National Institute for Health and Care Excellence

Final

Thyroid cancer: assessment and management

[E] Evidence review for efficacy of repeat fine needle aspiration cytology, active surveillance or discharge

NICE guideline NG230

Evidence reviews underpinning recommendation 1.2.11 in the NICE guideline

December 2022

Final

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/or their carer or guardian.

Local commissioners and/or 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 All NICE guidance is subject to regular review and may be updated or withdrawn.

Copyright

ISBN: 978-1-4731-4864-2

Contents

1 Repeat testing in people with benign or non-diagnostic FNAC

1.1 Review question

1.1.1 For people with fine-needle aspiration samples showing benign cytology or non-diagnostic features but with US / clinical findings indicative of increased risk of malignancy, is it clinically and cost effective to use diagnostic hemithyroidectomy, repeat FNAC, use active surveillance or discharge?

1.1.2 Introduction

The vast majority of thyroid nodules are the result of a benign process, however, thyroid cancer also, typically, presents as a thyroid nodule. In recent years, ultrasound examination of the thyroid has become the standard of care for the initial investigation of thyroid nodules, and where ultrasound shows any features, which are indeterminate or suspicious for malignancy, fine needle aspiration biopsy is undertaken.

Where the cytology is suggestive of a benign diagnosis, uncertainty then remains regarding the underlying reason for the indeterminate or suspicious features on ultrasound and the most appropriate further management to ensure that thyroid malignancy, when present, is diagnosed in a timely fashion, to prevent adverse patient outcomes. The current practice in most centres is to repeat an inadequate or benign FNA if the US/clinical features suggest increased risk of malignancy. This review seeks to determine the best management strategy for people with benign or inadequate findings on FNAC that are in conflict with the prior ultrasound results that suggest a risk of malignancy.

1.1.3 Summary of the protocol

For full details see the review protocol in Appendix A.

Table 1: PICO characteristics of review question

Population	People aged 16 or over who are suspected of having thyroid cancer. They have fine-needle aspiration samples showing benign cytology or non-diagnostic atypical features but have US / clinical findings indicative of increased risk of malignancy
Interventions	• diagnostic hemithyroidectomy • repeat FNAC • active surveillance • discharge without active surveillance
Comparison	Each other
Outcomes	• mortality • quality of life (any validated scores) • thyroid cancer diagnosis
Study design	• Systematic reviews • RCTs

1.1.4 Methods and process

This evidence review was developed using the methods and process described in Developing NICE guidelines: the manual. Methods specific to this review question are described in the review protocol in appendix A and the methods document.

Declarations of interest were recorded according to NICE's conflicts of interest policy.

1.1.5 Effectiveness evidence

1.1.5.1 Included studies

No relevant clinical studies comparing management strategies for people with indeterminate FNA were identified.

See also the study selection flow chart in Appendix C, study evidence tables in Appendix D, forest plots in Appendix E and GRADE tables in Appendix F.

1.1.5.2 Excluded studies

Fifteen studies were identified for full text eligibility assessment. Of these, 11 were narrative reviews or systematic reviews. All references were checked for potential inclusion however none were RCTs. Of the remaining 4, 3 studies were excluded due to having an incorrect study design (for example, non-randomised or no comparison group) and 1 was excluded as the population was unclear and the intervention was not included in the protocol.

See the excluded studies list in Appendix I.

1.1.6 Summary of studies included in the effectiveness evidence

No evidence was identified.

1.1.7 Summary of the effectiveness evidence

No evidence was identified.

1.1.8 Economic evidence

1.1.8.1 Included studies

One health economic study with the relevant comparison was included in this review.[2] This is summarised in the health economic evidence profile below (**Table 2**) and the health economic evidence table in 0Appendix H.

1.1.8.2 Excluded studies

One health economic study was excluded due to assessment of limited applicability or methodological limitations.[30] This is listed in Appendix I, with reasons for exclusion given.

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

1.1.9 Summary of included economic evidence

Table 2: Health economic evidence profile: Repeat FNAC, active surveillance or discharge

Study	Applicability	Limitations	Other comments	Incremental cost	Incremental effects	Cost effectiveness	Uncertainty
Furlan 2005[2] ([Canada])	Partially applicable [(a)]	Potentially serious limitations [(b)]	• Retrospective database analysis • Cost-consequence analysis • Population: Adults with suspected differentiated thyroid cancer and non-diagnostic or benign findings following FNA biopsy • Comparators: 1. Single FNA biopsy (includes the first cytological evaluation from individuals who later underwent sequential FNA biopsies. 2. Sequential FNA biopsy (at least 2 cytological aspirations from the same patient on different dates over ≥1 year) • Follow-up: 2 years	2-1: £33[(c)]	2-1: • Accuracy: 14.4% • Sensitivity: 10.2% • Specificity: 4.3%		Exclusion of all incidental microcarcinomas of 10mm or less found after thyroidectomy did not result in any significant differences in diagnostic accuracy (p=0.898); sensitivity (p=0.965); or specificity (0.594).

Abbreviations: ICER = incremental cost-effectiveness ratio; FNAC = Fine-needle aspiration cytology; NA = not applicable; NR = not reported; QALYs= quality-adjusted life years.

(a) Canadian healthcare context. Discount rate was not reported for a study with a time horizon of 2 years. QALYs were not reported.

(b) Patients were not randomized to treatment groups and had statistically significant differences in baseline patient characteristics between groups. Mortality, quality of life, and other patient outcomes were not included. Study is non-randomized, sample size is small and study was conducted at one hospital with unclear generalizability. Analysis did

not include costs of surgery and other downstream costs and consequences of diagnosis. Cost year not specified; assumed to be the date of the published source used to obtain unit costs. Sensitivity analyses were not included. Relatively old study: accuracy of FNAC may have been improved in the last two decades.

(c) 2004 Canadian dollars converted to 2004 UK pounds.[17]. Cost components incorporated: FNA biopsy

1.1.10 Economic model

A health economic model was developed to assess the cost-effectiveness of several diagnostic pathways for people who received a non-diagnostic (Thy1) or indeterminate (Thy3a and Thy3f) cytology. The full economic report can be viewed in the economic report published alongside the guideline

Population and strategies

Three population with different cancer prevalence were included in the analysis: Thy1, Thy3a and Thy3f. The strategies included are different for each cytology:

- Non-diagnostic Thy1
    - Repeat FNAC and selective use of diagnostic hemithyroidectomy (current practice)
    - Core needle biopsy (CNB) and selective use of diagnostic hemithyroidectomy
    - Routine use of diagnostic hemithyroidectomy
- Indeterminate Thy3a
    - Repeat FNAC and selective use of diagnostic hemithyroidectomy (current practice)
    - Repeat FNAC and selective use of molecular testing
    - CNB and selective use of diagnostic hemithyroidectomy
    - CNB and selective use of molecular testing
    - Routine use of molecular testing
    - Routine use of diagnostic hemithyroidectomy
- Indeterminate Thy3f
    - Routine use of diagnostic hemithyroidectomy (current practice)
    - Routine use of molecular testing

Molecular testing is not clinically indicated for people with a non-diagnostic Thy1, therefore it was not included for this population. Likewise, repeat sampling with FNAC or CNB is rarely of any use after Thy3f, so CNB and FNAC strategies were not included for Thy3f.

A cost-utility analysis was undertaken where quality-adjusted life years (QALYs) and costs from a current UK NHS and personal social services perspective were considered. The analysis was run each time separately for each cytology.

Model structure

- Two decision trees and one Markov model with 50 cycles were developed to estimate costs and consequences of each strategy
- Each diagnostic strategy had its own decision decision tree that was used to estimate the number of people ending up with a correct diagnosis, a missed diagnosis or an avoidable hemithyroidectomy.
- A second decision tree was used to determine the outcomes of surgery (hemithyroidectomy or completion thyroidectomy for people with malignancy). The possible outcomes are: dead, no complications, recurrent laryngeal nerve (RLN) injury, transient or permanent hypoparathyroidism, re-admission
- Depending on the outcomes of the surgical and diagnostic decision tree, people enter the Markov model in a specific state characterized by whether the node was benign or malignant, whether the cancer was misdiagnosed, whether the person received surgery, and whether surgery caused one of the two long-term complications: recurrent laryngeal nerve (RLN) injury and hypoparathyroidism.
- The Markov model was utilized to estimate long-term consequences of cancer, risk of recurrence and long-term costs of monitoring and health states.

- People with malignant nodules are at a risk of recurrence during each cycle, with a higher probability if their diagnoses was missed or the cancer is untreated, thus moving to a new state representing the progressed state of disease. This state is characterized by a higher cost, lower QoL and higher mortality

Data sources

- Accuracy data were estimated from the studies included in 1.5 review with the addition of external meta-analysis and studies when necessary. These were four studies on the accuracy of molecular tests[16, 18 10, 22] that were assessed through QUADAS tables and a meta-analysis on repeat sampling with FNAC or CNB[19]
- Surgery outcomes were taken from the latest BAETS audit[26]
- Recurrence rates were informed from HiLo trial[12]
- Mortality in the recurrent states or after a complication was estimated using published studies comparing people with recurrence or complication with the general population
- Disutility factors associated with each complication or adverse event were calculated using the study from Kebebew 2005[7]

Cost

- The price of the molecular tests were, when possible, obtained from the manufacturers or, alternatively, estimated from a published study[9]. Costs of sample packaging and shipping were also included
- Cost of FNAC and CNB were collected from the NHS Reference Costs 2019-2020[15]
- The costs of a hemithyroidectomy, completion thyroidectomy and total thyroidectomy were estimated using NHS Reference Costs combined with LOS data from BAETS audit[26]
- The cost associated with surgery complications and radioactive iodine ablation (RAI) were estimated using BNF[6] and PCA[5] for pharmaceutical costs, and NHS Reference Costs 2019-2020[15] for healthcare costs
- Thyroid hormone replacement costs were calculated using BNF[6] and PCA[5]
- The costs of the various tests needed for post-surgery monitoring were informed from NICE thyroid disease guideline[14] and NHS Reference Costs 2019/2020[15]

Results

The results for the non-diagnostic cytology Thy1 are illustrated in **Table 3**.

Table 3: Diagnostic performance – Thy1 (probabilistic)

Strategy	Cost per patient	QALYs per patient	NMB (£20,000)	Probability 1st ranked
Repeat FNAC & HT	£2,094 (£1,570 to £2,826)	17.17 (16.79 to 17.35)	£341,346 (£333,310 to £345,253)	2%
CNB & HT	£1,815 (£1,223 to £2,600)	17.18 (16.81 to 17.36)	£341,874 (£333,729 to £345,965)	98%
HT	£3,018 (£2,586 to £3,617)	17.13 (16.77 to 17.31)	£339,608 (£331,841 to £343,456)	0%

Abbreviations: CNB = core needle biopsy; FNAC = fine needle aspiration cytology; HT = hemithyroidectomy; NMB = net monetary benefit; QALY = quality-adjusted life year

In almost all simulations CNB was the most cost-effective strategy compared with repeat FNAC and hemithyroidectomy. None of the scenario analysis altered the conclusions.

Table 8 shows the probabilistic results for the indeterminate category Thy3a.

Table 4: Diagnostic performance – Thy3a (probabilistic)

Strategy	Cost per patient	QALYs per patient	NMB (£20,000)	Probability 1st ranked
Afirma-GSC	£5,318 (£4,830 to £5,879)	16.88 (16.28 to 17.20)	£332,198 (£319,779 to £338,890)	0%
ThyroSeq V3	£3,819 (£3,339 to £4,529)	16.87 (16.27 to 17.20)	£333,658 (£321,259 to £340,376)	0%
ThyGenx/Thyra MIR	£4,759 (£4,243 to £5,349)	16.88 (16.24 to 17.21)	£332,741 (£319,806 to £339,693)	0%
ThyroSeq V1	£4,045 (£3,500 to £4,653)	16.80 (16.15 to 17.17)	£332,013 (£318,665 to £339,586)	0%
FNAC & HT	£3,078 (£2,629 to £3,601)	16.90 (16.33 to 17.20)	£334,992 (£323,115 to £341,241)	18%
FNAC & MT[(a)]	£3,089 (£2,620 to £3,638)	16.90 (16.32 to 17.20)	£334,902 (£322,854 to £341,278)	18%
CNB & HT	£3,018 (£2,542 to £3,573)	16.90 (16.33 to 17.21)	£335,079 (£323,131 to £341,372)	34%
CNB & MT[(a)]	£3,025 (£2,527 to £3,597)	16.90 (16.32 to 17.21)	£335,021 (£322,989 to £341,380)	30%
HT	£3,784 (£3,400 to £4,247)	16.89 (16.32 to 17.19)	£333,935 (£322,279 to £340,139)	0%

Abbreviations: CNB = core needle biopsy; FNAC = fine needle aspiration cytology; HT = hemithyroidectomy; MT = molecular testing

(a) The molecular test in FNAC & MT and CNB & MT strategies is ThyroSeq V3

In none of the simulations routine use of molecular testing or routine hemithyroidectomy were found to be cost-effective. The two CNB strategies were more likely to be cost-effective compared with repeat FNAC strategies: 64% vs 36%. Comparing CNB and selective use of molecular testing (ThyroSeq V3) with CNB and selective use of hemithyroidectomy, the latter was found to be cheaper and slightly more effective, making it more cost-effective. Moreover, CNB and selective use of hemithyroidectomy was the most cost-effective strategy in a higher number of simulations compared to CNB and selective use of hemithyroidectomy (34% vs 30%).

The scenario analysis showed that when a lower prevalence of cancer was assumed and when the disutility associated with cancer recurrence was applied for one cycle only, CNB and selective use of molecular testing becomes the most cost-effective strategy.

Table 5 shows the probabilistic results for the indeterminate category Thy3f.

Table 5: Diagnostic performance – Thy3f (probabilistic)

Strategy	Cost per patient	QALYs per patient	NMB (£20,000)	Probability 1st ranked
Afirma-GSC	£5,867 (£5,223 to £6,609)	16.69 (15.87 to 17.12)	£327,882 (£310,894 to £336,889)	0%
ThyroSeq V3	£4,452 (£3,820 to £5,195)	16.78 (16.02 to 17.17)	£331,215 (£315,631 to £339,366)	57%
ThyGenx/ThyraMIR	£5,014 (£4,335 to £5,801)	16.60 (15.64 to 17.12)	£326,888 (£307,388 to £337,730)	1%
ThyroSeq V1	£4,485 (£3,800 to £5,271)	16.65 (15.81 to 17.11)	£328,601 (£311,248 to £338,006)	1%
HT	£4,137 (£3,630 to £4,765)	16.77 (16.03 to 17.15)	£331,361 (£316,111 to £339,142)	40%

Abbreviations: HT = hemithyroidectomy; NMB = net monetary benefit; QALY = quality-adjusted life year

Although routine use of molecular testing with GSC test ThyroSeq V3 was more likely to be cost-effective, routine use of hemithyroidectomy had a slightly higher average NMB suggesting that molecular testing had a higher degree of uncertainty compared with diagnostic hemithyroidectomy: whereas accuracy and feasibility of molecular testing in England is unclear, hemithyroidectomy has always 100% sensitivity and specificity. The scenario analysis found the results to be very dependent on the assumptions on the cost of molecular tests: when a higher price was used or when an additional FNAC was deemed necessary to extract the sample, routine use of diagnostic hemithyroidectomy became cost-effective instead.

Two threshold analyses on the cost of molecular tests and risk of malignancy (ROM) were conducted for Thy3a and Thy3f cytologies. Selective use of molecular tests after CNB became cost-effective for Thy3a when the prices drop between £1,500 (ThyGenX/ThyraMIR) and £1,100 (Afirma-GCS). ThyroSeq V3 has a price of £1,407 which is above its threshold price of £1,200 explaining why it was not found to be cost-effective in England. Only ThyroSeq V3 was found to be potentially cost-effective in England when its price drops below £1,500. This is just above the estimated price of £1,407 which explains the uncertainty of its cost-effectiveness in this cytology.

The threshold analysis on ROM found a threshold ROM of 23% and 35% for, respectively, CNB and selective use of molecular testing in Thy3a and routine use of molecular testing in Thy3f. These are very similar to the estimated cancer prevalence in these two categories, 25% in Thy3a and 31% in Thy3f, suggesting that even a slight change in the prevalence of cancer in one of these two cytologies may change the conclusion of this analysis.

1.1.11 Economic evidence statements

- One cost-consequence analysis found sequential FNAC to be more costly but more accurate compared to single FNAC in people with non-diagnostic or benign cytology. The analysis was assessed as partially applicable with potentially serious limitations.
- One original cost-utility analysis found CNB to dominate repeat FNAC and diagnostic hemithyroidectomy in people with non-diagnostic Thy1 and indeterminate Thy3a cytologies. The analysis was assessed as directly applicable with minor limitations.

1.1.12 The committee's discussion and interpretation of the evidence

1.1.12.1 The outcomes that matter most

The outcomes considered for this review were mortality, health related quality of life, and thyroid cancer diagnosis. For purposes of decision-making all outcomes were equally regraded as being of critical importance. No evidence was identified for any of these outcomes as no relevant articles were included in the review.

1.1.12.2 The quality of the evidence

No evidence was included in the review.

1.1.12.3 Benefits and harms

This section discusses all options for management following the initial FNAC results. It draws on data from this review, the review on the initial FNAC tests (evidence report D), the review on molecular testing (evidence report F) and committee consensus to make recommendations. No evidence was identified for repeat FNAC or molecular testing and the committee used a mixture of the evidence from the diagnostic accuracy review, the results from the diagnostic economic model and consensus to make recommendations. The committee agreed that although the evidence in report D is indirect evidence for this review, they did not anticipate the accuracy would be that different for repeat testing. They were happy to use it in the model.

Overall, the committee agreed that once the initial results from FNAC are reported it is common practice to do further testing of inadequate (Thy 1), benign (Thy2) and indeterminate (Thy 3) results to confirm diagnosis and avoid unnecessary surgery where possible. While FNAC is the test used following initial ultrasound the committee agreed that other options may offer a better management strategy when repeating tests.

Results that are suspicious of malignancy (Thy 4) or malignant (Thy 5) usually go straight to exploratory surgery.

The committee acknowledged that, in some cases, FNAC and CNB can produce artefacts that can simulate malignancy on histological assessment of subsequent resections. CNB can produce larger artefacts due to the larger size of its needle. Although this may require additional awareness when assessing resected nodules, the committee agreed that a trained physician would be perfectly able to discriminate real lesion from artefacts.

Thy1 - inadequate FNAC results

A Thy1 (inadequate) result is one where insufficient aspirate is available to permit any diagnosis – there is effectively a dearth of information. In such a case there is a clear need to offer repeat sampling and the committee agreed that this should be with either core needle biopsy or FNAC with ROSE, or FNAC alone if ROSE unavailable and CNB is unavailable or inappropriate. The committee recognised that core biopsy or FNAC with ROSE were the

preferred option for repeating non-diagnostic FNAC attempts, because they may be less prone to inadequate results. However, if this second attempt were also unsuccessful the committee recommended that the patient should be sent for a diagnostic hemithyroidectomy. This was made by consensus, on the basis that there had to be a limit to the number of repeated attempts, the number of such cases would be small, and that sending for surgery would be the most conservative and safe approach.

Thy 1c – cystic lesion

The committee agreed that repeat sampling with FNAC should be offered but to consider diagnostic hemithyroidectomy if the repeat sample is also Thy 1c and the ultrasound appearances are concerning. FNAC with ROSE and CNB were not considered useful after this category as Thy1c is non-adequate for cystic lesion and it is not operator- or technique-dependent.

Thy 2 and Thy 2c – benign FNAC results

Even taking the best result for direct smear, liquid based cytology and cellblock into consideration the committee were concerned that some people with malignancy would be missed. the best sensitivity from our analysis should a level of 0.937 and the committee agreed that this would lead to 6.3% of people with malignancy being missed. Calculation of the negative predictive value at the sensitivity of 0.937 and the specificity of 0.825, at our estimated prevalence of malignancy in people having FNAC of 11%, yielded a value of 0.991. This indicated that only 0.9% of people with a Thy 2 (benign) FNAC grade would have a true malignancy. At first sight, this ameliorated some of the concern, but on reflection it was realised that this favourable result was an artefact of the low prevalence of malignancy overall, and did not alter the fact that amongst those people with malignancy a significant proportion would remain undetected. However, a solution to this problem was provided through discussion of how failure to detect a truly malignant nodule may be partly mediated by random sampling error, where the aspirated sample, by chance, does not represent the population of predominantly malignant cells in the nodule. If this is the case, repetition of the aspiration is unlikely to lead to a second such error as the probability of two such errors occurring together would be subject to the multiplicative law. For example, a 0.1 probability of such an event happening once would imply only a 0.01 probability of two such consecutive events. Therefore, the committee suggested, by consensus, that a way of improving the sensitivity would be to repeat the ultrasound after any benign (Thy2) FNAC readings in those people who had been selected for FNAC based on suspicious US findings. These could be legitimately regarded as people in whom there were good ultrasonographic reasons to doubt the verisimilitude of the FNAC result. The committee agreed that this is in line with current practice If the second ultrasound was positive, then repeat sampling should be done by FNAC. If a benign reading were found on second FNAC result this would confirm the original FNAC result, but if a malignant reading were found this would strongly suggest initial sampling error and allow replacement of the original FNAC result, thereby improving sensitivity. The committee discussed how this would not represent a change in current practice and so would not necessarily introduce new costs. The committee agreed that core needle biopsy could be used as an alternative to FNAC, because although it is invasive and more expensive than FNAC, it can extract more material.

Thy 3a - Neoplasm possible - atypia

For any indeterminate result an additional sample should be offered to establish whether the nodule was malignant or not. The committee agreed that sampling should be repeated and made a consider recommendation for CNB (or FNAC if CNB is unavailable or inappropriate). The committee noted that Thy3a findings can often appear by chance in people that may often show Thy2 findings on a repeat sampling with CNB or FNAC. Therefore, this approach

would maximise specificity whilst ensuring that those who show a repeat Thy3a finding (or possibly a worse grading) are appropriately managed with surgery. If the result is still Thy 3a, the committee agreed that diagnostic hemithyroidectomy or active surveillance should be considered. While they noted that most people would not have a malignancy in this group and active surveillance could be an option, they also agreed that a minority of people may have malignancy and diagnostic accuracy would be the best approach should this be suspected.

Thy 3f - Neoplasm possible, suggesting follicular neoplasmThy 4 and Thy 5 – suspicion of malignancy or malignant

If Thy 3f, 4 or 5 were obtained (equivalent to Bethesda IV, V and VI) then surgery should be considered. The recommendations for Thy 3f, 4 and 5 were based on extrapolation of the diagnostic accuracy evidence that these groups would contain a large quantity of people with malignancy, and that failure to send them for surgery would risk missing people with malignancy. Although the economic model suggested molecular testing could be cost effective in Thy 3f the committee agreed that that repeat sampling is less useful. Diagnostic hemithyroidectomy is justified by the high risk of malignancy in this group (around 30%). There were also concerns that, if not followed up with surgery, final diagnosis after Thy 3f could take longer to happen. This would delay treatment for a potentially malignant tumour, create uncertainty for the person, and in some centres lead to a longer delay than is allowed by NHS cancer targets.

The committee discussed how the type of surgery undertaken would depend upon the FNAC grade, as well as other information. For FNAC grades denoting a higher probability of malignancy, such as Thy 4 and Thy 5, surgery would be undertaken with the assumption that malignancy was probably present, involving a diagnostic or therapeutic hemi-thyroidectomy, total thyroidectomy or isthmectomy, possibly combined with central and/or lateral neck dissection, according to the clinical circumstances. However, for Thy 3a and Thy 3f, where there would be greater doubt about the likelihood of malignancy, and particularly for Thy1, where the probability of malignancy would be low, the surgery would be performed initially for diagnostic reasons, and a hemi thyroidectomy or isthmectomy would be preferred.

1.1.12.4 1.1.12.4 Cost effectiveness and resource use

One health economics evidence was included. This was a cost consequence analysis on offering sequential FNAC (at least 2) instead of single FNAC to people with suspected thyroid cancer.

The analysis was assessed to be partially applicable as it was conducted in Canada. Moreover, the analysis had potentially serious limitations due to being not randomized and relying on very old sources. The analysis showed that at an increased price of £46, sequential FNAC increases the sensitivity by 10.2% and specificity by 4.2% compared to single FNAC.

There was no included study in the clinical review. The committee agreed that understanding the optimal diagnostic pathway for people who received a non-diagnostic or indeterminate FNAC was extremely important in England. Around 30% of FNAC samples result in an indeterminate cytology (Thy3a and Thy3f), and between 5 to 10% are non-diagnostic. Assuming 5,600 fine-needle aspirations performed by the NHS every year, as per the latest NHS Refence Costs 2019/2020, around 2,000 of those would require further management. Hence, any recommendation changing current practice would likely require a large use of NHS resource. Therefore, an economic plan to look at the most cost-effective diagnostic pathway after a non-diagnostic (Thy1) or indeterminate cytology (Thy3a or Thy3f) was approved by the committee and a lifetime model was developed incorporating several

strategies: for the non-diagnostic cytology Thy1, current practice of repeat FNAC and selective use of hemithyroidectomy was compared with core-needle biopsy (CNB) and selective use of hemithyroidectomy and routine use of hemithyroidectomy; for the indeterminate category Thy3a, current practice of repeat FNAC and selective use of hemithyroidectomy was compared with FNAC and selective use of molecular testing, CNB and selective use of hemithyroidectomy and molecular testing, routine hemithyroidectomy, and routine molecular testing; for the other indeterminate category Thy3f, the committee were aware that repeat FNAC or CNB can rarely be beneficial, therefore the analysis compared current practice (routine hemithyroidectomy) with routine molecular testing only.

The full economic model can be viewed in the Economic Report. The model found CNB cost effective in people who received Thy1 compared to the two alternatives, repeat FNAC and diagnostic hemithyroidectomy. In the probabilistic analysis, CNB had a higher Net Monetary Benefit (NMB) than repeat FNAC and a very high probability of being cost-effective: 99%. In none of the 10,000 Monte Carlo simulation hemithyroidectomy managed to be the most cost-effective strategy, which was widely expected by the committee as ROM in Thy1 is too low to justify a diagnostic surgery. The committee acknowledged that, for people with a non-diagnostic Thy1 cytology, a CNB or FNAC with ROSE can effectively reduce the likelihood of obtaining a second non-diagnostic cytology thus avoiding additional diagnostic surgeries and harms. The evidence on ROSE was discussed in Evidence Review D which found ROSE potentially very useful in centres with a concerningly high non-diagnostic rate. The committee noted that there is heterogeneity in clinical practice in the UK with some centres preferring CNB or ROSE in the management of Thy1 cytologies. Moreover, it is unlikely that in most centres CNB and ROSE are both available and implementing both tests everywhere could be fairly expensive and require a significant amount of time. Hence, the committee recommended either CNB or FNAC with ROSE as follow-up tests for people who received a Thy1 non-diagnostic cytology. This should ensure flexibility and allow most centres to adopt their preferred and most available technique in the management of Thy1. A second recommendation was made to consider FNAC alone if CNB and FNAC with ROSE are unavailable or inappropriate, i.e. the nodule is located too close to a blood vessel to use a large needle. For further inadequate results after repeat sampling, a consider recommendation for diagnostic hemithyroidectomy was made. This reflects the Committee's view that people with Thy1 should not always undergo surgery, even when sequential tests end up being non-diagnostic, as the risk of malignancy with this cytology remains relatively low and additional clinical considerations should assessed alongside when making a decision. The management of a Thy1c cystic lesion was separately defined in the recommendations as these lesions are non-diagnostic for cyst and, therefore, would not benefit from a CNB or FNAC with ROSE.

Similarly to Thy1, the analysis on Thy3a found that CNB was the most cost effective strategy compared to all the alternatives included. In the probabilistic analysis, the two CNB strategies, CNB with selective use of molecular testing and CNB with selective use of hemithyroidectomy, had together a probability of 65% of being cost effective and a higher NMB than each alternative, including FNAC and hemithyroidectomy which reflects current practice. In line with the evidence, the committee recommended CNB as the preferred method of sampling after Thy3a, although room was left to use FNAC instead based on local preference and availability as in the case of Thy1. In case of a further inconclusive result, the results of the model indicate that selective use of hemithyroidectomy is cost-effective compared to selective use of molecular testing with ThyroSeq V3. Consequently, the committee decided to recommend diagnostic hemithyroidectomy or active surveillance when repeat sampling gives a further inconclusive result.

Regarding the last category analysed Thy3f, routine use of molecular testing with ThyroSeq V3 had a higher probability of being cost-effective (57%) but had a lower probabilistic NMB than hemithyroidectomy highlighting the difference in uncertainty between these two strategies: whereas hemithyroidectomy had a minimal level of uncertainty as it represents a

perfectly sensitive and specific test, the same is not true for molecular tests whose true accuracy is uncertain and its cost in England is unclear being largely produced in the US. The Committee acknowledged that the uncertainty on cost-effectiveness of molecular testing after Thy3f would not allow to make a positive recommendation for this category, as it would represent a very important deviation from current practice in England, where Thy3f is generally followed-up with diagnostic hemithyroidectomy. Moreover, the risk of malignancy in Thy3f is very significant, as around 30% of Thy3f nodules are found to be malignant after surgery. The NHS has a 28 days target from referral to final diagnosis for cancer. As molecular tests are not widely available in the UK, a positive recommendation for molecular testing after Thy3f implies that a large number of samples would be shipped to laboratories overseas at least in the short term, which is likely to significantly increase time between GP referral and first treatment. This risks to lead several centres to fail in meeting the target which may negatively affect wellbeing and risk of recurrent/persistent disease in people with thyroid cancer.

Regarding the other RCPath categories Thy2, Thy4 and Thy5, the Committee made recommendations in line with current practice. Thy2 is the benign cytology and has the lowest risk of malignancy: 5%. The committee made a recommendation in line with current practice and latest BTS guidelines recommending repeat sampling with FNAC only if there are clinical concerns on ultrasound (EU TIRADS > 4). Thy4 and Thy5 are the suspicious and malignant categories with a risk of malignancy of, respectively, 79% and 98%. Given the very high cancer prevalence, the Committee recommended surgery, either diagnosis or therapeutic, which appears to be strongly justified from a health economics point of view due to the low number of false positives in these two categories.

1.1.13 Recommendations supported by this evidence review

This evidence review supports recommendations 1.2.11.

References

1. Castellana M, Trimboli P, Piccardo A, Giovanella L, Treglia G. Performance of 18f-fdg pet/ct in selecting thyroid nodules with indeterminate fine-needle aspiration cytology for surgery. A systematic review and a meta-analysis. Journal of Clinical Medicine. 2019; 8(9):28

2. Furlan JC, Bedard YC, Rosen IB. Single versus sequential fine-needle aspiration biopsy in the management of thyroid nodular disease. Canadian Journal of Surgery. 2005; 48(1):12-18

3. Gerhard R, Boerner SL. Evaluation of indeterminate thyroid cytology by second-opinion diagnosis or repeat fine-needle aspiration: which is the best approach? Acta Cytologica. 2015; 59(1):43-50

4. GRADE Working Group. The grading of recommendations assessment, development and evaluation (grade) working group website. 2011. Available from: http://www.gradeworkinggroup.org/ Last accessed:

5. Health and Social Care Information Centre. Prescription cost analysis, England [updated December 2021]. Available from: https://www.nhsbsa.nhs.uk/prescription-data/dispensing-data/prescription-cost-analysis-pca-data Last accessed:

6. Joint Formulary Committee. British National Formulary (BNF). 67th ed. London. British Medical Association and The Royal Pharmaceutical Society of Great Britain,. 2014. Available from: http://www.bnf.org.uk

7. Kebebew E, Greenspan FS, Clark OH, Woeber KA, McMillan A. Anaplastic thyroid carcinoma. Treatment outcome and prognostic factors. Cancer. 2005; 103(7):1330-1335

8. Khoncarly SM, Tamarkin SW, McHenry CR. Can ultrasound be used to predict malignancy in patients with a thyroid nodule and an indeterminate fine-needle aspiration biopsy? Surgery (United States). 2014; 156(4):967-971

9. Kuo JH, McManus C, Graves CE, Madani A, Khokhar MT, Huang B et al. Updates in the management of thyroid nodules. Current Problems in Surgery. 2019; 56(3):103-127

10. Labourier E, Shifrin A, Busseniers AE, Lupo MA, Manganelli ML, Andruss B et al. Molecular Testing for miRNA, mRNA, and DNA on Fine-Needle Aspiration improves the preoperative diagnosis of thyroid nodules with indeterminate cytology. The Journal of Clinical Endocrinology & Metabolism. 2015; 100(7):2743-2750

11. Lee J, Kim BK, Sul HJ, Kim JO, Lee J, Sun WY. Negative pressure is not necessary for using fine-needle aspiration biopsy to diagnose suspected thyroid nodules: a prospective randomized study. Annals of surgical treatment and research. 2019; 96(5):216-222

12. Mallick U, Harmer C, Yap B, Wadsley J, Clarke S, Moss L et al. Ablation with low-dose radioiodine and thyrotropin alfa in thyroid cancer. New England Journal of Medicine. 2012; 366(18):1674-1685

13. National Institute for Health and Care Excellence. Developing NICE guidelines: the manual. London. National Institute for Health and Care Excellence, 2014. Available

from:
http://www.nice.org.uk/article/PMG20/chapter/1%20Introduction%20and%20overview

14. National Institute for Health and Clinical Excellence. Thyroid disease: assessment and management NG145: Evidence Review [F] Monitoring thyroid disease. 2019. Available from: https://www.nice.org.uk/guidance/ng145

15. NHS England and NHS Improvement. National Cost Collection Data Publication 2019-2020. London. 2020. Available from: https://www.england.nhs.uk/wp-content/uploads/2021/06/National-Cost-Collection-2019-20-Report-FINAL.pdf

16. Nikiforov YE, Baloch ZW. Clinical validation of the ThyroSeq v3 genomic classifier in thyroid nodules with indeterminate FNA cytology. Cancer Cytopathology. 2019; 127(4):225-230

17. Organisation for Economic Co-operation and Development (OECD). Purchasing power parities (PPP). 2021. Available from: http://www.oecd.org/std/ppp Last accessed: 24/03/2022.

18. Patel K, Patel A, Shah K, Patel N. FNAC is a primary diagnostic tool in thyroid swelling. Biosciences Biotechnology Research Asia. 2014; 11(3):1873-1876

19. Pyo JS, Sohn JH, Kang G. Core needle biopsy is a more conclusive follow-up method than repeat fine needle aspiration for thyroid nodules with initially inconclusive results: A systematic review and meta-analysis. Journal of Pathology & Translational Medicine. 2016; 50(3):217-224

20. Qiu Y, Xing Z, Liu J, Peng Y, Zhu J, Su A. Diagnostic reliability of elastography in thyroid nodules reported as indeterminate at prior fine-needle aspiration cytology (FNAC): a systematic review and Bayesian meta-analysis. European Radiology. 2020; 30(12):6624-6634

21. Romitelli F, Di Stasio E, Santoro C, Iozzino M, Orsini A, Cesareo R. A comparative study of fine needle aspiration and fine needle non-aspiration biopsy on suspected thyroid nodules. Endocrine Pathology. 2009; 20(2):108-113

22. Shrestha RT, Evasovich MR, Amin K, Radulescu A, Sanghvi TS, Nelson AC et al. Correlation between histological diagnosis and mutational panel testing of thyroid nodules: A two-year institutional experience. Thyroid : official journal of the American Thyroid Association. 2016; 26(8):1068-1076

23. Straccia P, Rossi ED, Bizzarro T, Brunelli C, Cianfrini F, Damiani D et al. A meta-analytic review of the Bethesda System for Reporting Thyroid Cytopathology: Has the rate of malignancy in indeterminate lesions been underestimated? Cancer Cytopathology. 2015; 123(12):713-722

24. Suh CH, Baek JH, Kim KW, Sung TY, Kim TY, Song DE et al. The role of core-needle biopsy for thyroid nodules with initially nondiagnostic fine-needle aspiration results: A systematic review and meta-analysis. Endocrine Practice. 2016; 22(6):679-688

25. Suh CH, Baek JH, Park C, Choi YJ, Lee JH. The role of core needle biopsy for thyroid nodules with initially indeterminate results on previous fine-needle aspiration: A systematic review and meta-analysis. AJNR: American Journal of Neuroradiology. 2017; 38(7):1421-1426

26. The British Association of Endocrine & Thyroid Surgeons. Sixth National Audit Report. 2021. Available from: https://www.baets.org.uk/wp-content/uploads/BAETS-Sixth-Audit-Report-FINAL.pdf

27. Trimboli P, Giovanella L. Reliability of core needle biopsy as a second-line procedure in thyroid nodules with an indeterminate fine-needle aspiration report: a systematic review and meta-analysis. Ultrasonography. 2018; 37(2):121-128

28. Trimboli P, Treglia G, Sadeghi R, Romanelli F, Giovanella L. Reliability of real-time elastography to diagnose thyroid nodules previously read at FNAC as indeterminate: a meta-analysis. Endocrine. 2015; 50(2):335-343

29. Valerio E, Pastorello RG, Calsavara V, Porfirio MM, Engelman GG, Francisco Dalcin J et al. Should we wait 3 months for a repeat aspiration in non-diagnostic/indeterminate thyroid nodules? A cancer centre experience. Cytopathology. 2020; 31(6):525-532

30. van Roosmalen J, van Hemel B, Suurmeijer A, Groen H, Ruitenbeek T, Links TP et al. Diagnostic value and cost considerations of routine fine-needle aspirations in the follow-up of thyroid nodules with benign readings. Thyroid. 2010; 20(12):1359-1365

31. Vriens D, de Wilt JH, van der Wilt GJ, Netea-Maier RT, Oyen WJ, de Geus-Oei LF. The role of [18F]-2-fluoro-2-deoxy-d-glucose-positron emission tomography in thyroid nodules with indeterminate fine-needle aspiration biopsy: systematic review and meta-analysis of the literature. Cancer. 2011; 117(20):4582-4594

32. Wang N, Zhai H, Lu Y. Erratum: Is fluorine-18 fluorodeoxyglucose positron emission tomography useful for the thyroid nodules with indeterminate fine needle aspiration biopsy? A meta-analysis of the literature (Journal of Otolaryngology - Head and Neck Surgery (2014) 43 (43)). Journal of Otolaryngology Head and Neck Surgery. 2014; 43(November)

Appendices

Appendix A – Review protocols

1.1.13.1 Review protocol for management strategy in people with indeterminate FNA

Field	Content
PROSPERO registration number	CRD42021244449
Review title	The clinical and cost effectiveness of diagnostic hemithyroidectomy, repeat FNAC, active surveillance or discharge, in people with fine-needle aspiration samples showing benign cytology or non-diagnostic features, but with US / clinical findings indicative of increased risk of malignancy.
Review question	For people with fine-needle aspiration samples showing benign cytology or non-diagnostic features but with US / clinical findings indicative of increased risk of malignancy, is it clinically and cost effective to use diagnostic hemithyroidectomy, repeat FNAC, use active surveillance or discharge?
Objective	To determine the best management strategy for people with indeterminate findings on FNAC
Searches	The following databases (from inception) will be searched: • Cochrane Central Register of Controlled Trials (CENTRAL) • Cochrane Database of Systematic Reviews (CDSR) • Embase • MEDLINE

	Searches will be restricted by: • English language • Human studies • Letters and comments are excluded. Other searches: • Inclusion lists of relevant systematic reviews will be checked by the reviewer. The searches may be re-run 6 weeks before final committee meeting and further studies retrieved for inclusion if relevant. The full search strategies will be published in the final review. Medline search strategy to be quality assured using the PRESS evidence-based checklist (see methods chapter for full details).
Condition or domain being studied	Thyroid cancer
Population	Inclusion: People aged 16 or over who are suspected of having thyroid cancer. They have fine-needle aspiration samples showing benign cytology or non-diagnostic atypical features but have US / clinical findings indicative of increased risk of malignancy Exclusion: Children and young people under 16 years

Intervention	• diagnostic hemithyroidectomy • repeat FNAC • active surveillance • discharge without active surveillance
Comparator	Each other
Types of study to be included	• Systematic reviews • RCTs Non-randomised studies will be excluded.
Other exclusion criteria	Non-English language studies. Conference abstracts will be excluded as it is expected there will be sufficient full text published studies available. People with indications for surgery other than risk of malignancy (such as hyperthyroidism, pressure effects)
Context	Occasionally there is uncertainty when FNAC results are benign or non-diagnostic but high clinical suspicion of cancer remains based on clinical findings or US findings. This question aims to compare the possible courses of action in this scenario to enable recommendation of the best approach.
Primary outcomes (critical outcomes)	All outcomes are considered equally important for decision making and therefore have all been rated as critical: mortality • quality of life (any validated scores)

	• thyroid cancer diagnosis Time of follow up: longest available
Data extraction (selection and coding)	EndNote will be used for reference management, sifting, citations and bibliographies. Titles and/or abstracts of studies retrieved using the search strategy and those from additional sources will be screened for inclusion. 10% of the abstracts will be reviewed by two reviewers, with any disagreements resolved by discussion or, if necessary, a third independent reviewer. The full text of potentially eligible studies will be retrieved and will be assessed for eligibility in line with the criteria outlined above. • 10% of all evidence reviews are quality assured by a senior research fellow. This includes checking: • papers were included /excluded appropriately • a sample of the data extractions • correct methods are used to synthesise data • a sample of the risk of bias assessments Disagreements between the review authors over the risk of bias in particular studies will be resolved by discussion, with involvement of a third review author where necessary. An in-house developed database; EviBase, will be used for data extraction. A standardised form is followed to extract data from studies (see Developing NICE guidelines: the manual section 6.4) and for undertaking assessment of study quality. Summary evidence tables will be produced including information on: study setting; study population and participant demographics and baseline characteristics; details of the intervention and control interventions; study methodology' recruitment and missing data rates; outcomes and times of measurement; critical appraisal ratings.

Risk of bias (quality) assessment	Risk of bias will be assessed using the appropriate checklist as described in Developing NICE guidelines: the manual. For Intervention reviews the following checklist will be used according to study design being assessed: • Systematic reviews: Risk of Bias in Systematic Reviews (ROBIS) • Randomised Controlled Trial: Cochrane RoB (2.0)
Strategy for data synthesis	Where possible, data will be meta-analysed. Pairwise meta-analyses will be performed using Cochrane Review Manager (RevMan5) to combine the data given in all studies for each of the outcomes stated above. A fixed effect meta-analysis, with weighted mean differences for continuous outcomes and risk ratios for binary outcomes will be used, and 95% confidence intervals will be calculated for each outcome. Heterogeneity between the studies in effect measures will be assessed using the I^2 statistic and visually inspected. We will consider an I^2 value greater than 50% indicative of substantial heterogeneity. Sensitivity analyses will be conducted based on pre-specified subgroups using stratified meta-analysis to explore the heterogeneity in effect estimates. If this does not explain the heterogeneity, the results will be presented using random-effects. GRADE pro will be used to assess the quality of each outcome, taking into account individual study quality and the meta-analysis results. The 4 main quality elements (risk of bias, indirectness, inconsistency and imprecision) will be appraised for each outcome. Publication bias is tested for when there are more than 5 studies for an outcome. Other bias will only be taken into consideration in the quality assessment if it is apparent. Where meta-analysis is not possible, data will be presented and quality assessed individually per outcome.

	If sufficient data is available to make a network of treatments, WinBUGS will be used for network meta-analysis.
Analysis of sub-groups	<u>Stratification</u> None <u>Sub-grouping</u> If serious or very serious heterogeneity (I2>50%) is present within any stratum, sub-grouping will occur according to the following strategies: • Benign vs inadequate vs indeterminate (based on FNA findings) • Indeterminate vs suspicious (based on US findings) • Size of nodule (<4cm, >=4cm)
Type and method of review	☒ Intervention ☐ Diagnostic ☐ Prognostic ☐ Qualitative ☐ Epidemiologic ☐ Service Delivery ☐ Other (please specify)
Language	English
Country	England
Named contact	**Named contact**

	National Guideline Centre **Organisational affiliation of the review** National Institute for Health and Care Excellence (NICE) and the National Guideline Centre
Review team members	From the National Guideline Centre: Carlos Sharpin, Guideline lead Mark Perry, Senior systematic reviewer Alfredo Mariani, Health economist Lina Gulhane, Head of Information specialists
Funding sources/sponsor	This systematic review is being completed by the National Guideline Centre which receives funding from NICE.
Conflicts of interest	All guideline committee members and anyone who has direct input into NICE guidelines (including the evidence review team and expert witnesses) must declare any potential conflicts of interest in line with NICE's code of practice for declaring and dealing with conflicts of interest. Any relevant interests, or changes to interests, will also be declared publicly at the start of each guideline committee meeting. Before each meeting, any potential conflicts of interest will be considered by the guideline committee Chair and a senior member of the development team. Any decisions to exclude a person from all or part of a meeting will be documented. Any changes to a member's declaration of interests will be recorded in the minutes of the meeting. Declarations of interests will be published with the final guideline.
Collaborators	Development of this systematic review will be overseen by an advisory committee who will use the review to inform the development of evidence-based recommendations in line with section 3 of Developing NICE guidelines: the manual. Members of the guideline committee

	are available on the NICE website: https://www.nice.org.uk/guidance/indevelopment/gid-ng10150/documents
Other registration details	N/A
Reference/URL for published protocol	https://www.crd.york.ac.uk/PROSPERO/display_record.php?RecordID=244449
Dissemination plans	NICE may use a range of different methods to raise awareness of the guideline. These include standard approaches such as: • notifying registered stakeholders of publication • publicising the guideline through NICE's newsletter and alerts • issuing a press release or briefing as appropriate, posting news articles on the NICE website, using social media channels, and publicising the guideline within NICE.
Keywords	Thyroid cancer
Details of existing review of same topic by same authors	N/A
Additional information	N/A
Details of final publication	www.nice.org.uk

1.1.13.2 Review protocol health economic evidence

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 2005, 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).[13] **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 2005 or later but that depend on unit costs and resource data entirely or predominantly from before 2005 will be rated as ‘Not applicable’.
- Studies published before 2005 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 these reviews are detailed below and complied with the methodology outlined in Developing NICE guidelines: the manual, 2014 (updated 2020) https://www.nice.org.uk/process/pmg20/chapter/identifying-the-evidence-literature-searching-and-evidence-submission.

For more information, please see the Methodology review published as part of the accompanying documents for this guideline.

Clinical literature search strategy

This literature search strategy was used for the following reviews:

- For people with fine-needle aspiration samples showing benign cytology or non-diagnostic features but with US / clinical findings indicative of increased risk of malignancy, is it clinically and cost effective to use diagnostic hemithyroidectomy, repeat FNAC, use active surveillance or discharge?

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 6: Database parameters, filters and limits applied

Database	Dates searched	Search filters and limits applied
Medline (OVID)	1946 – 13 January 2022	Randomised controlled trials Systematic review studies Observational studies Diagnostic studies Exclusions (animal studies, letters, comments, editorials, case studies/reports, children) English language
Embase (OVID)	1974 – 13 January 2022	Randomised controlled trials Systematic review studies Observational studies Diagnostic studies Exclusions (animal studies, letters, comments, editorials, case studies/reports,

Database	Dates searched	Search filters and limits applied
		conference abstracts, children) English language
The Cochrane Library (Wiley)	Cochrane Database of Systematic Reviews to Issue 12 of 12, December 2021 Cochrane Central Register of Controlled Trials to Issue 12 of 12, December 2021	Exclusions (clinical trials, conference abstracts)
Epistemonikos (The Epistemonikos Foundation)	Inception – 13 January 2022	Systematic review Exclusions (Cochrane reviews) English language

Medline (Ovid) search terms

1.	exp Thyroid Neoplasms/
2.	(thyroid adj3 (cancer* or carcinom* or microcarcinoma* or tumo?r* or neoplasm* or metast* or adenoma* or adenocarcinom* or node* or nodul* or nodal or lump* or papillar* or swollen or swell* or anaplastic or sarcoma* or cyst* or malignan*)).ti,ab.
3.	DTC.ti,ab.
4.	((papillar* or anaplastic) adj2 (cancer* or carcinom* or tumo?r* or neoplasm* or metast* or adenoma* or adenocarcinom* or nodul* or node* or lump*)).ti,ab.
5.	or/1-4
6.	letter/
7.	editorial/
8.	news/
9.	exp historical article/
10.	Anecdotes as Topic/
11.	comment/
12.	case report/
13.	(letter or comment*).ti.
14.	or/6-13
15.	randomized controlled trial/ or random*.ti,ab.
16.	14 not 15
17.	animals/ not humans/
18.	exp Animals, Laboratory/
19.	exp Animal Experimentation/
20.	exp Models, Animal/
21.	exp Rodentia/
22.	(rat or rats or mouse or mice or rodent*).ti.
23.	or/16-22
24.	5 not 23
25.	limit 24 to english language
26.	(exp child/ or exp pediatrics/ or exp infant/) not (exp adolescent/ or exp adult/ or exp middle age/ or exp aged/)

27.	25 not 26
28.	exp Biopsy, Needle/
29.	((needle or core or puncture) adj3 (aspirat* or biops* or cytology)).ti,ab.
30.	(FNAC or FNA or FNAB or FNB or FNC or CNB).ti,ab.
31.	or/28-30
32.	27 and 31
33.	randomized controlled trial.pt.
34.	controlled clinical trial.pt.
35.	randomi#ed.ab.
36.	placebo.ab.
37.	randomly.ab.
38.	clinical trials as topic.sh.
39.	trial.ti.
40.	or/33-39
41.	Meta-Analysis/
42.	Meta-Analysis as Topic/
43.	(meta analy* or metanaly* or metaanaly* or meta regression).ti,ab.
44.	((systematic* or evidence*) adj3 (review* or overview*)).ti,ab.
45.	(reference list* or bibliograph* or hand search* or manual search* or relevant journals).ab.
46.	(search strategy or search criteria or systematic search or study selection or data extraction).ab.
47.	(search* adj4 literature).ab.
48.	(medline or pubmed or cochrane or embase or psychlit or psyclit or psychinfo or psycinfo or cinahl or science citation index or bids or cancerlit).ab.
49.	cochrane.jw.
50.	((multiple treatment* or indirect or mixed) adj2 comparison*).ti,ab.
51.	or/41-50
52.	32 and (40 or 51)
53.	Epidemiologic studies/
54.	Observational study/
55.	exp Cohort studies/
56.	(cohort adj (study or studies or analys* or data)).ti,ab.
57.	((follow up or observational or uncontrolled or non randomi#ed or epidemiologic*) adj (study or studies or data)).ti,ab.
58.	((longitudinal or retrospective or prospective) and (study or studies or review or analys* or cohort* or data)).ti,ab.
59.	Controlled Before-After Studies/
60.	Historically Controlled Study/
61.	Interrupted Time Series Analysis/
62.	(before adj2 after adj2 (study or studies or data)).ti,ab.
63.	exp case control study/
64.	case control*.ti,ab.
65.	Cross-sectional studies/
66.	(cross sectional and (study or studies or review or analys* or cohort* or data)).ti,ab.
67.	or/53-66
68.	32 and 67

69.	68 not 52
70.	exp "sensitivity and specificity"/
71.	(sensitivity or specificity).ti,ab.
72.	((pre test or pretest or post test) adj probability).ti,ab.
73.	(predictive value* or PPV or NPV).ti,ab.
74.	likelihood ratio*.ti,ab.
75.	likelihood function/
76.	((area under adj4 curve) or AUC).ti,ab.
77.	(receive* operat* characteristic* or receive* operat* curve* or ROC curve*).ti,ab.
78.	(diagnos* adj3 (performance* or accurac* or utilit* or value* or efficien* or effectiveness)).ti,ab.
79.	gold standard.ab.
80.	exp Diagnostic errors/
81.	(false positiv* or false negativ*).tw.
82.	or/70-81
83.	32 and 82
84.	83 not (52 or 69)

Embase (Ovid) search terms

1.	exp Thyroid Cancer/
2.	(thyroid adj3 (cancer* or carcinom* or microcarcinoma* or tumo?r* or neoplasm* or metast* or adenoma* or adenocarcinom* or node* or nodul* or nodal or lump* or papillar* or swollen or swell* or anaplastic or sarcoma* or cyst* or malignan*)).ti,ab.
3.	DTC.ti,ab.
4.	((papillar* or anaplastic) adj2 (cancer* or carcinom* or tumo?r* or neoplasm* or metast* or adenoma* or adenocarcinom* or nodul* or node* or lump*)).ti,ab.
5.	or/1-4
6.	letter.pt. or letter/
7.	note.pt.
8.	editorial.pt.
9.	case report/ or case study/
10.	(letter or comment*).ti.
11.	(conference abstract or conference paper).pt.
12.	or/6-11
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 or rodent*).ti.
22.	or/14-21
23.	5 not 22
24.	limit 23 to english language
25.	(exp child/ or exp pediatrics/) not (exp adult/ or exp adolescent/)
26.	24 not 25

27.	exp Needle Biopsy/
28.	((needle or core or puncture) adj3 (aspirat* or biops* or cytology)).ti,ab.
29.	(FNAC or FNA or FNAB or FNB or FNC or CNB).ti,ab.
30.	or/27-29
31.	26 and 30
32.	random*.ti,ab.
33.	factorial*.ti,ab.
34.	(crossover* or cross over*).ti,ab.
35.	((doubl* or singl*) adj blind*).ti,ab.
36.	(assign* or allocat* or volunteer* or placebo*).ti,ab.
37.	crossover procedure/
38.	single blind procedure/
39.	randomized controlled trial/
40.	double blind procedure/
41.	or/32-40
42.	systematic review/
43.	Meta-Analysis/
44.	(meta analy* or metanaly* or metaanaly* or meta regression).ti,ab.
45.	((systematic* or evidence*) adj3 (review* or overview*)).ti,ab.
46.	(reference list* or bibliograph* or hand search* or manual search* or relevant journals).ab.
47.	(search strategy or search criteria or systematic search or study selection or data extraction).ab.
48.	(search* adj4 literature).ab.
49.	(medline or pubmed or cochrane or embase or psychlit or psyclit or psychinfo or psycinfo or cinahl or science citation index or bids or cancerlit).ab.
50.	cochrane.jw.
51.	((multiple treatment* or indirect or mixed) adj2 comparison*).ti,ab.
52.	or/42-51
53.	31 and (41 or 52)
54.	Clinical study/
55.	Observational study/
56.	family study/
57.	longitudinal study/
58.	retrospective study/
59.	prospective study/
60.	cohort analysis/
61.	follow-up/
62.	cohort*.ti,ab.
63.	61 and 62
64.	(cohort adj (study or studies or analys* or data)).ti,ab.
65.	((follow up or observational or uncontrolled or non randomi#ed or epidemiologic*) adj (study or studies or data)).ti,ab.
66.	((longitudinal or retrospective or prospective) and (study or studies or review or analys* or cohort* or data)).ti,ab.
67.	(before adj2 after adj2 (study or studies or data)).ti,ab.
68.	exp case control study/

69.	case control*.ti,ab.
70.	cross-sectional study/
71.	(cross sectional and (study or studies or review or analys* or cohort* or data)).ti,ab.
72.	or/54-60,63-71
73.	31 and 72
74.	73 not 53
75.	exp "sensitivity and specificity"/
76.	(sensitivity or specificity).ti,ab.
77.	((pre test or pretest or post test) adj probability).ti,ab.
78.	(predictive value* or PPV or NPV).ti,ab.
79.	likelihood ratio*.ti,ab.
80.	((area under adj4 curve) or AUC).ti,ab.
81.	(receive* operat* characteristic* or receive* operat* curve* or ROC curve*).ti,ab.
82.	diagnostic accuracy/
83.	diagnostic test accuracy study/
84.	gold standard.ab.
85.	exp diagnostic error/
86.	(false positiv* or false negativ*).ti,ab.
87.	differential diagnosis/
88.	(diagnos* adj3 (performance* or accurac* or utilit* or value* or efficien* or effectiveness or precision or validat* or validity or differential or error*)).ti,ab.
89.	or/75-88
90.	31 and 89
91.	90 not (53 or 74)

Cochrane Library (Wiley) search terms

#1.	MeSH descriptor: [Thyroid Neoplasms] explode all trees
#2.	(thyroid near/3 (cancer* or carcinom* or microcarcinoma* or tumo?r* or neoplasm* or metast* or adenoma* or adenocarcinom* or node* or nodul* or nodal or lump* or papillar* or swollen or swell* or anaplastic or sarcoma* or cyst* or malignan*)):ti,ab
#3.	DTC:ti,ab
#4.	((papillar* or anaplastic) near/2 (cancer* or carcinom* or tumo?r* or neoplasm* or metast* or adenoma* or adenocarcinom* or nodul* or node* or lump*)):ti,ab
#5.	#1 or #2 or #3 or #4
#6.	MeSH descriptor: [Biopsy, Needle] explode all trees
#7.	(needle or core or puncture) near/3 (aspirat* or biops* or cytology):ti,ab
#8.	(FNAC or FNA or FNAB or FNB or FNC or CNB):ti,ab
#9.	#6 or #7 or #8
#10.	#5 and #9
#11.	conference:pt or (clinicaltrials or trialsearch):so
#12.	#10 not #11

Epistemonikos search terms

1.	(title:((title:(thyroid AND (cancer* OR neoplasm* OR nodule* OR carcinoma*)) OR abstract:(thyroid AND (cancer* OR neoplasm* OR nodule* OR carcinoma*))) AND (title:(needle OR puncture OR biops* OR aspirat*) OR abstract:(needle OR puncture OR biops* OR aspirat*))) OR abstract:((title:(thyroid AND (cancer* OR neoplasm* OR nodule* OR carcinoma*)) OR abstract:(thyroid AND (cancer* OR neoplasm* OR

	nodule* OR carcinoma*))) AND (title:(needle OR puncture OR biops* OR aspirat*) OR abstract:(needle OR puncture OR biops* OR aspirat*))))

Health Economics literature search strategy

Health economic evidence was identified by conducting searches using terms for a broad Thyroid Cancer population. The following databases were searched: NHS Economic Evaluation Database (NHS EED - this ceased to be updated after 31st March 2015), Health Technology Assessment database (HTA - this ceased to be updated from 31st March 2018) and The International Network of Agencies for Health Technology Assessment (INAHTA). Searches for recent evidence were run on Medline and Embase from 2014 onwards for health economics, and all years for quality-of-life studies.

Table 2: Database parameters, filters and limits applied

Database	Dates searched	Search filters and limits applied
Medline (OVID)	Health Economics 1 January 2014 – 16 December 2021	Health economics studies Quality of life studies
	Quality of Life 1946 – 16 December 2021	Exclusions (animal studies, letters, comments, editorials, case studies/reports, conference abstracts) English language
Embase (OVID)	Health Economics 1 January 2014 – 16 December 2021	Health economics studies Quality of life studies
	Quality of Life 1974 – 16 December 2021	Exclusions (animal studies, letters, comments, editorials, case studies/reports, conference abstracts) English language
NHS Economic Evaluation Database (NHS EED) (Centre for Research and Dissemination - CRD)	Inception –31st March 2015	
Health Technology Assessment Database (HTA) (Centre for Research and Dissemination – CRD)	Inception – 31st March 2018	
The International Network of Agencies for Health Technology Assessment (INAHTA)	Inception - 16 December 2021	English language

Medline (Ovid) search terms

1.	exp Thyroid Neoplasms/
2.	(thyroid adj4 (cancer* or carcinom* or tumo?r* or neoplasm* or metast* or adenoma* or adenocarcinom* or nod* or lump* or papillar* or follicul* or lymphoma* or anaplastic)).ti,ab.

3.	((papillar* or follicul* or medullary or anaplastic) adj4 (cancer* or carcinom* or tumo?r* or neoplasm* or metast* or adenoma* or adenocarcinom* or nod* or lump* or lymphoma*)).ti,ab.
4.	or/1-3
5.	letter/
6.	editorial/
7.	news/
8.	exp historical article/
9.	Anecdotes as Topic/
10.	comment/
11.	case report/
12.	(letter or comment*).ti.
13.	or/5-12
14.	randomized controlled trial/ or random*.ti,ab.
15.	13 not 14
16.	animals/ not humans/
17.	exp Animals, Laboratory/
18.	exp Animal Experimentation/
19.	exp Models, Animal/
20.	exp Rodentia/
21.	(rat or rats or mouse or mice).ti.
22.	or/15-21
23.	4 not 22
24.	limit 23 to english language
25.	economics/
26.	value of life/
27.	exp "costs and cost analysis"/
28.	exp Economics, Hospital/
29.	exp Economics, medical/
30.	Economics, nursing/
31.	economics, pharmaceutical/
32.	exp "Fees and Charges"/
33.	exp budgets/
34.	budget*.ti,ab.
35.	cost*.ti.
36.	(economic* or pharmaco?economic*).ti.
37.	(price* or pricing*).ti,ab.
38.	(cost* adj2 (effectiv* or utilit* or benefit* or minimi* or unit* or estimat* or variable*)).ab.
39.	(financ* or fee or fees).ti,ab.
40.	(value adj2 (money or monetary)).ti,ab.
41.	or/25-40
42.	24 and 41
43.	quality-adjusted life years/
44.	sickness impact profile/
45.	(quality adj2 (wellbeing or well being)).ti,ab.
46.	sickness impact profile.ti,ab.

47.	disability adjusted life.ti,ab.
48.	(qal* or qtime* or qwb* or daly*).ti,ab.
49.	(euroqol* or eq5d* or eq 5*).ti,ab.
50.	(qol* or hql* or hqol* or h qol* or hrqol* or hr qol*).ti,ab.
51.	(health utility* or utility score* or disutilit* or utility value*).ti,ab.
52.	(hui or hui1 or hui2 or hui3).ti,ab.
53.	(health* year* equivalent* or hye or hyes).ti,ab.
54.	discrete choice*.ti,ab.
55.	rosser.ti,ab.
56.	(willingness to pay or time tradeoff or time trade off or tto or standard gamble*).ti,ab.
57.	(sf36* or sf 36* or short form 36* or shortform 36* or shortform36*).ti,ab.
58.	(sf20 or sf 20 or short form 20 or shortform 20 or shortform20).ti,ab.
59.	(sf12* or sf 12* or short form 12* or shortform 12* or shortform12*).ti,ab.
60.	(sf8* or sf 8* or short form 8* or shortform 8* or shortform8*).ti,ab.
61.	(sf6* or sf 6* or short form 6* or shortform 6* or shortform6*).ti,ab.
62.	or/52-70
63.	24 and 62

Embase (Ovid) search terms

1.	exp Thyroid Cancer/
2.	(thyroid adj4 (cancer* or carcinom* or tumo?r* or neoplasm* or metast* or adenoma* or adenocarcinom* or nod* or lump* or papillar* or follicul* or lymphoma* or anaplastic)).ti,ab.
3.	((papillar* or follicul* or medullary or anaplastic) adj4 (cancer* or carcinom* or tumo?r* or neoplasm* or metast* or adenoma* or adenocarcinom* or nod* or lump* or lymphoma*)).ti,ab.
4.	or/1-3
5.	letter.pt. or letter/
6.	note.pt.
7.	editorial.pt.
8.	case report/ or case study/
9.	(letter or comment*).ti.
10.	or/5-9
11.	randomized controlled trial/ or random*.ti,ab.
12.	10 not 11
13.	animal/ not human/
14.	nonhuman/
15.	exp Animal Experiment/
16.	exp Experimental Animal/
17.	animal model/
18.	exp Rodent/
19.	(rat or rats or mouse or mice).ti.
20.	or/12-19
21.	4 not 20
22.	limit 21 to english language
23.	health economics/
24.	exp economic evaluation/

25.	exp health care cost/
26.	exp fee/
27.	budget/
28.	funding/
29.	budget*.ti,ab.
30.	cost*.ti.
31.	(economic* or pharmaco?economic*).ti.
32.	(price* or pricing*).ti,ab.
33.	(cost* adj2 (effectiv* or utilit* or benefit* or minimi* or unit* or estimat* or variable*)).ab.
34.	(financ* or fee or fees).ti,ab.
35.	(value adj2 (money or monetary)).ti,ab.
36.	or/23-35
37.	22 and 36
38.	quality-adjusted life years/
39.	"quality of life index"/
40.	short form 12/ or short form 20/ or short form 36/ or short form 8/
41.	sickness impact profile/
42.	(quality adj2 (wellbeing or well being)).ti,ab.
43.	sickness impact profile.ti,ab.
44.	disability adjusted life.ti,ab.
45.	(qal* or qtime* or qwb* or daly*).ti,ab.
46.	(euroqol* or eq5d* or eq 5*).ti,ab.
47.	(qol* or hql* or hqol* or h qol* or hrqol* or hr qol*).ti,ab.
48.	(health utility* or utility score* or disutilit* or utility value*).ti,ab.
49.	(hui or hui1 or hui2 or hui3).ti,ab.
50.	(health* year* equivalent* or hye or hyes).ti,ab.
51.	discrete choice*.ti,ab.
52.	rosser.ti,ab.
53.	(willingness to pay or time tradeoff or time trade off or tto or standard gamble*).ti,ab.
54.	(sf36* or sf 36* or short form 36* or shortform 36* or shortform36*).ti,ab.
55.	(sf20 or sf 20 or short form 20 or shortform 20 or shortform20).ti,ab.
56.	(sf12* or sf 12* or short form 12* or shortform 12* or shortform12*).ti,ab.
57.	(sf8* or sf 8* or short form 8* or shortform 8* or shortform8*).ti,ab.
58.	(sf6* or sf 6* or short form 6* or shortform 6* or shortform6*).ti,ab.
59.	or/37-58
60.	22 and 59

NHS EED and HTA (CRD) search terms

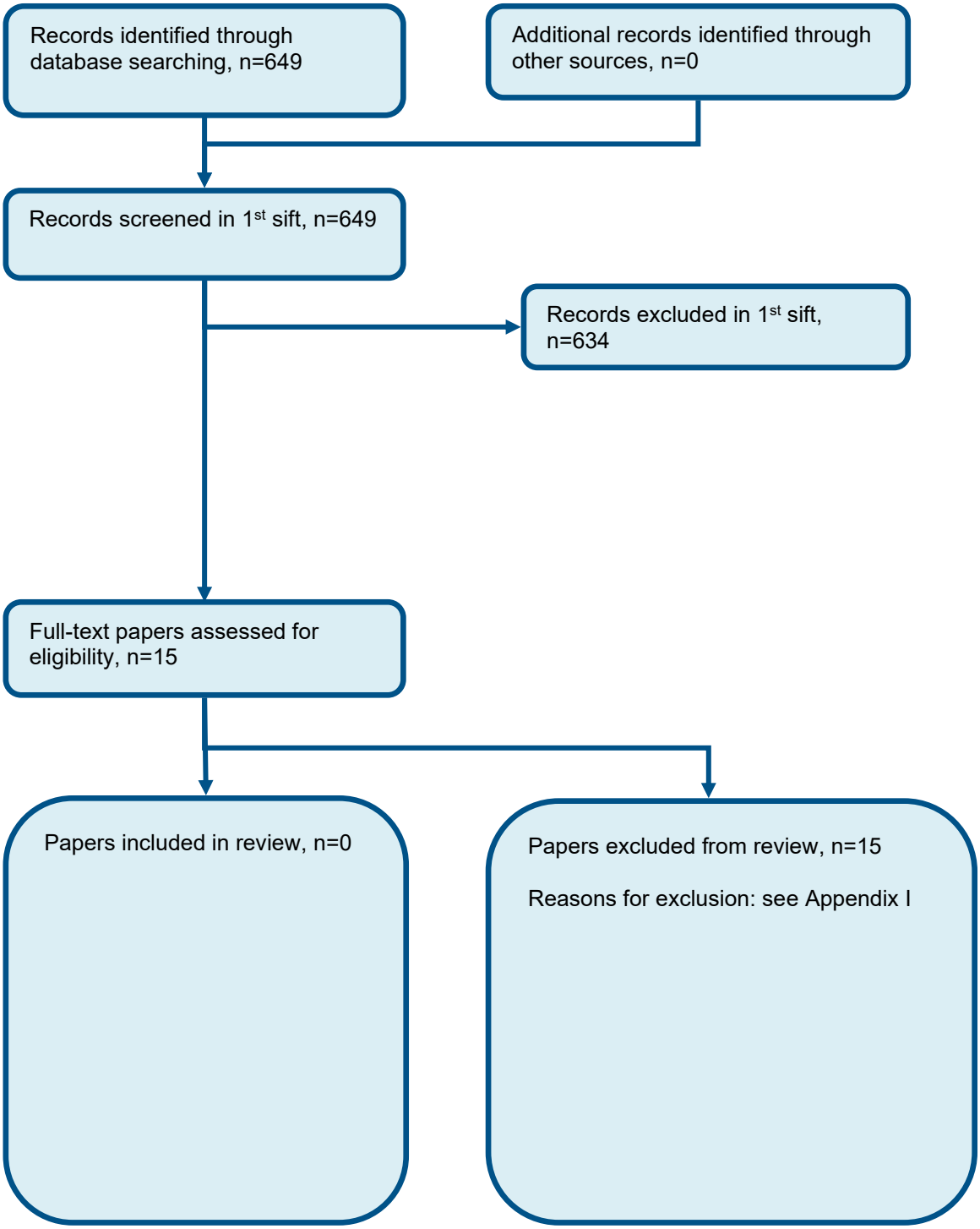
#1.	MeSH DESCRIPTOR Thyroid Neoplasms EXPLODE ALL TREES
#2.	((thyroid NEAR4 (cancer* or carcinom* or tumour* or tumor* or neoplasm* or metast* or adenoma* or adenocarcinom* or nod* or lump* or papillar* or follicul* or lymphoma* or anaplastic)))
#3.	(((papillar* or follicul* or medullary or anaplastic) NEAR4 (cancer* or carcinom* or tumour* or tumor* or neoplasm* or metast* or adenoma* or adenocarcinom* or nod* or lump* or lymphoma*)))
#4.	#1 OR #2 OR #3

INHATA search terms

1.	(Thyroid Neoplasms)[mh] OR (thyroid neoplasms) AND (thyroid cancers)

Appendix C – Effectiveness evidence study selection

Figure 1: Flow chart of clinical study selection for the review of management strategy in people with indeterminate FNA



Appendix D – Effectiveness evidence

No evidence was identified.

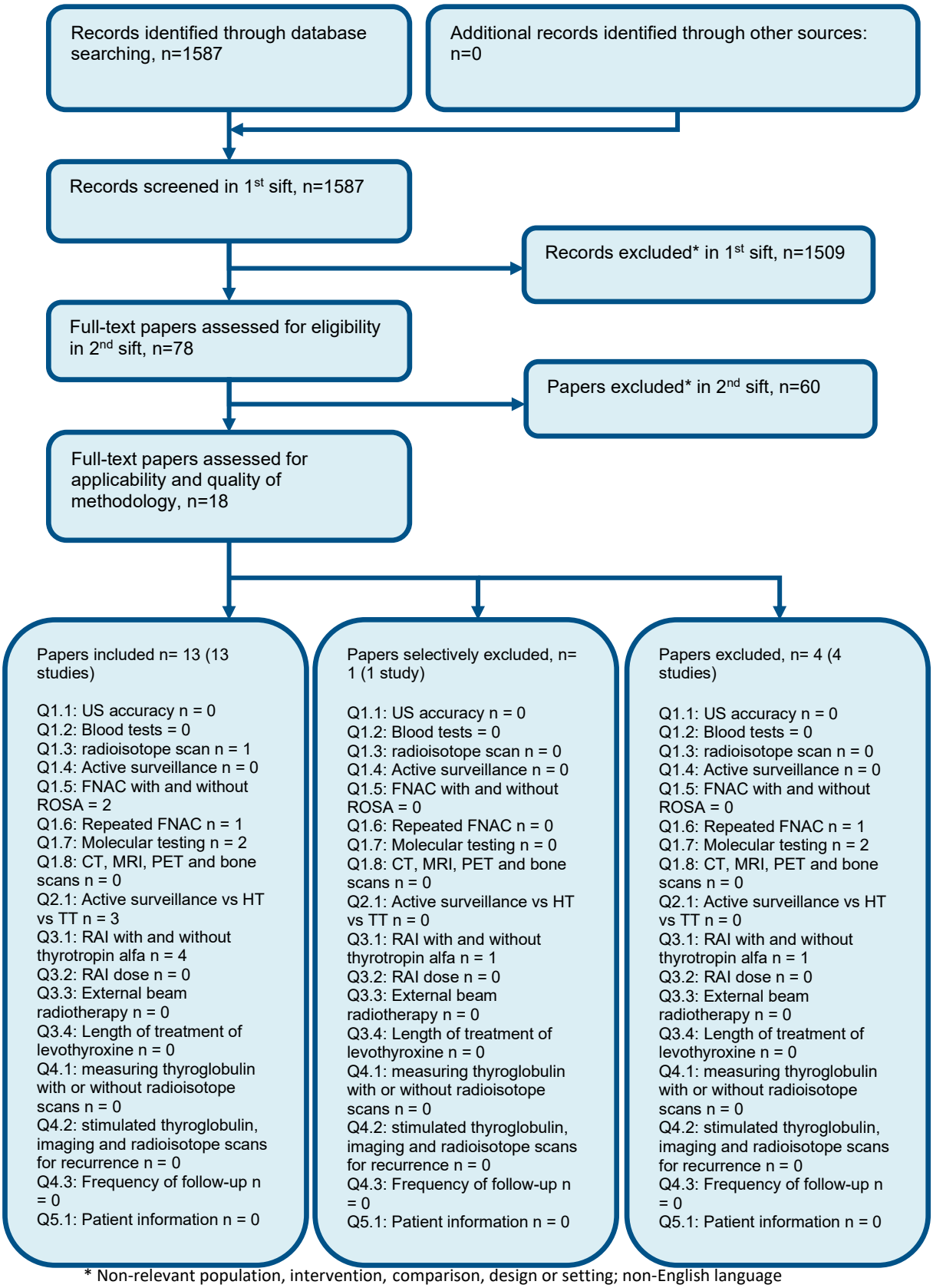
Appendix E – Forest plots

No evidence was identified.

Appendix F – GRADE and/or GRADE-CERQual tables

No evidence was identified.

Appendix G – Economic evidence study selection



* Non-relevant population, intervention, comparison, design or setting; non-English language

Appendix H – Economic evidence tables

Study	Furlan 2005[2]			
Study details	**Population & interventions**	**Costs**	**Health outcomes**	**Cost effectiveness**
Economic analysis: Cost consequence (health outcome: None) **Study design:** Retrospective chart review of patients who underwent surgery at one hospital between 1998 and 2000 **Approach to analysis:** FNA biopsy reports of 'suspicious for malignancy', 'follicular lesion', and 'cellular atypia' were considered positive and used to calculate accuracy, sensitivity, specificity, and costs. **Perspective:** Canadian healthcare system **Time horizon:** 2 years **Discounting:**	**Population:** Adults with suspected differentiated thyroid cancer and non-diagnostic or benign findings following FNA biopsy **Cohort settings:** Median age: 47 years Male: 16% N: 268 **Intervention 1:** Single FNA biopsy (includes the first cytological evaluation from individuals who later underwent sequential FNA biopsies. **Intervention 2:** Sequential FNA biopsy (at least 2 cytological aspirations from the same patient on different dates over ≥1 year)	**Total costs (mean per patient):** Intervention 1: £47 Intervention 2: £80 Incremental (2−1): £33 (95% CI: NR; p=NR) **Currency & cost year:** 2004 Canadian dollars (presented here as 2004 UK pounds[(a)]) **Cost components incorporated:** FNA biopsy	**Diagnostic outcomes (mean per group):** Intervention 1: • Accuracy: 63.8% • Sensitivity: 73.8% • Specificity: 69.0% Intervention 2: • Accuracy: 78.2% • Sensitivity: 84.0% • Specificity: 73.3% Incremental (2-1) • Accuracy: 14.4% • Sensitivity: 10.2% • Specificity: 4.3%	Sequential FNA biopsy was 70% more costly compared with single FNA biopsy and resulted in fewest number of false diagnoses. **Analysis of uncertainty:** Exclusion of all incidental microcarcinomas of 10mm or less found after thyroidectomy did not result in any significant differences in diagnostic accuracy (p=0.898); sensitivity (p=0.965); or specificity (0.594).

Costs: NR Outcomes: NR				
Data sources				
Health outcomes: Accuracy of FNA biopsy was defined as the ratio between the number of true results and number of patients undergoing aspiration. Sensitivity was defined as the ratio between the number of patients with a positive cytology result and cancer at histology (true-positive results), and the number of patients with carcinoma at definitive histopathological report (true-positive plus false-negative results).Specificity was defined as the ratio between the number of patients with a negative cytological diagnosis as well as no tumour at histology (true negative results) and the number of patients with no carcinoma at histology (true-negative plus false negative results). **Quality-of-life weights:** NA **Cost sources:** Healthcare costs were estimated using the Ontario Health Insurance Plan schedule of benefits and fees.				
Comments				
Source of funding: The Head and Neck Cancer Foundation **Limitations:** Canadian healthcare context. Discount rate was not reported for a study with a time horizon of 2 years. Patients were not randomized to treatment groups and had statistically significant differences in baseline patient characteristics between groups. Mortality, quality of life, and other patient outcomes were not included. Study is non-randomized, sample size is small and study was conducted at one hospital with unclear generalizability. Analysis did not include costs of surgery and other downstream costs and consequences of diagnosis. Cost year not specified; assumed to be the date of the published source used to obtain unit costs. Sensitivity analyses were not included. The study is relatively old. Accuracy of FNAC may have been improved during the last two decades **Other:** None				
Overall applicability:(b) Partially applicable **Overall quality:**(c) Potentially serious limitations				

Abbreviations: 95% CI= 95% confidence interval; da= deterministic analysis; EQ-5D= Euroqol 5 dimensions (scale: 0.0 [death] to 1.0 [full health], negative values mean worse than death); ICER= incremental cost-effectiveness ratio; NA = not applicable; NR= not reported; pa= probabilistic analysis; QALYs= quality-adjusted life years; rhTSH = recombinant human thyroid stimulating hormone

(a) Converted using 2004/05 purchasing power parities[17]

(b) Directly applicable / Partially applicable / Not applicable

(c) Minor limitations / Potentially serious limitations / Very serious limitations

Appendix I – Excluded studies

I.1 Clinical studies

Table 7: Studies excluded from the clinical review

Study	Exclusion reason
Castellana 2019[1]	Systematic review: study designs inappropriate
Gerhard 2015[3]	Systematic review: study designs inappropriate
Khoncarly 2014[8]	Incorrect study design: case series
Lee 2019[11]	Not review population. Inappropriate comparison
Pyo 2016[19]	Systematic review: study designs inappropriate
Qiu 2020[20]	Systematic review: study designs inappropriate. Incorrect interventions
Romitelli 2009[21]	Wrong comparison: fine needle aspiration versus fine needle non-aspiration
Straccia 2015[23]	Systematic review: study designs inappropriate
Suh 2016[24]	Systematic review: study designs inappropriate
Suh 2017[25]	Systematic review: study designs inappropriate
Trimboli 2015[28]	Systematic review: study designs inappropriate
Trimboli 2018[27]	Systematic review: study designs inappropriate
Valerio 2020[29]	Incorrect study design: retrospective observational study
Vriens 2011[31]	Systematic review: study designs inappropriate
Wang 2014[32]	Systematic review: study designs inappropriate

I.2 Health Economic studies

Published health economic studies that met the inclusion criteria (relevant population, comparators, economic study design, published 2005 or later and not from non-OECD country or USA) but that were excluded following appraisal of applicability and methodological quality are listed below. See the health economic protocol for more details.

Table 8: Studies excluded from the health economic review

Reference	Reason for exclusion
Van Roosmalen[4]	Excludes as rated with very serious limitations. Patients were not randomized to either treatment group and the analysis was not adjusted for confounding factors.