National Institute for Health and Care Excellence

Final

Thyroid cancer: assessment and management

[F] Evidence review for molecular testing

NICE guideline NG230

Evidence reviews underpinning recommendation 1.2.15 and the research recommendation on molecular testing 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 All NICE guidance is subject to regular review and may be updated or withdrawn.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: 978-1-4731-4865-9

Contents

1 Molecular testing

1.1 Review question

1.1.1 For people who have had fine-needle aspiration samples, where the FNAC sample was adequate but was not able to differentiate between benign and malignant samples, what is the clinical and cost effectiveness of molecular testing?

1.1.2 Introduction

Approximately one quarter of patients who undergo FNAC of their thyroid nodule, will receive an 'indeterminate' result i.e. one which is unable to give a clear benign or malignant result and which requires diagnostic surgery in most cases. Most of these patients will subsequently be found to have benign disease whereas up to 25% will be diagnosed with cancer and may then need further surgery. The sequencing of the somatic genomics of thyroid cancer in the Thyroid Cancer Gene Atlas and subsequent studies has permitted the identification of key abnormalities in DNA, RNA and mRNA associated with thyroid cancers. This information, along with the advent of 'next generation sequencing' which allows multiple DNA segments to be read simultaneously, has been utilised to attempt to reclassify cytologically indeterminate thyroid nodules as benign or malignant. This is known as 'molecular testing' and is offered by several commercial companies based in the USA. Its purpose is to avoid diagnostic surgery and allow optimal therapeutic surgery to be performed at the first operation. However, the clinical accuracy and cost effectiveness of this technology in a UK setting is unknown.

This review seeks to determine the effectiveness of molecular testing for people with indeterminate results on fine-needle aspiration.

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 with thyroid nodules with fine-needle aspiration or core biopsy samples where the sample was adequate but it was not possible to differentiate between benign and malignant disease.
Interventions	Any molecular testing after FNAC, whether commercial or non-commercial (the findings from each molecular test added to those of FNAC to give a fuller and more detailed picture). These will include, but not be restricted to: • ThyroSeq • Affirma • ThyGenNEXT/ThyraMiR • Rosetta GX-Reveal • Non proprietary / non commercial molecular tests with published data • NHS GLH panels: (M9.1) BRAF, H- K- N-RAS, (M9.2) RET, NTRK1/2+3 , (M9.3) RET, (M9.4) NTRK1/3+3
Comparisons	• Each other • FNAC alone without subsequent molecular testing.

Outcomes	All outcomes are considered equally important for decision making and therefore have all been rated as critical. • mortality • quality of life • disease progression
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 molecular testing after FNAC with each other or FNAC only 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

Twenty studies were identified for full text eligibility assessment, but none met the protocol criteria. For details, please 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

Two health economic studies with the relevant comparison was included in this review.[7, 18] This is summarised in the health economic evidence profile below (**Table 2**) and the health economic evidence tables in Appendix H.

1.1.8.2 Excluded studies

Two economic studies relating to this review question were identified but was excluded due to limited applicability.[1, 4] 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: Diagnostic lobectomy vs molecular test

Study	Applicability	Limitations	Other comments	Incremental cost	Incremental effects	Cost effectiveness	Uncertainty
Lee 2014 [7] (Canada)	Partially applicable [(a)]	Minor limitations [(b)]	• Microsimulation model • Cost-utility analysis (QALYs) • Population: Adults with low-risk thyroid nodules and indeterminate findings following two fine needle aspiration biopsy and cytology evaluations • Comparators: 1. Standard management 2. Routine GEC 3. Routine GEC followed by selective use of GMP 4. Routine GMP 5. Routine GMP followed by selective use of GEC • Time horizon: 1 year	2-1: £1,570[(c)] 3-2: -£204[(c)] 4-3: -£910[(c)] 5-4: £1,016[(c)]	2-1: 0.07 3-2: 0.05 4-3: -0.15 5-4: -0.04	Intervention 2 vs intervention 1: £22,428 Intervention 3 vs intervention 2: dominant Intervention 3 vs intervention 1: £13,649 Intervention 4: dominated Intervention 5: dominated	Probability Intervention 3 cost effective (£20/30k threshold): 22% Standard management has the highest probability of being cost effective at all thresholds. Uncertainty: Results were most sensitive to the cost of GEC and GPM, and variations in the probability of malignancy
Ronen 2021[18] (UK)	Directly applicable [(d)]	Potentially serious limitation [(e)]	• Decision tree model • Cost-comparison analysis • Population: Adults with indeterminate cytology after FNAC: Thy3a and	GSC molecular test costs £576[(f)] per person	NA	Intervention 1 is cost saving	Probability Intervention 3 cost effective (£20/30k threshold): NA Uncertainty:

Study	Applicability	Limitations	Other comments	Incremental cost	Incremental effects	Cost effectiveness	Uncertainty
			Thy3f. Only results for Thy3f were extracted. • Comparators: 1. Diagnostic hemithyroidectomy 2. GSC molecular test • Time horizon: 1 year				The threshold analysis showed that routine GSC tests would become cost effective in England at a unit price below £2,177

Abbreviations: DSA= deterministic sensitivity analysis; ICER= incremental cost-effectiveness ratio; GEC = gene expression classifier; GMP = gene mutation panel; GSC= gene sequencing classifier; NA= not applicable; NR= not reported; PSA= probabilistic sensitivity analysis; QALYs= quality-adjusted life years; RAI= radioactive iodine ablation; ROM= risk of malignancy.

(a) Canadian and US perspectives reported with Canadian perspective used as the base case in the evidence table. Method of utility valuation and utility weight tariff not reported.

(b) There was considerable uncertainty surrounding the results. Accuracy of tests based on old generation studies.

(c) 2013 Canadian dollars converted to 2013 UK pounds.[17]. Cost components incorporated: GEC and GMP, surgery, RAI adjuvant treatment, treatment of complications, TSH suppression, surveillance, benign nodules follow-up

(d) Hemithyroidectomy does not reflect current practice for Thy3a in England. Consequently, this analysis can be applied to Thy3f only.

(e) No PSA or DSA were conducted. The model structure does not allow to incorporate long-term consequences of missed diagnosis or avoidable surgeries. No attempt to establish a baseline ROM. Some costs were not included (e.g. RAI). The same cost of a GSC test was applied to all countries included in the analysis even though GSC is not commonly available in the UK and additional costs would need to be sustained such as storing and shipping the sample.

(f) 2020 American Dollars converted to 2020 UK Pounds using 2020 average exchange rate. Cost components incorporated: GSC, thyroidectomy costs (including surgical and hospital bed costs)

1.1.10 Economic model

A health economic model was developed to assess the cost-effectiveness of four different molecular testing in England after a Thy3a and Thy3f cytology. The results of the model can be viewed in evidence review E and the full report can be separately consulted in the “Cost-utility analysis: Most cost-effective diagnostic pathways for people with non-diagnostic or indeterminate cytology”.

1.1.11 Economic evidence statements

One cost-effectiveness analysis found routine GEC followed by selective use of GMP to be cost-effective compared to standard management or other molecular testing strategies in people with indeterminate nodules. The analysis was assessed as partially applicable with minor limitations.

One cost-consequence analysis fond hemithyroidectomy to be cost saving compared to GSC molecular test in people with Thy3f nodules. The analysis was assessed as partially applicable with potentially serious limitations.

One original cost-utility analysis found that:

- In people with Thy3a, routine use of molecular testing is dominated by repeat sampling with core need biopsy (CNB). If a further inconclusive result (Thy1 or Thy3a) is obtained through repeat sampling, selective use of molecular testing is dominated by selective use of hemithyroidectomy.
- In people with Thy3f, it is uncertain whether routine use of molecular testing with ThyroSeq V3 is cost-effective compared to routine use of hemithyroidectomy.

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

Protocol specified outcomes of mortality, quality of life (any validated tools) and disease progression were all deemed critical and were therefore of equal importance to decision making. One RCT was found but it was a diagnostic study comparing sensitivity and specificity of molecular tests. The committee were interested in whether the tests were of benefit on health outcomes rather than comparing diagnostic accuracy.

1.1.12.2 The quality of the evidence

The review found no evidence pertaining to the clinical effectiveness of molecular testing after FNAC compared to a different type of molecular testing after FNAC or FNAC without subsequent molecular testing in people with an adequate FNA or core biopsy sample but where it was not possible to differentiate between benign and malignant disease.

1.1.12.3 Benefits and harms

There was no clinical evidence found to inform a recommendation. The committee noted that this is an area of interest with considerable potential impact on future practice, but that there is a current lack of evidence to guide a consensus recommendation. Therefore, given the lack of any high quality randomised controlled trials in this area, the committee decided to make a research recommendation to address the question: "For people who have had fine-needle aspiration samples, where the FNAC sample was adequate but was not able to differentiate between benign and malignant samples, what is the clinical and cost effectiveness of molecular testing?"

There is health economic evidence to suggest molecular tests can be considered. The main discussion around this is in evidence report F on repeat testing. This evidence report discusses what to do following the initial FNAC and compares repeat FNAC testing, core needle biopsy, molecular testing, diagnostic hemithyroidectomy and active surveillance as part of a diagnostic economic model.

1.1.12.4 Cost effectiveness and resource use

Four health economics studies were fully assessed but only two were ultimately included for this research question as one had a wrong comparator (MIBI scintigraphy) whereas the other was rated not applicable as a large part of costs derived from productivity losses and healthcare costs could not be estimated separately.

One of the health economics study included was a cost-utility analysis comparing two different molecular tests, the Gene expression classifier (GEC) and gene mutation panel GMP, and different combinations of both, with Canadian current practice after an indeterminate FNAC (Thy3a and Thy3F), which is diagnostic hemithyroidectomy. The analysis was assessed as partially applicable, being conducted in Canada, and with minor limitations mostly deriving from the fact that old sources (2010/2011) were used to estimate costs and accuracy of molecular tests even though, in recent years, new tests have been introduced in the market and technological improvements have increased overall accuracy. The deterministic analysis found GEC followed by a selective use of GMP to be the most cost-effective strategy with a ICER below NICE threshold of £20,000, however the authors found large uncertainty. The probabilistic analysis found that standard management (diagnostic lobectomy) has the higher probability of being more cost effective at every threshold.

A second cost-comparison analysis was conducted from an UK NHS perspective and it compared the more recent GSC molecular test with hemithyroidectomy. As GSC molecular testing are the most commonly used being more advanced than previous generation GEC test, this analysis was considered directly applicable. However, the comparator hemithyroidectomy does not reflect current practice after Thy3a in England, as Thy3a cytologies are commonly followed up with a repeat FNAC and, if necessary, and Multidisciplinary Team (MDT) discussion. Consequently, the analysis was assessed to reflect only Thy3f with a risk of malignancy (ROM) of around 30%. The analysis found that, at a cancer prevalence of 30%, GSC molecular test was more expensive than diagnostic hemithyroidectomy with an additional cost per patient of £494. A threshold analysis on the price of a GSC test found that routine test would become cost effective in England only with a price below £2,177.

The committee acknowledged that cost-effectiveness of molecular testing was very dependent on the setting where the analysis was conducted. In countries where thyroid cancer management is expensive, like the US, molecular testing is likely to be very cost effective whereas in health care systems where cancer management is cheaper, like Canada or the UK, its cost effectiveness is more uncertain. The committee recognized that molecular testing for people with indeterminate FNAC (Thy3) may potentially be cost effective in England if it is able to reduce unnecessary surgeries in people with benign nodules, although it was noted that the price set by manufactures reflects the high cost of the US healthcare system and may not meet cost effectiveness criteria in other countries. The committee agreed that a health economics study on molecular testing in England would help defining prices that would make the test cost effective, hence possibly encouraging manufactures to negotiate lower price to have access to the British market. Hence, the plan to conduct an UK cost-utility analysis on the most cost-effective diagnostic pathway after an indeterminate FNAC (Thy3a and Thy3f) and looking at all molecular tests available in the market was discussed and approved by the committee.

The results of the model indicate that CNB and selective use of molecular testing is not cost-effective compared to CNB and selective use of hemithyroidectomy. The threshold analysis shows that the price of the test should be below £1,200, which is higher than the estimated price of the cheapest test, ThyroSeq V3. In people with Thy3f, it is uncertain whether routine use of molecular testing with ThyroSeq V3 is cost-effective compared to routine use of hemithyroidectomy due to the high uncertainty associated with the first strategy. The committee were aware that the health economic evidence was not strong enough to support

a recommendation changing current practice in this cytology. A short description of the model and the full discussion are reported in Evidence Review E. The full economic report can be consulted in a separated document.

1.1.12.5 Other factors the committee took into account

The committee were aware that molecular testing is a new and rapidly developing area. While these tests are not readily available in the UK they anticipate that there value will

1.1.13 Recommendations supported by this evidence review

This evidence review supports recommendation 1.2.15 and the research recommendation on molecular tests.

References

1. Aydemirli MD, Snel M, van Wezel T, Ruano D, Obbink CMH, van den Hout WB et al. Yield and costs of molecular diagnostics on thyroid cytology slides in the Netherlands, adapting the Bethesda classification. Endocrinology, and Diabetes & Metabolism. 2021; 4(4):e00293

2. Borowczyk M, Szczepanek-Parulska E, Olejarz M, Wieckowska B, Verburg FA, Debicki S et al. Correction to: Evaluation of 167 gene expression classifier (gec) and thyroseq v2 diagnostic accuracy in the preoperative assessment of indeterminate thyroid nodules: Bivariate/hroc meta-analysis (endocrine pathology, (2018), 10.1007/s12022-018-9560-5). Endocrine Pathology. 2019;

3. Borowczyk M, Szczepanek-Parulska E, Olejarz M, Wieckowska B, Verburg FA, Debicki S et al. Evaluation of 167 gene expression classifier (gec) and thyroseq v2 diagnostic accuracy in the preoperative assessment of indeterminate thyroid nodules: Bivariate/hroc meta-analysis. Endocrine Pathology. 2019; 30(1):8-15

4. Heinzel A, Muller D, Behrendt FF, Giovanella L, Mottaghy FM, Verburg FA. Thyroid nodules with indeterminate cytology: molecular imaging with 99mTc-methoxyisobutylisonitrile (MIBI) is more cost-effective than the Afirma gene expression classifier. European Journal of Nuclear Medicine and Molecular Imaging. 2014; 41(8):1497-1500

5. Hitscherich KJ, Hernandez M, Tiesi G, Hanlon A, Sienko D. Detecting thyroid cancer: Comparative analysis of Afirma genomic sequence classifier versus thyroseq v3 genomic classifier. Annals of Surgical Oncology. 2020; 27:S91-S92

6. Kebebew E, Duh QY, Clark OH. Total thyroidectomy or thyroid lobectomy in patients with low-risk differentiated thyroid cancer: surgical decision analysis of a controversy using a mathematical model. World Journal of Surgery. 2000; 24(11):1295-1302

7. Lee L, How J, Tabah RJ, Mitmaker EJ. Cost-effectiveness of molecular testing for thyroid nodules with atypia of undetermined significance cytology. Journal of Clinical Endocrinology and Metabolism. 2014; 99(8):2674-2682

8. Livhits MJ, Kuo EJ, Leung AM, Rao J, Levin M, Douek ML et al. Gene expression classifier vs targeted next-generation sequencing in the management of indeterminate thyroid nodules. Journal of Clinical Endocrinology and Metabolism. 2018; 103(6):2261-2268

9. Livhits MJ, Zhu CY, Kuo EJ, Nguyen DT, Kim J, Tseng CH et al. Effectiveness of molecular testing techniques for diagnosis of indeterminate thyroid nodules: A randomized clinical trial. JAMA oncology. 2021; 7(1):70-77

10. Mernagh P, Suebwongpat A, Silverberg J, Weston A. Cost-effectiveness of using recombinant human thyroid-stimulating hormone before radioiodine ablation for thyroid cancer: the Canadian perspective. Value in Health. 2010; 13(2):180-187

11. Najafzadeh M, Marra CA, Lynd LD, Wiseman SM. Cost-effectiveness of using a molecular diagnostic test to improve preoperative diagnosis of thyroid cancer. Value in Health. 2012; 15(8):1005-1012

12. 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

13. Ngo HTT, Nguyen TPX, Vu TH, Jung CK, Hassell L, Kakudo K et al. Impact of molecular testing on the management of indeterminate thyroid nodules among western and asian countries: A systematic review and meta-analysis. Endocrine Pathology. 2021; 32(2):269-279

14. Nicholson KJ, Yip L. An update on the status of molecular testing for the indeterminate thyroid nodule and risk stratification of differentiated thyroid cancer. Current Opinion in Oncology. 2018; 30(1):8-15

15. Nishino M, Nikiforova M. Update on molecular testing for cytologically indeterminate thyroid nodules. Archives of Pathology and Laboratory Medicine. 2018; 142(4):446-457

16. Noureldine SI, Najafian A, Han PA, Olson MT, Genther DJ, Schneider EB et al. Evaluation of the effect of diagnostic molecular testing on the surgical decision-making process for patients with thyroid nodules. JAMA Otolaryngology - Head and Neck Surgery. 2016; 142(7):676-682

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. Ronen O, Oichman M. National differences in cost analysis of Afirma Genomic sequencing classifier. Clinical Endocrinology. 2021; 94(4):717-724

19. Rossi ED, Pantanowitz L, Faquin WC. The role of molecular testing for the indeterminate thyroid FNA. Genes. 2019; 10(10):736

20. Schumm MA, Nguyen DT, Kim J, Tseng CH, Chow AY, Shen N et al. Longitudinal assessment of quality of life following molecular testing for indeterminate thyroid nodules. Annals of Surgical Oncology. 2021; 28(13):8872-8881

21. Sciacchitano S, Lavra L, Ulivieri A, Magi F, De Francesco GP, Bellotti C et al. Comparative analysis of diagnostic performance, feasibility and cost of different test-methods for thyroid nodules with indeterminate cytology. Oncotarget. 2017; 8(30):49421-49442

22. Vargas-Salas S, Martinez JR, Urra S, Dominguez JM, Mena N, Uslar T et al. Genetic testing for indeterminate thyroid cytology: review and meta-analysis. Endocrine-Related Cancer. 2018; 25(3):R163-R177

23. Vuong HG, Nguyen TPX, Hassell LA, Jung CK. Diagnostic performances of the afirma gene sequencing classifier in comparison with the gene expression classifier: A meta-analysis. Cancer Cytopathology. 2021; 129(3):182-189

24. Wang Z, Chen JQ, Liu JL, Qin XG. Clinical impact of BRAF mutation on the diagnosis and prognosis of papillary thyroid carcinoma: a systematic review and meta-analysis. European Journal of Clinical Investigation. 2016; 46(2):146-157

25. Yip L, Sosa JA. Molecular-directed treatment of differentiated thyroid cancer: Advances in diagnosis and treatment. JAMA Surgery. 2016; 151(7):663-670

26. Zhang X. Value of molecular tests in cytologically indeterminate lesions of thyroid. Archives of Pathology and Laboratory Medicine. 2015; 139(12):1484-1490

27. Zhu CY, Donangelo I, Gupta D, Nguyen DT, Ochoa JE, Yeh MW et al. Outcomes of indeterminate thyroid nodules managed nonoperatively after molecular testing. Journal of Clinical Endocrinology and Metabolism. 2021; 106(3):e1240-e1247

Appendices

Appendix A – Review protocols

A.1 Review protocol for molecular testing

Field	Content
PROSPERO registration number	Not registered
Review title	The clinical and cost effectiveness of molecular testing for people with fine-needle aspiration samples where the FNAC sample was adequate but was not able to differentiate between benign and malignant samples.
Review question	For people who have had fine-needle aspiration samples, where the FNAC sample was adequate but was not able to differentiate between benign and malignant samples, what is the clinical and cost effectiveness of molecular testing?
Objective	To determine the best management strategy for people with indeterminate results on fine-needle aspiration samples leaving the possibility of differentiated thyroid cancer
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

Field	Content
	• 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 with thyroid nodules with fine-needle aspiration or core biopsy samples where the sample was adequate but it was not possible to differentiate between benign and malignant disease. Exclusion: Children and young people under 16 years
Intervention/Exposure/Test	Any molecular testing after FNAC, whether commercial or non-commercial (the findings from each molecular test added to those of FNAC to give a fuller and more detailed picture). These will include, but not be restricted to:

Field	Content
	• ThyroSeq • Affirma • ThyGenNEXT/ThyraMiR • Rosetta GX-Reveal • Non proprietary / non commercial molecular tests with published data • NHS GLH panels: (M9.1) BRAF, H- K- N-RAS, (M9.2) RET, NTRK1/2+3 , (M9.3) RET, (M9.4) NTRK1/3+3
Comparator/Reference standard/Confounding factors	• Each other • FNAC alone without subsequent molecular testing. Therefore, comparisons will be one test versus another test, or each test versus usual care
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.
Context	FNAC may sometimes be unable to differentiate between benign and malignant samples and in such a case it has been suggested that molecular testing may be a useful method of clarification. It is important to assess how useful molecular testing is, and whether it is a cost effective addition to evaluation.
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

Field	Content
	• disease progression 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. A standardised form will be used to extract data from studies (see Developing NICE guidelines: the manual section 6.4). 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 through discussion, with involvement of a third review author where necessary.
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:

Field	Content
	• 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.

Field	Content
Analysis of sub-groups	Stratification None Sub-grouping If serious or very serious heterogeneity (I2>50%) is present within any stratum, sub-grouping will occur according to the following strategies: • US findings: benign / indeterminate / suspicious • Gender (male v female) • Age (<55 vs ≥55)
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

Field	Content
	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

Field	Content
Other registration details	N/A
Reference/URL for published protocol	N/A
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	
Details of existing review of same topic by same authors	N/A
Additional information	N/A
Details of final publication	www.nice.org.uk

A.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).[12] **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 review:

- For people who have had fine-needle aspiration samples, where the FNAC sample was adequate but was not able to differentiate between benign and malignant samples, what is the clinical and cost effectiveness of molecular testing to diagnose or rule out thyroid cancer?

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 3: 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 Exclusions (animal studies, letters, comments, editorials, case studies/reports, children) English language
Embase (OVID)	1974 – 13 January 2022	Randomised controlled trials Systematic review studies Exclusions (animal studies, letters, comments, editorials, case studies/reports, 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	Inception – 13 January 2022	Systematic review

Database	Dates searched	Search filters and limits applied
(The Epistemonikos Foundation)		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.	Molecular Diagnostic Techniques/
29.	Genetic Testing/
30.	Pharmacogenomic testing/
31.	(ThyroSeq or ThyGenX or ThyGeNEXT or ThyraMIR or Rosetta* or Afirma or ThyroSPEC or ThyroPrint or miRInform or AmpliSeq or mir-THYpe).ti,ab.
32.	((molecular or mutation* or gene or genetic or genom* or multigene* or pharmacogen*) adj3 (test* or diagnos* or technique* or analys* or marker* or biomarker* or profil* or assay* or panel* or classifi* or expression or sequenc*)).ti,ab.
33.	(molecular adj (approach* or genetic* or characteri#ation or alteration* or signature* or abnormalit* or cytology or pathology or cytopathology)).ti,ab.

34.	((next generation or massive* parallel or high throughput) adj2 sequenc*).ti,ab.
35.	((DNA or RAS or KRAS or HRAS or BRAF* or B-RAF* or TERT or protooncogene* or oncogene*) adj5 mutat*).ti,ab.
36.	(RNA* or mRNA* or miRNA* or microRNA*).ti,ab.
37.	or/28-36
38.	27 and 37
39.	randomized controlled trial.pt.
40.	controlled clinical trial.pt.
41.	randomi#ed.ab.
42.	placebo.ab.
43.	randomly.ab.
44.	clinical trials as topic.sh.
45.	trial.ti.
46.	or/39-45
47.	Meta-Analysis/
48.	Meta-Analysis as Topic/
49.	(meta analy* or metanaly* or metaanaly* or meta regression).ti,ab.
50.	((systematic* or evidence*) adj3 (review* or overview*)).ti,ab.
51.	(reference list* or bibliograph* or hand search* or manual search* or relevant journals).ab.
52.	(search strategy or search criteria or systematic search or study selection or data extraction).ab.
53.	(search* adj4 literature).ab.
54.	(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.
55.	cochrane.jw.
56.	((multiple treatment* or indirect or mixed) adj2 comparison*).ti,ab.
57.	or/47-56
58.	38 and (46 or 57)

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.	Molecular diagnosis/
28.	Genetic screening/
29.	Pharmacogenetic testing/
30.	(ThyroSeq or ThyGenX or ThyGeNEXT or ThyraMIR or Rosetta* or Afirma or ThyroSPEC or ThyroPrint or miRInform or AmpliSeq or mir-THYpe).ti,ab.
31.	((molecular or mutation* or gene or genetic or genom* or multigene* or pharmacogen*) adj3 (test* or diagnos* or technique* or analys* or marker* or biomarker* or profil* or assay* or panel* or classifi* or expression or sequenc*)).ti,ab.
32.	(molecular adj (approach* or genetic* or characteri#ation or alteration* or signature* or abnormalit* or cytology or pathology or cytopathology)).ti,ab.
33.	((next generation or massive* parallel or high throughput) adj2 sequenc*).ti,ab.
34.	((DNA or RAS or KRAS or HRAS or BRAF* or B-RAF* or TERT or protooncogene* or oncogene*) adj5 mutat*).ti,ab.
35.	(RNA* or mRNA* or miRNA* or microRNA*).ti,ab.
36.	or/27-35
37.	26 and 36
38.	random*.ti,ab.
39.	factorial*.ti,ab.
40.	(crossover* or cross over*).ti,ab.
41.	((doubl* or singl*) adj blind*).ti,ab.
42.	(assign* or allocat* or volunteer* or placebo*).ti,ab.
43.	crossover procedure/
44.	single blind procedure/
45.	randomized controlled trial/
46.	double blind procedure/
47.	or/38-46
48.	systematic review/
49.	Meta-Analysis/
50.	(meta analy* or metanaly* or metaanaly* or meta regression).ti,ab.
51.	((systematic* or evidence*) adj3 (review* or overview*)).ti,ab.
52.	(reference list* or bibliograph* or hand search* or manual search* or relevant journals).ab.
53.	(search strategy or search criteria or systematic search or study selection or data extraction).ab.
54.	(search* adj4 literature).ab.
55.	(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.
56.	cochrane.jw.

57.	((multiple treatment* or indirect or mixed) adj2 comparison*).ti,ab.
58.	or/48-57
59.	37 and (47 or 58)

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: [Molecular Diagnostic Techniques] this term only
#7.	MeSH descriptor: [Genetic Testing] this term only
#8.	MeSH descriptor: [Pharmacogenomic Testing] this term only
#9.	(ThyroSeq or ThyGenX or ThyGeNEXT or ThyraMIR or Rosetta* or Afirma or ThyroSPEC or ThyroPrint or miRInform or AmpliSeq or mir-THYpe):ti,ab
#10.	((molecular or mutation* or gene or genetic or genom* or multigene* or pharmacogen*) near/3 (test* or diagnos* or technique* or analys* or marker* or biomarker* or profil* or assay* or panel* or classifi* or expression or sequenc*)):ti,ab
#11.	(molecular next (approach* or genetic* or characteri?ation or alteration* or signature* or abnormalit* or cytology or pathology or cytopathology)):ti,ab
#12.	((generation or massive* parallel or high throughput) near/2 sequenc*):ti,ab
#13.	((DNA or RAS or KRAS or HRAS or BRAF* or B-RAF* or TERT or protooncogene* or oncogene*) near/5 mutat*):ti,ab
#14.	(RNA* or mRNA* or miRNA* or microRNA*):ti,ab
#15.	#6 or #7 or #8 or #9 or #10 or #11 or #12 or #13 or #14
#16.	#5 and #15
#17.	conference:pt or (clinicaltrials or trialsearch):so
#18.	#16 not #17

Epistemonikos search terms

1.	(title:(thyroid AND cancer* OR neoplasm* OR nodule* OR carcinoma*) OR abstract:(thyroid AND cancer* OR neoplasm* OR nodule* OR carcinoma*)) AND (title:(indeterminate OR nondiagnostic OR molecular OR mutation* OR ThyroSeq OR ThyGenX OR ThyGeNEXT OR ThyraMIR OR Rosetta* OR Afirma OR ThyroSPEC OR ThyroPrint OR miRInform OR AmpliSeq OR mir-THYpe OR RAS OR KRAS OR HRAS OR BRAF* OR B-RAF* OR TERT OR RNA* OR mRNA* OR miRNA* OR microRNA*) OR abstract:(indeterminate OR nondiagnostic OR molecular OR mutation* OR ThyroSeq OR ThyGenX OR ThyGeNEXT OR ThyraMIR OR Rosetta* OR Afirma OR ThyroSPEC OR ThyroPrint OR miRInform OR AmpliSeq OR mir-THYpe OR RAS OR KRAS OR HRAS OR BRAF* OR B-RAF* OR TERT OR RNA* OR mRNA* OR miRNA* OR microRNA*))

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

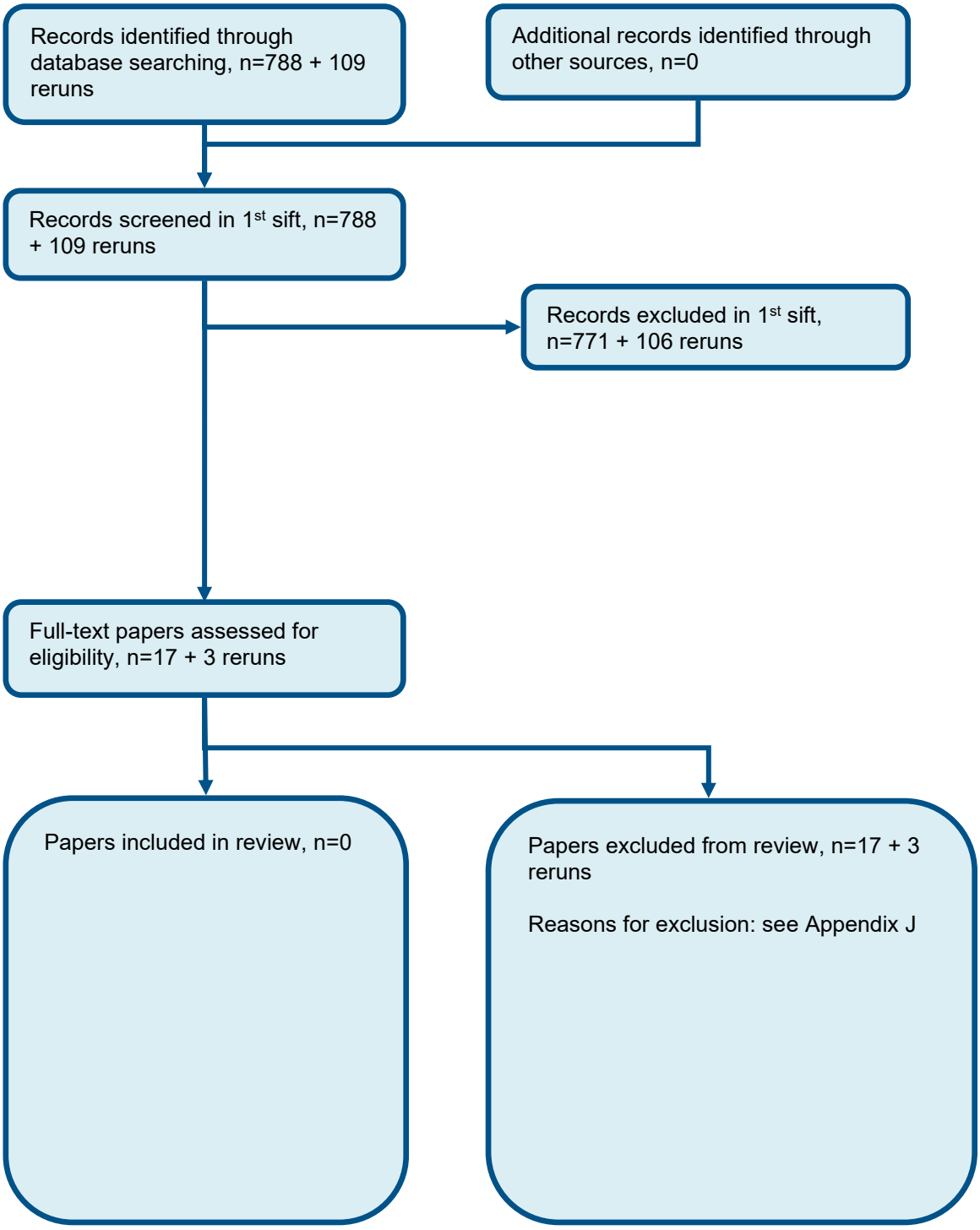
#1.	MeSH DESCRIPTOR Thyroid Neoplasms EXPLODE ALL TREES
#2.	((thyroid NEAR4 (cancer* or carcinom* or tumour* or tumor* or neoplasm* or metast* or adenoma* or adenocarcinom* or nod* or lump* or papillar* or follicul* or lymphoma* or anaplastic)))
#3.	(((papillar* or follicul* or medullary or anaplastic)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



Appendix D – Effectiveness evidence

No evidence was identified.

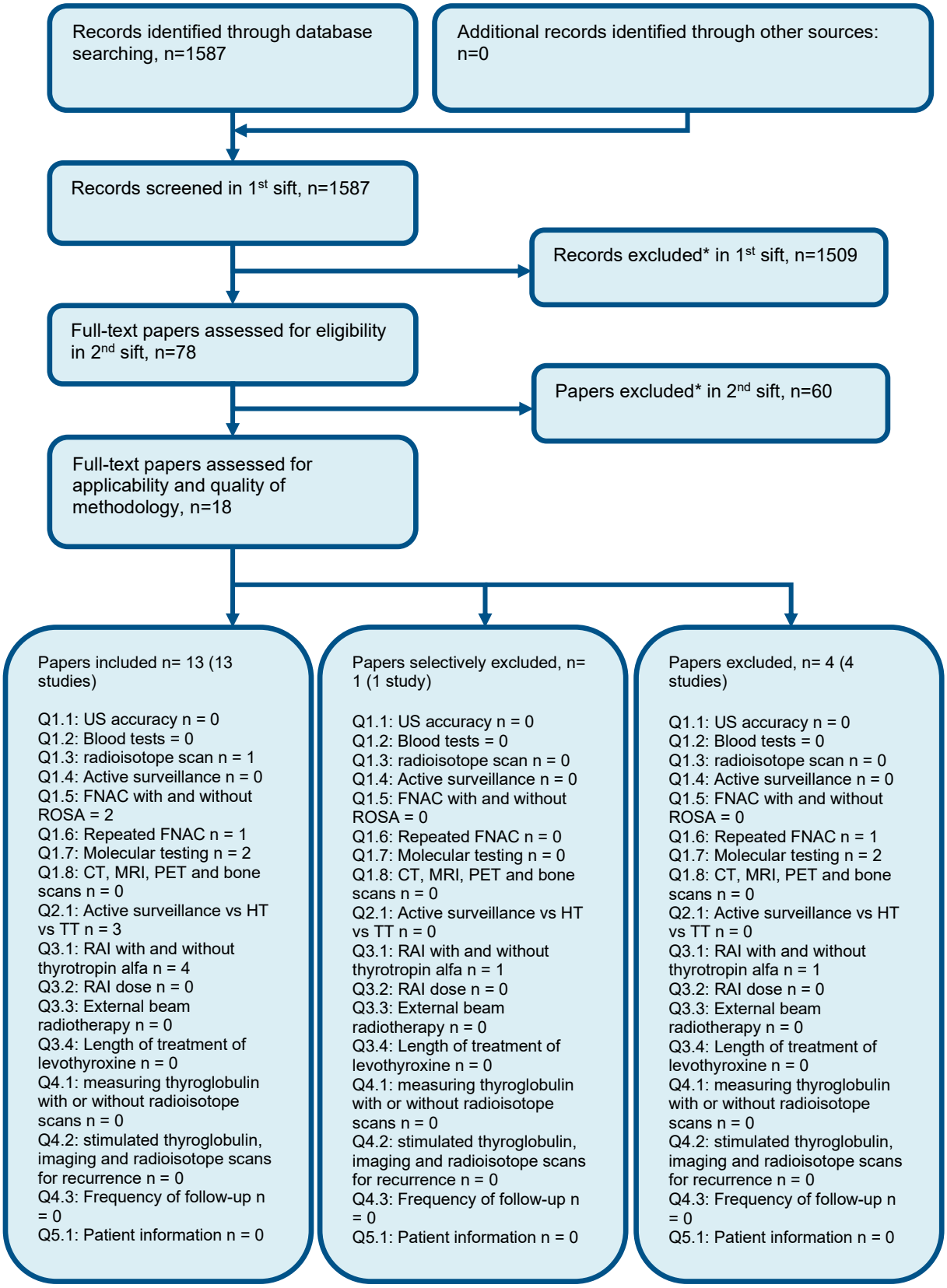
Appendix E – Forest plots

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	Lee 2014[7]			
Study details	**Population & interventions**	**Costs**	**Health outcomes**	**Cost effectiveness**
Economic analysis: CUA (health outcome: QALYs) **Study design:** Patient microsimulation model **Approach to analysis:** Microsimulation model with 1 million patients and 4 health states (clinical follow-up, surveillance, recurrence, death) **Perspective:** Canadian healthcare perspective with Quebec as the reference province **Time horizon:** 1 year **Discounting:** Costs: 3% Outcomes: 3%	**Population:** Adults with low-risk thyroid nodules and indeterminate findings following two fine needle aspiration biopsy and cytology evaluations **Cohort settings:** Median age: 54 years Male: 19% **Intervention 1:** Standard management **Intervention 2:** Routine GEC **Intervention 3:** Routine GEC followed by selective use of GMP **Intervention 4:** Routine GMP	**Total costs (mean per patient):** Intervention 1: £6,277 Intervention 2: £7,847 Intervention 3: £7,641 Intervention 4: £6,732 Intervention 5: £7,748 Incremental 2-1: £1,570 (95% CI: £1,529 to £1,610; p=NR) Incremental (3−2): -£204 (95% CI: -£264 to £146; p=NR) Incremental (4−3): -£910 (95% CI: -£1,014 to -£805; p=NR) Incremental (5−4): £1,016 (95% CI: £934 to £1,098 NR; p=NR) **Currency & cost year:** 2013 Canadian dollars (presented here as 2013 UK pounds[(a)])	**QALYs (mean per patient):** Intervention 1: 17.18 Intervention 2: 17.25 Intervention 3: 17.28 Intervention 4: 17.02 Intervention 5: 17.12 Incremental (2−1): 0.07 (95% CI: -0.04, 0.19; p=NR) Incremental (3−2): 0.05 (95% CI: -0.03, 0.14; p=NR) Incremental (4−3): -0.15 (95% CI: -0.26, -0.04; p=NR) Incremental (5−4): -0.04 (95% CI: -0.15, 0.07; p=NR)	**ICER:** Intervention 2 vs intervention 1: £22,428 Intervention 3 vs intervention 2: dominant Intervention 3 vs intervention 1: £13,649 Intervention 4: dominated Intervention 5: dominated Probability that Intervention 3 was cost effective (£20k/30k threshold): 22% Standard management has the highest probability of being cost effective at all thresholds. **Analysis of uncertainty:** Results were most sensitive to the cost of GEC and GPM, and variations in the probability of malignancy

	Intervention 5: Routine GMP followed by selective use of GEC	**Cost components incorporated:** GEC and GMP, surgery, RAI adjuvant treatment, treatment of complications, TSH suppression, surveillance, benign nodules follow-up		
Data sources				
Health outcomes: Relative risk of mortality obtained from Najafzdeh 2012[11] and Mazzaferri 1994. Relative risk of recurrence with delayed diagnosis obtained from yeh 2004. Probability of malignancy in AUS lesions obtained from Wang 2011. Annual probability of recurrence after TT and RAI obtained from Bilimoria 2007. Probability of mortality with recurrent disease obtained from Rouxel 2004. Probability of complications after thyroid surgery obtained from Landerholm 2014 and Giordano 2012. Diagnostic test characteristics obtained from Alexander 2012 and Nikiforov 2011. **Quality-of-life weights:** Health state utility values for unilateral RLN palsy, bilateral RLN palsy, Hypoparathyroidism and Recurrence were obtained from Kebebew 2000[6]. Disease-free after HT, Disease-free after TT + RAI were obtained from Esnaola 2001. Pre-RAI (post-surgery) and Post-RAI 0-4 weeks were obtained from a study by Mernagh 2010[10]. **Cost sources:** The cost of GEC and GMP were obtained from the manufacturers. Overall costs of thyroid lobectomy, total thyroidectomy, completion thyroidectomy, neck dissection procedures, including the operation and subsequent hospitalisation, were obtained for the province of Quebec from the Canadian Institute for Health Information 2013 case mix groups and Régie de l'Assurance Maladie du Québec 2013. The cost of radioactive iodine ablation, annual surveillance, and follow-up was obtained from the Régie de l'Assurance Maladie du Québec 2013 and Ministère de la Santé et des Services Sociaux 2013. Other costs and physician fees were obtained from the Régie de l'Assurance Maladie du Québec 2013 and Ministère de la Santé et des Services Sociaux 2013. Medication costs were based on Quebec provincial wholesale costs from the Ministère de la Santé et des Services Sociaux 2013.			
Comments				
Source of funding: NR **Limitations:** Canadian and US perspectives reported with Canadian perspective used as the base case in the evidence table. Accuracy of tests based on 2011 and 2012 data where old generation tests were used. Method of utility valuation and utility weight tariff not reported. There was considerable uncertainty surrounding the results which was explored with limited sensitivity analyses that were difficult to interpret. **Other:** None				
Overall applicability:(b) Partially applicable **Overall quality:**(c) Minor limitations				

Abbreviations: 95% CI= 95% confidence interval; AUS= atypia of undetermined significance; GEC= gene expression classifier; GMP = gene mutation testing; ICER= incremental cost-effectiveness ratio; NA= not applicable; NR= not reported; pa= probabilistic analysis; QALYs= quality-adjusted life years; RAI= Radioactive Iodine Ablation.
(a) Converted using 2013/14 purchasing power parities[17]
(b) Directly applicable / Partially applicable / Not applicable
(c) Minor limitations / Potentially serious limitations / Very serious limitations

Study	**Ronen 2021[18]**

Study details	Population & interventions	Costs	Health outcomes	Cost effectiveness
Economic analysis: Cost comparison analysis **Study design:** Decision tree model **Approach to analysis:** A decision tree model was developed to calculate the cost of GSC testing of indeterminate thyroid nodules in different countries. Two threshold analyses on prevalence of cancer and GSC cost were conducted **Perspective:** UK NHS **Time horizon:** 1 year **Discounting:** Costs: NR Outcomes: NR	**Population:** Adults with indeterminate cytology after FNAC: Thy3a and Thy3f. Only results of Thy3f were extracted as hemithyroidectomy is current practice for Thy3f only in the UK **Cohort settings:** Median age: NR Male: NR **Intervention 1** Diagnostic hemithyroidectomy **Intervention 2:** GSC molecular test	**Total costs (mean per patient):** With prevalence 30% Thy3f: Intervention 1: £5,198 Intervention 2: £5,774 Incremental 2-1: £576 **Currency & cost year:** 2020 American dollars (presented here as 2020 UK pounds[(a)]) **Cost components incorporated:** GSC, thyroidectomy costs (including surgical and hospital bed costs)	NA	Hemithyroidectomy is cost saving compared to GSC molecular test at a prevalence of 30% (Thy3f) **Analysis of uncertainty:** The threshold analysis showed that routine GSC tests would become cost effective in England at a price below £2,177
Data sources				
Health outcomes: NA. **Quality-of-life weights:** NA. **Cost sources:** UK healthcare costs were collected from National UK Tariff whereas the cost of GSC test from published literature.				
Comments				
Source of funding: NR **Limitations:** Except the two threshold analyses, a probabilistic or deterministic sensitivity analysis was not conducted. The comparator (hemithyroidectomy) does not reflect current practice for Thy3a in England as BTA guidelines requires Thy3a to receive a repeat FNAC first,				

and therefore the analysis can be applied to Thy3f only. The structure of the model does not allow to incorporate long-term consequences (e.g. complication of surgery) which could affect an impact on cost-effectiveness conclusion. There was no attempt to establish a baseline ROM despite there are multiple sources and meta-analyses available. Some relevant costs (e.g. RAI and post-surgery management) appear to be missing from the total costs. The same GSC cost was applied to all the countries of the analysis even though GSC is not usually available in the UK and would require additional costs, such as packing and shipping the sample abroad.

Other: None

Overall applicability:[a] Directly applicable **Overall quality:**[b] Potentially serious limitations

Abbreviations: 95% CI= 95% confidence interval; FNAC= Fine needle aspiration cytology; GSC= gene sequencing classifier; NA = not applicable; NR= not reported; RAI= Radioactive Iodine Ablation; ROM= Risk of malignancy.

(a) Converted using 2020 average exchange rate.

(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 4: Studies excluded from the clinical review

Study	Exclusion reason
Borowczyk 2019[3]	Systematic review: study designs inappropriate
Borowczyk 2019[2]	Systematic review: study designs inappropriate
Hitscherich 2020[5]	Abstract only
Livhits 2018[8]	No relevant outcomes
Livhits, 2021[9]	Randomised trial comparing molecular tests but no relevant outcomes
Ngo 2021[13]	Systematic review: study designs inappropriate
Nicholson 2018[14]	Incorrect study design (narrative review)
Nishino 2018[15]	Incorrect study design (narrative review)
Noureldine 2016[16]	Incorrect study design (case control)
Rossi 2019[19]	Incorrect study design (narrative review)
Schumm, 2021[20]	Did not compare outcomes between molecular tests
Sciacchitano 2017[21]	Incorrect study design (narrative review)
Vargas-salas 2018[22]	Systematic review: study designs inappropriate
Vuong 2021[23]	Systematic review: study designs inappropriate
Wang 2016[24]	Systematic review: study designs inappropriate. Incorrect population
Yip 2016[25]	Systematic review: study designs inappropriate
Zhang 2015[26]	Incorrect study design (narrative review)
Zhu, 2021 [27]	Did not compare outcomes between molecular tests

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 5: Studies excluded from the health economic review

Reference	Reason for exclusion
Heinzel 2014[4]	Excludes as rated not applicable. The comparator of the study, MIBI thyroid scintigraphy, is an off-label use and rarely performed in the UK so cannot be used to assess the cost effectiveness of molecular testing.
Aidemirly 2021[1]	Excluded as rated not applicable with very serious limitation. The analysis did not distinguish between healthcare and productivity costs that were presented together. As the latter were presumably very high, it is hard to rely on the conclusions of the analysis.

Appendix J – Research recommendations – full details

J.1.1 Research recommendation

In fine-needle aspiration cytology (FNAC) samples that are adequate but indeterminate (Thy3a and Thy3f), what is the clinical and cost effectiveness of molecular testing?

J.1.2 Why this is important

There are currently no randomised controlled trials comparing clinical outcomes when molecular tests are used compared to usual care in people who have adequate but indeterminate FNAC results. Diagnostic accuracy data are available, suggesting that molecular tests can accurately differentiate between malignant and benign nodules in this patient group, but the downstream effects on patient reported outcomes such as quality of life are unknown. Such data would facilitate health economic modelling and possibly support the case for the introduction of a potentially useful but expensive test.

J.1.3 Rationale for research recommendation

Importance to 'patients' or the population	Knowledge of any beneficial effects of molecular tests on quality of life may allow the introduction of molecular tests into standard care if these benefits can be shown to be cost effective. The possibility that this research recommendation could lead to the introduction of a diagnostic strategy that can improve quality of life is clearly of importance to patients.
Relevance to NICE guidance	The efficacy of molecular tests has been considered in this guideline, but we did not find any RCTs. The development of such RCTs is therefore required.
Relevance to the NHS	Molecular tests have the potential to improve accuracy of diagnosis, and reduce both the number of people who have a missed diagnosis and those that undergo unnecessary surgery. This will lead to better health outcomes for more patients.
National priorities	None known
Current evidence base	There is currently no RCT evidence. Some diagnostic evidence, which is outside the remit of the guideline scope, suggests that molecular tests have sufficiently high sensitivity and specificity to enable them to improve diagnosis, and they therefore have the potential, awaiting formal testing, to improve downstream outcomes for people with thyroid cancer.
Equality considerations	None known

J.1.4 Modified PICO table

Population	People who have had fine-needle aspiration samples, where the FNAC sample was adequate but was not able to differentiate between benign and malignant samples
Intervention	Molecular testing (with appropriate stratification for different types of test according to discretion of research team)
Comparator	Usual care
Outcome	Quality of life, progression, mortality
Study design	RCT
Timeframe	Long term
Additional information	This will be a randomised diagnostic trial, where diagnostic accuracy will not be directly evaluated. Instead, the indirect effects of accurate determination of nodule status on downstream patient reported health outcomes will be evaluated.