiration of thyroid with histopathological correlation. Journal of Environmental Pathology, Toxicology and Oncology. 2013; 32(2):149-55

22. Lee J, Lee SY, Cha SH, Cho BS, Kang MH, Lee OJ. Fine-needle aspiration of thyroid nodules with macrocalcification. Thyroid. 2013; 23(9):1106-12

23. Lee SW, Lee HJ, Kim HJ, Lee J, Park JY, Kim SH et al. Combined categorical reporting systems of US and cytology findings for thyroid nodules: guidance on repeat fine-needle aspiration cytology. Radiology. 2013; 266(3):956-63

24. Lee YH, Kim BH, Suh SI, Seo HS, Seo BK, Cho KR et al. Comparison of cytological results obtained by repeated US-guided fine-needle aspiration biopsies of thyroid nodules: focus on the rate of malignancy and diagnostic concordance. Diagnostic Cytopathology. 2009; 37(7):492-7

25. Leung VA, Kirpalani A, Mnatzakanian G, Colak E, Vlachou PA. Effect of a Biopsy Center on Adequacy Rates of Thyroid Nodule Fine-Needle Aspiration. AJR American Journal of Roentgenology. 2017; 209(2):358-362

26. Lew JI, Rodgers SE, Solorzano CC. Developments in the use of ultrasound for thyroid cancer. Current Opinion in Oncology. 2010; 22(1):11-6

27. Li C, Zhan W, Yi F, Zheng B, Zhou Y, Zhao R et al. Fine needle aspiration cytology guided by ultrasound in the diagnosis of subcentimetre thyroid nodules. Springerplus. 2016; 5:876

28. Lin JD, Huang BY, Weng HF, Jeng LB, Hsueh C. Thyroid ultrasonography with fine-needle aspiration cytology for the diagnosis of thyroid cancer. Journal of Clinical Ultrasound. 1997; 25(3):111-8

29. Mehrotra P, Viswanathan H, Johnson SJ, Wadehra V, Richardson DL, Lennard TW. Ultrasound guidance improves the adequacy of our preoperative thyroid cytology but not its accuracy. Cytopathology. 2006; 17(3):137-44

30. Melany M, Chen S. Thyroid Cancer: Ultrasound Imaging and Fine-Needle Aspiration Biopsy. Endocrinology and Metabolism Clinics of North America. 2017; 46(3):691-711

31. Mirshemirani A, Roshanzamir F, Tabari AK, Ghorobi J, Salehpoor S, Gorji FA. Thyroid nodules in childhood: a single institute experience. Iranian Journal of Pediatrics. 2010; 20(1):91-6

32. Mittendorf EA, Tamarkin SW, McHenry CR. The results of ultrasound-guided fine-needle aspiration biopsy for evaluation of nodular thyroid disease. Surgery. 2002; 132(4):648-53; discussion 653-4

33. Moon HG, Jung EJ, Park ST, Ha WS, Choi SK, Hong SC et al. Role of ultrasonography in predicting malignancy in patients with thyroid nodules. World Journal of Surgery. 2007; 31(7):1410-6

34. Muruganandham K, Sistla SC, Elangovan S, Verma SK. Routine ultrasound-guided aspiration cytology for evaluation of palpable thyroid nodules in an endemic area: is it justified? Journal of otolaryngology - head & neck surgery. 2009; 38(2):222-226

35. Nachiappan AC, Metwalli ZA, Hailey BS, Patel RA, Ostrowski ML, Wynne DM. The thyroid: review of imaging features and biopsy techniques with radiologic-pathologic correlation. Radiographics. 2014; 34(2):276-93

36. Nam-Goong IS, Kim HY, Gong G, Lee HK, Hong SJ, Kim WB et al. Ultrasonography-guided fine-needle aspiration of thyroid incidentaloma: correlation with pathological findings. Clinical Endocrinology. 2004; 60(1):21-8

37. National Institute for Health and Care Excellence. Developing NICE guidelines: the manual [updated October 2018]. London. National Institute for Health and Care Excellence, 2014. Available from: http://www.nice.org.uk/article/PMG20/chapter/1%20Introduction%20and%20overview

38. Newkirk KA, Ringel MD, Jelinek J, Mark A, Wartofsky L, Deeb ZE et al. Ultrasound-guided fine-needle aspiration and thyroid disease. Otolaryngology - Head & Neck Surgery. 2000; 123(6):700-5

39. Ogawa Y, Kato Y, Ikeda K, Aya M, Ogisawa K, Kitani K et al. The value of ultrasound-guided fine-needle aspiration cytology for thyroid nodules: an assessment of its diagnostic potential and pitfalls. Surgery Today. 2001; 31(2):97-101

40. Organisation for Economic Co-operation and Development (OECD). Purchasing power parities (PPP). 2012. Available from: http://www.oecd.org/std/ppp Last accessed: 15/02/2019

41. Peng L, Gu MJ. Diagnostic value of conventional and ultrasound-guided fine-needle aspiration biopsy for thyroid nodules: a meta-analysis. Academic Journal of Second Military Medical University. 2007; 28(9):968-972

42. Persichetti A, Di Stasio E, Guglielmi R, Bizzarri G, Taccogna S, Misischi I et al. Predictive Value of Malignancy of Thyroid Nodule Ultrasound Classification Systems: A Prospective Study. Journal of Clinical Endocrinology and Metabolism. 2018; 103(4):1359-1368

43. Rorive S, D'Haene N, Fossion C, Delpierre I, Abarguia N, Avni F et al. Ultrasound-guided fine-needle aspiration of thyroid nodules: stratification of malignancy risk using follicular proliferation grading, clinical and ultrasonographic features. European Journal of Endocrinology. 2010; 162(6):1107-15

44. Schwartz J, How J, Lega I, Cote J, Gologan O, Rivera JA et al. Ultrasound-guided fine-needle aspiration thyroid biopsies in the otolaryngology clinic. Journal of Otolaryngology - Head and Neck Surgery. 2010; 39(4):356-60

45. Singh Ospina N, Brito JP, Maraka S, Espinosa de Ycaza AE, Rodriguez-Gutierrez R, Gionfriddo MR et al. Diagnostic accuracy of ultrasound-guided fine needle aspiration biopsy for thyroid malignancy: systematic review and meta-analysis. Endocrine. 2016; 53(3):651-61

46. Takashima S, Fukuda H, Kobayashi T. Thyroid nodules: clinical effect of ultrasound-guided fine-needle aspiration biopsy. Journal of Clinical Ultrasound. 1994; 22(9):535-42

47. Witt RL, Sukumar VR, Gerges F. Surgeon-performed ultrasound-guided FNAC with on-site cytopathology improves adequacy and accuracy. Laryngoscope. 2015; 125(7):1633-6

48. Yang GC, Liebeskind D, Messina AV. Ultrasound-guided fine-needle aspiration of the thyroid assessed by Ultrafast Papanicolaou stain: data from 1135 biopsies with a two- to six-year follow-up. Thyroid. 2001; 11(6):581-9

49. Young JK, Lumapas CG, Mirasol R. Sonographically guided fine-needle aspiration biopsy of thyroid nodules: Correlation between cytologic and histopathologic findings. Phillippine Journal of Internal Medicine. 2011; 49(1):8-14

50. Zawawi F, Mosli MH, Zawawi ST. Should ultrasound-guided fine needle aspiration be considered a first-line technique in assessing a thyroid nodule? Otolaryngologia Polska. 2016; 70(1):49-53

Appendices

Appendix A: Review protocols

Table 8:

ID	Field	Content
I	Review question	Should a fine-needle aspiration biopsy (FNAB) be under ultrasound guidance?
II	Type of review question	Diagnostic A review of health economic evidence related to the same review question was conducted in parallel with this review. For details see the health economic review protocol for this NICE guideline.
III	Objective of the review	To determine the accuracy of FNAB with and without ultrasound
IV	Eligibility criteria – population / disease / condition / issue / domain	• People presenting with euthyroid thyroid enlargement with preliminary investigation suggesting need for biopsy
V	Eligibility criteria – intervention(s) / exposure(s) / prognostic factor(s)	• FNAB without ultrasound • FNAB with ultrasound
VI	Eligibility criteria – comparator(s) / control or reference (gold) standard	• Reference standard will be malignant status as confirmed by core biopsy or surgery/subsequent development of cancer in case of false negatives that are not further investigated
VII	Outcomes and prioritisation	• Sensitivity, specificity, PPV, NPV of tests for diagnosing thyroid cancer Sensitivity will be prioritised for decision making
VIII	Eligibility criteria – study design	• Diagnostic accuracy studies • Prospective studies prioritised, retrospective studies included if insufficient prospective studies identified • Evidence will be extracted according to the following hierarchy, lower levels will only be considered if insufficient evidence for decision making is found for higher levels: o Studies in which entire population gets FNAB without ultrasound and with ultrasound o Studies in which FNAB with ultrasound and without ultrasound are compared in the same setting o Studies in which only one of FNAB with ultrasound or without ultrasound is assessed
IX	Other inclusion exclusion criteria	• Excluding two gate study design

X	Proposed sensitivity / subgroup analysis, or meta-regression	None specified
XI	Selection process – duplicate screening / selection / analysis	A sample of at least 10% of the abstract lists were double-sifted by a senior research fellow and discrepancies rectified, with committee input where consensus could not be reached, for more information please see the separate Methods report for this guideline.
XII	Data management (software)	• Endnote was used for bibliography, citations, sifting and reference management • Pairwise meta-analyses were performed using Cochrane Review Manager (RevMan5). • GRADEpro was used to assess the quality of evidence for each outcome. • WinBUGS was used for meta-analysis of accuracy outcomes
XIII	Information sources – databases and dates	• Medline, Embase and the Cochrane library
XIV	Identify if an update	Not an update
XV	Author contacts	https://www.nice.org.uk/guidance/indevelopment/gid-ng10074
XVI	Highlight if amendment to previous protocol	Not an amendment
XVII	Search strategy – for one database	For details please see appendix B
XVIII	Data collection process – forms / duplicate	A standardised evidence table format will be used, and published as appendix D of the evidence report.
XIX	Data items – define all variables to be collected	For details please see evidence tables in Appendix D (clinical evidence tables) or G (health economic evidence tables).
XX	Methods for assessing bias at outcome / study level	QUADAS-2 checklists were used to critically appraise individual studies. For details please see section 6.2 of Developing NICE guidelines: the manual The risk of bias across all available evidence was evaluated for each outcome using an adaptation of the 'Grading of Recommendations Assessment, Development and Evaluation (GRADE) toolbox' developed by the international GRADE working group http://www.gradeworkinggroup.org/
XXI	Criteria for quantitative synthesis	For details please see section 6.4 of Developing NICE guidelines: the manual.
XXII	Methods for quantitative analysis – combining studies and exploring	For details please see the separate Methods report for this guideline.

	(in)consistency	
XXIII	Meta-bias assessment – publication bias, selective reporting bias	For details please see section 6.2 of Developing NICE guidelines: the manual.
XXIV	Confidence in cumulative evidence	For details please see sections 6.4 and 9.1 of Developing NICE guidelines: the manual.
XXV	Rationale / context – what is known	For details please see the introduction to the evidence review.
XXVI	Describe contributions of authors and guarantor	A multidisciplinary committee developed the evidence review. The committee was convened by the National Guideline Centre (NGC) and chaired by Sarah Fishburn line with section 3 of Developing NICE guidelines: the manual. Staff from NGC undertook systematic literature searches, appraised the evidence, conducted meta-analysis and cost-effectiveness analysis where appropriate, and drafted the evidence review in collaboration with the committee. For details please see Developing NICE guidelines: the manual.
XXVII	Sources of funding / support	NGC is funded by NICE and hosted by the Royal College of Physicians.
XXVIII	Name of sponsor	NGC is funded by NICE and hosted by the Royal College of Physicians.
XXIX	Roles of sponsor	NICE funds NGC to develop guidelines for those working in the NHS, public health and social care in England.
XXX	PROSPERO registration number	Not registered

Table 9: Health economic review protocol

Review question	All questions – health economic evidence
Objectives	To identify health economic studies relevant to any of the review questions.
Search criteria	• Populations, interventions and comparators must be as specified in the clinical review protocol above. • Studies must be of a relevant health economic study design (cost–utility analysis, cost-effectiveness analysis, cost–benefit analysis, cost–consequences analysis, comparative cost analysis). • Studies must not be a letter, editorial or commentary, or a review of health economic evaluations. (Recent reviews will be ordered although not reviewed. The bibliographies will be checked for relevant studies, which will then be ordered.) • Unpublished reports will not be considered unless submitted as part of a call for evidence. • Studies must be in English.
Search strategy	A health economic study search will be undertaken using population-specific terms and a health economic study filter – see Appendix B: below.
Review strategy	Studies not meeting any of the search criteria above will be excluded. Studies published before 2003, abstract-only studies and studies from non-OECD countries or the USA will also be excluded. Each remaining study will be assessed for applicability and methodological limitations using the NICE economic evaluation checklist which can be found in appendix H of Developing NICE guidelines: the manual (2014).[37] **Inclusion and exclusion criteria** • If a study is rated as both 'Directly applicable' and with 'Minor limitations' then it will be included in the guideline. A health economic evidence table will be completed and it will be included in the health economic evidence profile. • If a study is rated as either 'Not applicable' or with 'Very serious limitations' then it will usually be excluded from the guideline. If it is excluded then a health economic evidence table will not be completed and it will not be included in the health economic evidence profile. • If a study is rated as 'Partially applicable', with 'Potentially serious limitations' or both then there is discretion over whether it should be included. **Where there is discretion** The health economist will make a decision based on the relative applicability and quality of the available evidence for that question, in discussion with the guideline committee if required. The ultimate aim is to include health economic studies that are helpful for decision-making in the context of the guideline and the current NHS setting. If several studies are considered of sufficiently high applicability and methodological quality that they could all be included, then the health economist, in discussion with the committee if required, may decide to include only the most applicable studies and to selectively exclude the remaining studies. All studies excluded on the basis of applicability or methodological limitations will be listed with explanation in the excluded health economic studies appendix below. The health economist will be guided by the following hierarchies. *Setting:* • UK NHS (most applicable). • OECD countries with predominantly public health insurance systems (for example, France, Germany, Sweden). • OECD countries with predominantly private health insurance systems (for example, Switzerland).

	• Studies set in non-OECD countries or in the USA will be excluded before being assessed for applicability and methodological limitations. *Health economic study type:* • Cost–utility analysis (most applicable). • Other type of full economic evaluation (cost–benefit analysis, cost-effectiveness analysis, cost–consequences analysis). • Comparative cost analysis. • Non-comparative cost analyses including cost-of-illness studies will be excluded before being assessed for applicability and methodological limitations. *Year of analysis:* • The more recent the study, the more applicable it will be. • Studies published in 2003 or later but that depend on unit costs and resource data entirely or predominantly from before 2003 will be rated as 'Not applicable'. • Studies published before 2003 will be excluded before being assessed for applicability and methodological limitations. *Quality and relevance of effectiveness data used in the health economic analysis:* • The more closely the clinical effectiveness data used in the health economic analysis match with the outcomes of the studies included in the clinical review the more useful the analysis will be for decision-making in the guideline.

Appendix B: Literature search strategies

The literature searches for this review are detailed below and complied with the methodology outlined in Developing NICE guidelines: the manual 2014, updated 2018 https://www.nice.org.uk/guidance/pmg20/resources/developing-nice-guidelines-the-manual-pdf-72286708700869

For more detailed information, please see the Methodology Review.

B.1 Clinical search literature search strategy

Searches were constructed using a PICO framework where population (P) terms were combined with Intervention (I) and in some cases Comparison (C) terms. Outcomes (O) are rarely used in search strategies for interventions as these concepts may not be well described in title, abstract or indexes and therefore difficult to retrieve. Search filters were applied to the search where appropriate.

Table 10: Database date parameters and filters used

Database	Dates searched	Search filter used
Medline (OVID)	1946 – 07 January 2019	Exclusions Randomised controlled trials Observational studies Diagnostic tests studies
Embase (OVID)	1974 – 07 January 2019	Exclusions Randomised controlled trials Observational studies Diagnostic tests studies
The Cochrane Library (Wiley)	Cochrane Reviews to 2019 Issue 1 or 12 CENTRAL to 2019 Issue 1 or 12 DARE, and NHSEED to 2015 Issue 2 of 4 HTA to 2016 Issue 2 of 4	None

Medline (Ovid) search terms

1.	exp thyroid diseases/
2.	hyperthyroid*.ti,ab.
3.	hypothyroid*.ti,ab.
4.	thyrotoxicosis.ti,ab.
5.	(thyroid adj3 (swell* or dysfunction* or enlarg* or nodule* or node* or disease* or condition* or disorder*)).ti,ab.
6.	or/1-5
7.	letter/
8.	editorial/
9.	news/
10.	exp historical article/
11.	Anecdotes as Topic/
12.	comment/
13.	case report/
14.	(letter or comment*).ti.

15.	or/7-14
16.	randomized controlled trial/ or random*.ti,ab.
17.	15 not 16
18.	animals/ not humans/
19.	exp Animals, Laboratory/
20.	exp Animal Experimentation/
21.	exp Models, Animal/
22.	exp Rodentia/
23.	(rat or rats or mouse or mice).ti.
24.	or/17-23
25.	6 not 24
26.	Biopsy, Fine-Needle/
27.	(FNA or FNAB or FNA biops* or fine needle aspiration or fine needle aspiration biops* or fine-needle aspiration or fine-needle aspiration biops* or (palpation guid* adj3 aspiration)).ti,ab.
28.	26 or 27
29.	25 and 28
30.	randomized controlled trial.pt.
31.	controlled clinical trial.pt.
32.	randomi#ed.ti,ab.
33.	placebo.ab.
34.	randomly.ti,ab.
35.	Clinical Trials as topic.sh.
36.	trial.ti.
37.	or/30-36
38.	exp "sensitivity and specificity"/
39.	(sensitivity or specificity).ti,ab.
40.	((pre test or pretest or post test) adj probability).ti,ab.
41.	(predictive value* or PPV or NPV).ti,ab.
42.	likelihood ratio*.ti,ab.
43.	likelihood function/
44.	((area under adj4 curve) or AUC).ti,ab.
45.	(receive* operat* characteristic* or receive* operat* curve* or ROC curve*).ti,ab.
46.	(diagnos* adj3 (performance* or accurac* or utilit* or value* or efficien* or effectiveness)).ti,ab.
47.	gold standard.ab.
48.	or/38-47
49.	Epidemiologic studies/
50.	Observational study/
51.	exp Cohort studies/
52.	(cohort adj (study or studies or analys* or data)).ti,ab.
53.	((follow up or observational or uncontrolled or non randomi#ed or epidemiologic*) adj (study or studies or data)).ti,ab.
54.	((longitudinal or retrospective or prospective or cross sectional) and (study or studies or review or analys* or cohort* or data)).ti,ab.
55.	Controlled Before-After Studies/

56.	Historically Controlled Study/
57.	Interrupted Time Series Analysis/
58.	(before adj2 after adj2 (study or studies or data)).ti,ab.
59.	or/49-58
60.	exp case control study/
61.	case control*.ti,ab.
62.	or/60-61
63.	59 or 62
64.	Cross-sectional studies/
65.	(cross sectional and (study or studies or review or analys* or cohort* or data)).ti,ab.
66.	or/64-65
67.	59 or 66
68.	59 or 62 or 66
69.	29 and (37 or 48 or 68)
70.	limit 69 to English language

Embase (Ovid) search terms

1.	exp thyroid diseases/
2.	hyperthyroid*.ti,ab.
3.	hypothyroid*.ti,ab.
4.	thyrotoxicosis.ti,ab.
5.	(thyroid adj3 (swell* or dysfunction* or enlarg* or nodule* or node* or disease* or condition* or disorder*)).ti,ab.
6.	or/1-5
7.	letter.pt. or letter/
8.	note.pt.
9.	editorial.pt.
10.	case report/ or case study/
11.	(letter or comment*).ti.
12.	or/7-11
13.	randomized controlled trial/ or random*.ti,ab.
14.	12 not 13
15.	animal/ not human/
16.	nonhuman/
17.	exp Animal Experiment/
18.	exp Experimental Animal/
19.	animal model/
20.	exp Rodent/
21.	(rat or rats or mouse or mice).ti.
22.	or/14-21
23.	6 not 22
24.	fine needle aspiration biopsy/
25.	(FNA or FNAB or FNA biops* or fine needle aspiration or fine needle aspiration biops* or fine-needle aspiration or fine-needle aspiration biops* or (palpation guid* adj3 aspiration)).ti,ab.
26.	24 or 25

27.	23 and 26
28.	random*.ti,ab.
29.	factorial*.ti,ab.
30.	(crossover* or cross over*).ti,ab.
31.	((doubl* or singl*) adj blind*).ti,ab.
32.	(assign* or allocat* or volunteer* or placebo*).ti,ab.
33.	crossover procedure/
34.	single blind procedure/
35.	randomized controlled trial/
36.	double blind procedure/
37.	or/28-36
38.	exp "sensitivity and specificity"/
39.	(sensitivity or specificity).ti,ab.
40.	((pre test or pretest or post test) adj probability).ti,ab.
41.	(predictive value* or PPV or NPV).ti,ab.
42.	likelihood ratio*.ti,ab.
43.	((area under adj4 curve) or AUC).ti,ab.
44.	(receive* operat* characteristic* or receive* operat* curve* or ROC curve*).ti,ab.
45.	(diagnos* adj3 (performance* or accurac* or utilit* or value* or efficien* or effectiveness)).ti,ab.
46.	diagnostic accuracy/
47.	diagnostic test accuracy study/
48.	gold standard.ab.
49.	or/38-48
50.	Clinical study/
51.	Observational study/
52.	family study/
53.	longitudinal study/
54.	retrospective study/
55.	prospective study/
56.	cohort analysis/
57.	follow-up/
58.	cohort*.ti,ab.
59.	57 and 58
60.	(cohort adj (study or studies or analys* or data)).ti,ab.
61.	((follow up or observational or uncontrolled or non randomi#ed or epidemiologic*) adj (study or studies or data)).ti,ab.
62.	((longitudinal or retrospective or prospective or cross sectional) and (study or studies or review or analys* or cohort* or data)).ti,ab.
63.	(before adj2 after adj2 (study or studies or data)).ti,ab.
64.	or/50-56,59-63
65.	exp case control study/
66.	case control*.ti,ab.
67.	or/65-66
68.	64 or 67
69.	cross-sectional study/

70.	(cross sectional and (study or studies or review or analys* or cohort* or data)).ti,ab.
71.	or/69-70
72.	64 or 71
73.	64 or 67 or 71
74.	27 and (37 or 49 or 73)
75.	limit 74 to English language

Cochrane Library (Wiley) search terms

#1.	MeSH descriptor: [Thyroid Diseases] explode all trees
#2.	hyperthyroid*:ti,ab
#3.	hypothyroid*:ti,ab
#4.	thyrotoxicosis:ti,ab
#5.	(thyroid near/3 (swell* or dysfunction* or enlarg* or nodule* or node* or disease* or condition* or disorder*)) ti,ab
#6.	(or #1-#5)
#7.	MeSH descriptor: [Biopsy, Fine-Needle] explode all trees
#8.	(FNA or FNAB or FNA biops* or fine needle aspiration or fine needle aspiration biops* or fine-needle aspiration or fine-needle aspiration biops* or (palpation guid* near/3 aspiration)):ti,ab
#9.	#7 or #8
#10.	#6 and #9

B.2 Health Economics literature search strategy

Health economic evidence was identified by conducting a broad search relating to a thyroid disease population in NHS Economic Evaluation Database (NHS EED – this ceased to be updated after March 2015) and the Health Technology Assessment database (HTA) with no date restrictions. NHS EED and HTA databases are hosted by the Centre for Research and Dissemination (CRD). Additional searches were run on Medline and Embase for health economics, economic modelling and quality of life studies.

Table 11: Database date parameters and filters used

Database	Dates searched	Search filter used
Medline	2014 – 07 January 2019	Exclusions Health economics studies Health economics modelling studies Quality of life studies
Embase	2014 – 07 January 2019	Exclusions Health economics studies Health economics modelling studies Quality of life studies
Centre for Research and Dissemination (CRD)	HTA - Inception – 07 January 2019 NHSEED - Inception to March 2015	None

Medline (Ovid) search terms

1.	exp thyroid diseases/
2.	hyperthyroid*.ti,ab.

3.	hypothyroid*.ti,ab.
4.	thyrotoxicosis.ti,ab.
5.	(thyroid adj3 (swell* or dysfunction* or enlarg* or nodule* or node* or disease* or condition* or disorder*)).ti,ab.
6.	or/1-5
7.	letter/
8.	editorial/
9.	news/
10.	exp historical article/
11.	Anecdotes as Topic/
12.	comment/
13.	case report/
14.	(letter or comment*).ti.
15.	or/7-14
16.	randomized controlled trial/ or random*.ti,ab.
17.	15 not 16
18.	animals/ not humans/
19.	exp Animals, Laboratory/
20.	exp Animal Experimentation/
21.	exp Models, Animal/
22.	exp Rodentia/
23.	(rat or rats or mouse or mice).ti.
24.	or/17-23
25.	6 not 24
26.	limit 25 to English language
27.	Economics/
28.	Value of life/
29.	exp "Costs and Cost Analysis"/
30.	exp Economics, Hospital/
31.	exp Economics, Medical/
32.	Economics, Nursing/
33.	Economics, Pharmaceutical/
34.	exp "Fees and Charges"/
35.	exp Budgets/
36.	budget*.ti,ab.
37.	cost*.ti.
38.	(economic* or pharmaco?economic*).ti.
39.	(price* or pricing*).ti,ab.
40.	(cost* adj2 (effective* or utilit* or benefit* or minimi* or unit* or estimat* or variable*)).ab.
41.	(financ* or fee or fees).ti,ab.
42.	(value adj2 (money or monetary)).ti,ab.
43.	or/27-42
44.	exp models, economic/
45.	*Models, Theoretical/

46.	*Models, Organizational/
47.	markov chains/
48.	monte carlo method/
49.	exp Decision Theory/
50.	(markov* or monte carlo).ti,ab.
51.	econom* model*.ti,ab.
52.	(decision* adj2 (tree* or analy* or model*)).ti,ab.
53.	or/44-52
54.	quality-adjusted life years/
55.	sickness impact profile/
56.	(quality adj2 (wellbeing or well being)).ti,ab.
57.	sickness impact profile.ti,ab.
58.	disability adjusted life.ti,ab.
59.	(qal* or qtime* or qwb* or daly*).ti,ab.
60.	(euroqol* or eq5d* or eq 5*).ti,ab.
61.	(qol* or hql* or hqol* or h qol* or hrqol* or hr qol*).ti,ab.
62.	(health utility* or utility score* or disutilit* or utility value*).ti,ab.
63.	(hui or hui1 or hui2 or hui3).ti,ab.
64.	(health* year* equivalent* or hye or hyes).ti,ab.
65.	discrete choice*.ti,ab.
66.	rosser.ti,ab.
67.	(willingness to pay or time tradeoff or time trade off or tto or standard gamble*).ti,ab.
68.	(sf36* or sf 36* or short form 36* or shortform 36* or shortform36*).ti,ab.
69.	(sf20 or sf 20 or short form 20 or shortform 20 or shortform20).ti,ab.
70.	(sf12* or sf 12* or short form 12* or shortform 12* or shortform12*).ti,ab.
71.	(sf8* or sf 8* or short form 8* or shortform 8* or shortform8*).ti,ab.
72.	(sf6* or sf 6* or short form 6* or shortform 6* or shortform6*).ti,ab.
73.	or/54-72
74.	26 and (43 or 53 or 73)

Embase (Ovid) search terms

1.	exp thyroid diseases/
2.	hyperthyroid*.ti,ab.
3.	hypothyroid*.ti,ab.
4.	thyrotoxicosis*.ti,ab.
5.	(thyroid adj3 (swell* or dysfunction* or enlarg* or nodule* or node* or disease* or condition* or disorder*)).ti,ab.
6.	or/1-5
7.	letter.pt. or letter/
8.	note.pt.
9.	editorial.pt.
10.	case report/ or case study/
11.	(letter or comment*).ti.
12.	or/7-11

13.	randomized controlled trial/ or random*.ti,ab.
14.	12 not 13
15.	animal/ not human/
16.	nonhuman/
17.	exp Animal Experiment/
18.	exp Experimental Animal/
19.	animal model/
20.	exp Rodent/
21.	(rat or rats or mouse or mice).ti.
22.	or/14-21
23.	6 not 22
24.	limit 23 to English language
25.	health economics/
26.	exp economic evaluation/
27.	exp health care cost/
28.	exp fee/
29.	budget/
30.	funding/
31.	budget*.ti,ab.
32.	cost*.ti.
33.	(economic* or pharmaco?economic*).ti.
34.	(price* or pricing*).ti,ab.
35.	(cost* adj2 (effective* or utilit* or benefit* or minimi* or unit* or estimat* or variable*)).ab.
36.	(financ* or fee or fees).ti,ab.
37.	(value adj2 (money or monetary)).ti,ab.
38.	or/25-37
39.	statistical model/
40.	exp economic aspect/
41.	39 and 40
42.	*theoretical model/
43.	*nonbiological model/
44.	stochastic model/
45.	decision theory/
46.	decision tree/
47.	monte carlo method/
48.	(markov* or monte carlo).ti,ab.
49.	econom* model*.ti,ab.
50.	(decision* adj2 (tree* or analy* or model*)).ti,ab.
51.	or/41-50
52.	quality adjusted life year/

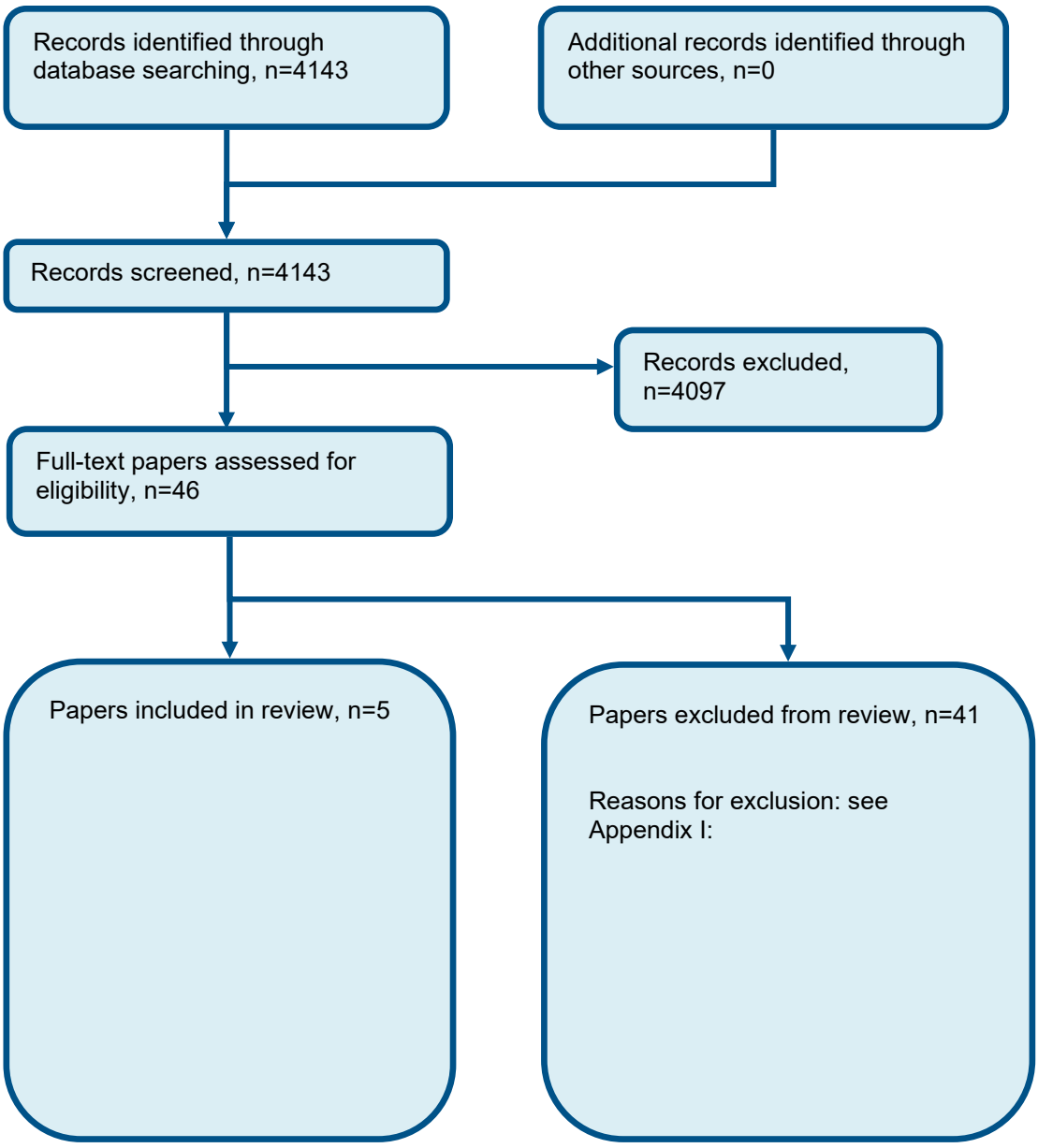
53.	"quality of life index"/
54.	short form 12/ or short form 20/ or short form 36/ or short form 8/
55.	sickness impact profile/
56.	(quality adj2 (wellbeing or well being)).ti,ab.
57.	sickness impact profile.ti,ab.
58.	disability adjusted life.ti,ab.
59.	(qal* or qtime* or qwb* or daly*).ti,ab.
60.	(euroqol* or eq5d* or eq 5*).ti,ab.
61.	(qol* or hql* or hqol* or h qol* or hrqol* or hr qol*).ti,ab.
62.	(health utility* or utility score* or disutilit* or utility value*).ti,ab.
63.	(hui or hui1 or hui2 or hui3).ti,ab.
64.	(health* year* equivalent* or hye or hyes).ti,ab.
65.	discrete choice*.ti,ab.
66.	rosser.ti,ab.
67.	(willingness to pay or time tradeoff or time trade off or tto or standard gamble*).ti,ab.
68.	(sf36* or sf 36* or short form 36* or shortform 36* or shortform36*).ti,ab.
69.	(sf20 or sf 20 or short form 20 or shortform 20 or shortform20).ti,ab.
70.	(sf12* or sf 12* or short form 12* or shortform 12* or shortform12*).ti,ab.
71.	(sf8* or sf 8* or short form 8* or shortform 8* or shortform8*).ti,ab.
72.	(sf6* or sf 6* or short form 6* or shortform 6* or shortform6*).ti,ab.
73.	or/52-72
74.	24 and (38 or 51 or 73)

NHS EED and HTA (CRD) search terms

#1.	MeSH DESCRIPTOR Thyroid Diseases EXPLODE ALL TREES
#2.	hyperthyroid*
#3.	hypothyroid*
#4.	thyrotoxicosis*
#5.	(thyroid adj3 (swell* or dysfunction* or enlarg* or nodule* or node* or disease* or condition* or disorder*))
#6.	#1 OR #2 OR #3 OR #4 or #5

Appendix C: Clinical evidence selection

Figure 2: Flow chart of clinical study selection for the review of FNAB with or without ultrasound



Appendix D: Clinical evidence tables

Reference	**Cesur 2006** [11]
Study type	Prospective study
Study methodology	Data source: prospective recruitment of patients with single or multiple nodular goiter admitted to outpatient thyroid clinic Recruitment: consecutive patients meeting inclusion criteria
Number of patients	n = 215
Patient characteristics	Age, mean (SD): 48.7 (13.5) years Gender (male to female ratio): 36:179 Ethnicity: not reported Setting: Imaging and Intervention Laboratory of Endocrinology and Metabolic Diseases Department, Ankara University medical School Country: Turkey Inclusion criteria: palpable nodules with maximal diameter between 1 and 2.5 cm Exclusion criteria: having more than four nodules, nodules <1 cm or >2.5 cm, hot nodules as determined by scintigraphic studies or in close proximity to a large nodule
Target condition(s)	Thyroid nodules
Index test(s) and reference standard	Index tests PGFBAB: patients put in supine position, neck extended backwards, skin preparation by 70% alcohol, local anaesthetic not used; while fixating nodule with two fingers of one hand, a 23-gauge (0.6 mm) needle attached to 10-mL syringe was introduced with other hand, aspiration was performed with a drive in one direction, nodule was aspirated with suction and needle was withdrawn. UGFNAC: performed in the same session, by the same operator, 5 minutes after PGFNAC, without patients getting up from the examination table. After determining the location of the nodule by ultrasound, the 'abdominal approach': in which the operator is positioned in the right side of the patient close to the abdomen was used. After ultrasound gel removal and skin preparation with antiseptic solution,

<table>
<tr><td>Reference</td><td colspan="5">Cesur 2006 [11]</td></tr>
<tr><td></td><td colspan="5">transducer was placed over the nodule in sagittal position in one hand, needle introduced with other hand, transducer released and aspiration carried out.

<u>Reference standard</u>
Surgery was performed in fifteen of 215 patients or 26 of 285 nodules because of suspicious or malignant cytology results in study included (n=9 patients) or study excluded nodules (n=2 patients) or due to their own decision (n=2 patients). Two cases where the reason for surgery was primary hyperparathyroidism were excluded.

Time between measurement of index tests: 5 minutes</td></tr>
<tr><td rowspan="4">2×2 table</td><td>UGFNAC</td><td>Reference standard +</td><td>Reference standard −</td><td>Total</td><td rowspan="4">Notes: FP & FN results only available in % for 26 surgical nodules in 13 patients</td></tr>
<tr><td>Index test +</td><td>6</td><td>2</td><td>8</td></tr>
<tr><td>Index test −</td><td>1</td><td>17</td><td>18</td></tr>
<tr><td>Total</td><td>7</td><td>19</td><td>26</td></tr>
<tr><td rowspan="4">2×2 table</td><td>PGFNAC</td><td>Reference standard +</td><td>Reference standard −</td><td>Total</td><td rowspan="4">Notes: FP & FN results only available in % for 26 surgical nodules in 13 patients</td></tr>
<tr><td>Index test +</td><td>4</td><td>3</td><td>7</td></tr>
<tr><td>Index test −</td><td>3</td><td>16</td><td>19</td></tr>
<tr><td>Total</td><td>7</td><td>19</td><td>26</td></tr>
<tr><td>Statistical measures</td><td colspan="5"><u>Index text UGFNAC</u>
Sensitivity : 85.7%
Specificity: 89.5%
PPV: 75%
NPV: 94.4%
Total number of people with inadequate sample: 61 (21.4%)

<u>Index text PGFNAC</u>
Sensitivity : 57.1%
Specificity 84.2%
PPV: 57.1%
NPV: 84.2%
Total number of people with inadequate sample: 92 (32.3%)

<u>Overall</u>
Total number of nodules with positive result: 7
Total number of nodules with negative result: 19</td></tr>
</table>

Reference	**Cesur 2006** [11]
Source of funding	Not reported
Limitations	Risk of bias: serious; high risk of bias in flow and timing Indirectness: none
Comments	

Reference	**Jalan 2017** [18]
Study type	Case series
Study methodology	Data source: all patients presenting with complains of thyroid swelling Recruitment: consecutive; all patients meeting criteria between 2011 and 2013
Number of patients	n = 84; both index tests: n=36.
Patient characteristics	Age, mean (SD): (range 8-71; majority 21-40) Gender (male to female ratio): 14:70 Ethnicity: not reported Setting: department of Pathology, BSMCH Country: India Inclusion criteria: patients with complaints of thyroid swelling at the department of Pathology between 2011 and 2013 Exclusion criteria: no age and sex criteria utilised to select cases
Target condition(s)	Thyroid lesions
Index test(s) and reference standard	Index tests PGFNAC: FNA was done using 25-gauge needle fitted to 10 ml syringe with patient in supine or sitting posture with neck extended; no aspiration technique was followed

Reference	**Jalan 2017** [18]				
	UGFNAC: FNA was repeated under ultrasound guidance. A minimum of four slides were smeared for each aspirate. Smears with at least six clusters of follicular cells, with at least 10 follicular cells each, were considered adequate for reporting. Papanicolaou (Pap) and May-Grunwald-Giemsa (MGG) staining were used. Reference standard: Histopathology: 40 patients, 18 from PGFNAC and 22 from UGFNAC group underwent surgery Time between measurement of index test and reference standard: not specified; index tests were conducted				
2×2 table	**UGFNAC**	Reference standard +	Reference standard −	Total	N= 36
	Index test +	8	1	9	Histological findings: 13 non-neoplastic lesions 9 neoplastic lesions
	Index test −	0	13	13	
	Total	8	14	22	
2×2 table	**PGFNAC**	Reference standard +	Reference standard −	Total	N=48
	Index test +	5	1	7	Histological findings: 11 non-neoplastic lesions 7 neoplastic lesions
	Index test −	2	10	11	
	Total	7	11	18	
Statistical measures	Index text UGFNAC Sensitivity : 100% Specificity: 92.31% Number of inadequate smears: 1 Index text PGFNAC Sensitivity : 71.43% Specificity: 90.91% Number of inadequate smears: 5				
Source of funding	Not reported				
Limitations	Risk of bias: serious; high risk of bias in flow and timing, unclear risk of bias in patient selection Indirectness: none				
Comments					

Reference	**Krishnappa 2013** [21]
Study type	Prospective
Study methodology	Data source: not specified; patients with thyroid lesions (96.8% presenting with swelling on the front of the neck) Recruitment: unclear
Number of patients	n = 91
Patient characteristics	Age, mean (range) : 38.5 (8-80) Gender (male to female ratio): 16:75 Ethnicity: not reported Setting: Department of pathology, Karnataka Institute of Medical Sciences Country: India Inclusion criteria: cases with thyroid lesions Exclusion criteria: not specified
Target condition(s)	Thyroid nodules
Index test(s) and reference standard	Index tests PGFNAC Several smears made for each case, some stained using routine method, others air dried and stained with Wright's stain. When obtained, fluid was aspirated using a syringe attached to the aspiration needle, examined macroscopically and then centrifuged. UGFNAC PGFNAC process was repeated under ultrasound guidance Reference standard: Histopathology: 25 patients underwent surgery, including subtotal thyroidectomy, lobectomy and isthmectomy. Removed specimens were examined histopathologically.

Reference	**Krishnappa 2013** [21]				
	Time between measurement of index test and reference standard: not specified				
2×2 table	**UGFNAC**	Reference standard +	Reference standard −	Total	UGFNAC (n=91): 68 cases with negative result (nonneoplastic), 21 cases with positive result, 2 unsatisfactory aspirates Surgery (n=10 positive/neoplastic cases): 9 with positive result, 1 with negative (nodular goiter)
	Index test +	9	1	10	
	Index test −	2	13	15	
	Total	11	14	25	
2×2 table	**PGFNAC**	Reference standard +	Reference standard −	Total	PGFNA (n=91): 67 cases with negative results,18 cases with positive, 6 unsatisfactory aspirates Surgery (n=7 positive cases): 6 with positive result, 1 with negative (nodular goiter)
	Index test +	6	1	7	
	Index test −	5	13	18	
	Total	11	14	25	
Statistical measures	Index text UGFNAC Sensitivity : 81.81% Specificity : 92.85% PPV: 90% NPV: 86.66% Number of people with inadequate sample: 2 (2.2%) Index text PGFNAC Sensitivity : 54.54% Specificity: 92.85% PPV: 85.71% NPV: 86.66% Number of people with inadequate sample: 6 (10.9%) FN, TN estimated				
Source of funding	Not specified				
Limitations	Risk of bias: serious; high risk of bias in patient selection, flow and timing Indirectness: none				
Comments	Statistical values (Sensitivity, Specificity, PPV, NPV, FN, FP) calculated for neoplastic lesions				

Reference	**Takashima 1994** [46]
Study type	Unclear but most likely prospective
Study methodology	Data source: patients referred to radiology department to confirm histopathologic diagnosis between 1989 and 1992 by other departments of Osaka University Hospital Recruitment: consecutive
Number of patients	n = 210
Patient characteristics	Age, mean (range): 53 (12-88) Gender (male to female ratio): 30:180 Ethnicity: not specified Setting: Department of Radiology, Osaka University Hospital Country: Japan Inclusion criteria: Exclusion criteria: cystic lesions less than 0.5 cm in diameter
Target condition(s)	Thyroid nodules
Index test(s) and reference standard	Index tests PGFNAC: performed in 62 nodules (57 patients) UGFNAC; performed in 268 nodules (all 210 patients) with a 22-gauge needle with a 5-MHz linear-array probe in a free-hand fashion. Smears were stained with both Papanicolaou and May-Giemsa methods. Nodules were classified as: 1. malignant, 2. suspicious, 3.cellular atypia, benign or insufficient material. Lesions in the first three categories were considered malignant. Reference standard: Histopathology: histopathologic confirmation following surgical removal was obtained for 133 nodules (99 aspirated with ultrasound

Reference	**Takashima 1994** [46]				
	guidance, 34 aspirated with palpation guidance) Time between measurement of index test and reference standard: not specified				
2×2 table	**UGFNAC**	Reference standard +	Reference standard −	Total	UGFNAC: 268 nodules (73 patients) Histopathology: obtained for 67 nodules (59 patients) with positive results (malignant), 32 nodules (14 patients) with negative result (benign)
	Index test +	64	3	67	
	Index test −	3	29	32	
	Total	67	32	99	
2×2 table	**PGFNAC**	Reference standard +	Reference standard −	Total	PGFNAC: 62 nodules (57 patients) Histopathology: obtained for 34/62 nodules (30 patients), 23 with positive results, 11 with negative results
	Index test +	21	1	22	
	Index test −	3	9	12	
	Total	24	10	34	
Statistical measures	Index text UGFNAC Sensitivity : 96% Specificity: 91% PPV 96% NPV 91% Inadequate sample: 10 nodules (3.7%) Index text PGFNAC Sensitivity : 88% Specificity: 90% PPV 95% NPV 75% Inadequate sample: 12 nodules (19%)				
Source of funding	Not specified				
Limitations	Risk of bias: serious; high risk of bias in flow and timing Indirectness: serious; high concern for patient selection (34% thyroid disease)				
Comments					

Reference	**Zawawi 2016** [50]				
Study type	Retrospective cohort				
Study methodology	Data source: retrospective chart review of patients undergoing thyroidectomies in tertiary health care facility Recruitment: consecutive				
Number of patients	n = 150				
Patient characteristics	Age, mean: 41.6 Gender (male to female ratio): 32:118 Ethnicity: not specified Setting: tertiary health care facility Country: Saudi Arabia Inclusion criteria: patients undergoing thyroidectomies at tertiary health care facility Exclusion criteria: not specified				
Target condition(s)	Thyroid nodules				
Index test(s) and reference standard	Index tests: PGFNAC: 151 aspirations performed, details not specified UGFNAC: 77 aspirations performed, details not specified Reference standard: Histopathology: thyroidectomy Time between measurement of index test and reference standard: not specified				
2×2 table	**UGFNAC**	Reference standard +	Reference standard −	Total	UGFNA: n=77, 22 positive result, 7 negative result Histopathology: number of people for who
	Index test +	15	6	21	
	Index test −	4	18	22	
	Total	19	24	43	

Reference	**Zawawi 2016** [50]				
					results available not specified
2×2 table	**PGFNAC**	Reference standard +	Reference standard −	Total	PGFNA: n=151, 31 negative result, 9 positive result Histopathology: number of people for who results available not specified
	Index test +	17	23	40	
	Index test −	7	24	31	
	Total	24	47	71	
Statistical measures	Index text UGFNAC Sensitivity : 78.9% Specificity: 75% PPV: 71.4% NPV: 81.8% Inadequate sample: 8 cytologies Index text PGFNAC Sensitivity : 70.8% Specificity: 51% PPV: 42.5% NPV: 77.4% Inadequate sample: 26 cytologies TP, FP, TN, FN estimated from SN, SP, PPV, NPV				
Source of funding	Not specified				
Limitations	Risk of bias: serious; high risk of bias in flow and timing, unclear risk of bias in index test Indirectness: none				
Comments	Index tests potentially conducted on different people				

Appendix E: Coupled sensitivity and specificity forest plots and sROC curves

E.1 Coupled sensitivity and specificity forest plots

Figure 3: UGFNAC

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Cesur 2006	6	2	1	17	0.86 [0.42, 1.00]	0.89 [0.67, 0.99]
Jalan 2017	8	1	0	13	1.00 [0.63, 1.00]	0.93 [0.66, 1.00]
Krishnappa 2013	9	1	2	13	0.82 [0.48, 0.98]	0.93 [0.66, 1.00]
Takashima 1994	64	3	3	13	0.96 [0.87, 0.99]	0.81 [0.54, 0.96]
Zawawi 2016	15	6	4	18	0.79 [0.54, 0.94]	0.75 [0.53, 0.90]

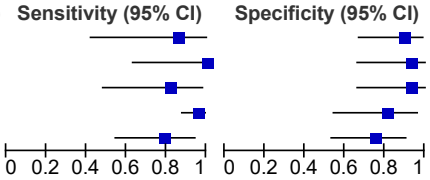
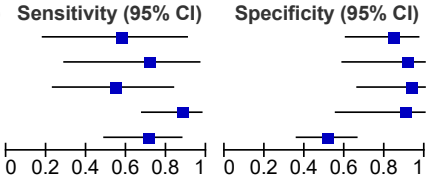


Figure 4: PGFNAC

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Cesur 2006	4	3	3	16	0.57 [0.18, 0.90]	0.84 [0.60, 0.97]
Jalan 2017	5	1	2	10	0.71 [0.29, 0.96]	0.91 [0.59, 1.00]
Krishnappa 2013	6	1	5	13	0.55 [0.23, 0.83]	0.93 [0.66, 1.00]
Takashima 1994	21	1	3	9	0.88 [0.68, 0.97]	0.90 [0.55, 1.00]
Zawawi 2016	17	23	7	24	0.71 [0.49, 0.87]	0.51 [0.36, 0.66]

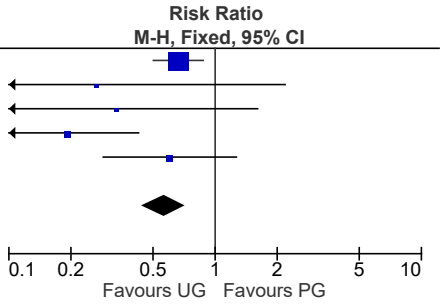


E.2 Inadequate results

Figure 5: Inadequate results

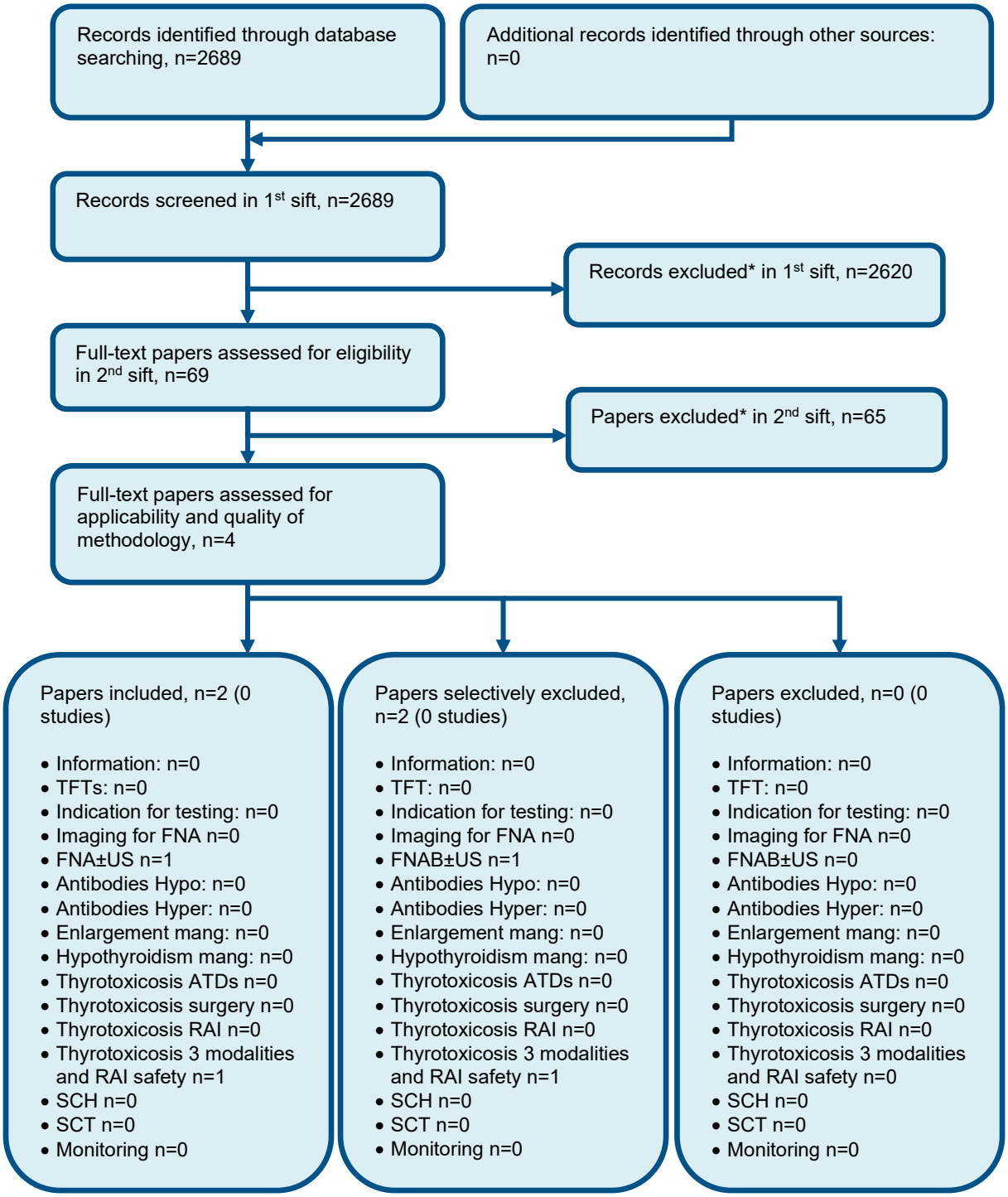
Study or Subgroup	UG Events	UG Total	PG Events	PG Total	Weight	Risk Ratio M-H, Fixed, 95% CI
Cesur 2006	61	285	92	285	66.0%	0.66 [0.50, 0.88]
Jalan 2017	1	36	5	48	3.1%	0.27 [0.03, 2.18]
Krishnappa 2013	2	91	6	91	4.3%	0.33 [0.07, 1.61]
Takashima 1994	10	268	12	62	14.0%	0.19 [0.09, 0.43]
Zawawi 2016	8	77	26	151	12.6%	0.60 [0.29, 1.27]
Total (95% CI)		**757**		**637**	**100.0%**	**0.56 [0.44, 0.72]**
Total events	82		141			

Heterogeneity: Chi² = 9.29, df = 4 (P = 0.05); I² = 57%
Test for overall effect: Z = 4.66 (P < 0.00001)



Appendix F: Health economic evidence selection

Figure 6: Flow chart of health economic study selection for the guideline



* Non-relevant population, intervention, comparison, design or setting; non-English language
TFT; thyroid function test, FNA; fine-needle aspiration, US; ultrasound, RAI; radioactive iodine, ATDs; antithyroid drugs, Mang; management, SCH; Subclinical hypothyroidism, SCT; Subclinical thyrotoxicosis.

Appendix G: Health economic evidence tables

Study	Cesur 2006 [4, 114, 114, 11]			
Study details	**Population & interventions**	**Costs**	**Health outcomes**	**Cost effectiveness**
Economic analysis: CEA **Study design:** Within trial analysis **Approach to analysis:** Outcomes and resource use from the same trial **Perspective:** Turkish hospital healthcare sector **Time horizon:** Length of treatment **Discounting:** N/A	**Population:** Adults admitted to the outpatient thyroid clinic with nodular goiter (single or multiple) **Patient characteristics:** n=215 patients (285 thyroid nodules) Mean age: 48.7 Male: 36 (17%) **Intervention 1:** Palpation-guided fine-needle aspiration biopsy (PGFNAB) **Intervention 2:** Ultrasound-guided fine-needle aspiration biopsy (UGFNAB)	**Total costs (mean per patient)** Intervention 1: £51 Intervention 2: £64 Incremental (2−1): £13 (95% CI: NR; p=NR) **Currency & cost year:** Currency and cost year unclear, assumed to be 2006 US dollars (presented here as 2006 UK pounds[(a)])] **Cost components incorporated:** The prices of thyroid ultrasonography, PGFNAC, UGFNAC and cytologic examinations.	**Key outcomes:** **True positives:** **Mean** Intervention 1: 0.019 Intervention 2: 0.028 Incremental (2−1): 0.009 (95% CI: NR; p=NR)	**Cancer detected (Intervention 2 versus Intervention 1):** £1,361 per extra cancer detected **Analysis of uncertainty:** No sensitivity analysis was conducted.
Data sources				
Health outcomes: Within trial analysis: single trial of 215 patients (285 nodules) in Ankara hospital. **Quality-of-life weights:** NA. **Cost sources:** Cohort analysis: Hospitals in Ankara, Turkey.				
Comments				
Source of funding: NR. **Limitations:** Turkish hospital health service perspective; outcomes were not valued using QALYs. Data taken from single study				

of 215 patients; currency and cost year not stated, costs taken from private and state hospitals in Turkey; sensitivity analysis not undertaken. **Other:** only 26 patients underwent surgery with no clear inclusion criteria.	
Overall applicability:(b) Partially applicable	**Overall quality:**(c) Potentially serious limitations

Abbreviations: CEA: cost-effectiveness analysis; 95% CI: 95% confidence interval; NR: not reported; N/A: Not applicable; UK: United Kingdom
(a) Converted using 2006 purchasing power parities[40]
(b) Directly applicable / Partially applicable / Not applicable
(c) Minor limitations / Potentially serious limitations / Very serious limitations

Appendix H: Health economic analysis

None

Appendix I: Excluded studies

I.1 Excluded clinical studies

Table 12: Studies excluded from the clinical review

Reference	Exclusion reason
Aksu 2014[1]	Wrong study design: separate sample for each index test
Al Maqbali 2012[2]	Wrong study design: sample consisting of non-diagnostic first FNAC results; diagnostic accuracy calculated based on successive FNACs majority of which under US guidance
Bohacek 2012[3]	Wrong study design: UGFNAC only assessed
Braga 2001[5]	Wrong study design: UGFNAC only assessed
Cai 2006[6]	Wrong study design: separate sample for each index test
Cam 2014[7]	Inappropriate comparison (Ultrasound vs CCT vs DW-MRI)
Can 2008[9]	Wrong study design: separate sample for each index test
Can 2009[8]	No relevant outcomes
Carmeci 1998[10]	Wrong study design: separate sample for each index test
Danese 1998[12]	Wrong study design: separate sample for each index test
de Meer 2012[13]	No relevant outcomes
Deandrea 2002[14]	Inappropriate comparison (palpation vs FNA)
Esfahanian 2016[15]	Wrong study design: UGFNAC only assessed
Hatada 1998[16]	Wrong study design: separate sample for each index test
Izquierdo 2006[17]	Wrong study design: separate sample for each index test
Kawai 2012[19]	Inappropriate comparison (Ultrasound vs FNA)
Kimoto 1999[20]	Wrong study design: UGFNAC only assessed
Lee 2013[23]	Wrong study design: UGFNAC only assessed
Lee 2009[24]	Wrong study design: repeated UGFNAC only assessed
Lee 2013[22]	Wrong study design: UGFNAC only assessed
Leung 2017[25]	Inappropriate comparison (FNA before vs after biopsy centre implementation)
Lew 2010[26]	Non-systematic review
Li 2016[27]	Wrong study design: UGFNAC only assessed
Lin 1997[28]	Wrong study design: UGFNAC only assessed
Mehrotra 2006[29]	Wrong study design: separate sample for each index test
Melany 2017[30]	Non-systematic review
Mirshemirani 2010[31]	Inappropriate comparison (Ultrasound vs FNA)
Mittendorf 2002[32]	Wrong study design: separate sample for each index test
Moon 2007[33]	Inappropriate comparison (Ultrasound vs FNA)
Muruganandham 2009[34]	Wrong study design: separate sample for each index test
Nachiappan 2014[35]	Non-systematic review
Nam-Goong 2004[36]	Wrong study design: UGFNAC only assessed
Newkirk 2000[38]	Wrong study design: UGFNAC only assessed
Ogawa 2001[39]	Wrong study design: UGFNAC only assessed
Peng 2007[41]	Not in English
Rorive 2010[43]	Wrong study design: UGFNAC only assessed
Schwartz 2010[44]	Wrong study design: UGFNAC only assessed (on sample with non-diagnostic PGFNAC)

Reference	Exclusion reason
Singh Ospina 2016[45]	SR not matching PICO
Witt 2015[47]	Wrong study design: separate sample for each index test
Yang 2001[48]	Wrong study design: UGFNAC only assessed
Young 2011[49]	Wrong study design: diagnostic results from UGFNAC only

I.2 Excluded health economic studies

Table 13: Studies excluded from the health economic review

Reference	Reason for exclusion
Can 2009 [8]	This study was assessed as not applicable, as the population did not match the clinical protocol and no relevant outcomes were recorded.