National Institute for Health and Care Excellence

Final

Thyroid Cancer

Cost-utility analysis: Most cost-effective diagnostic pathways for people with a non-diagnostic or indeterminate cytology

NICE guideline NG230

Economic analysis report

December 2022

Final

This analysis was developed by the National Guideline Centre

NICE accredited
www.nice.org.uk/accreditation

Disclaimer

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

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

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 All NICE guidance is subject to regular review and may be updated or withdrawn.

Copyright
© NICE 2022. All rights reserved. Subject to Notice of rights.

ISBN: 978-1-4731-4915-1

Contents

1 Introduction

Fine-needle aspiration cytology (FNAC) is the preferred first-line test to detect malignancy in nodules at initial presentation. Through an US-guide FNAC, a very fine needle is inserted into the lump or nodule using an ultrasound scan to guide the procedure and a sample of cells are extracted. The sample is then examined under the microscope to determine whether the cells are cancerous or benign[47] using the Royal College of Pathology classification system[7].

In most cases, the results are benign, and people are usually discharged. In other cases, when RCPath score is either Thy4 or Thy5, the nodule is considered either suspicious of malignancy or malignant, and a hemithyroidectomy or a total thyroidectomy are generally recommended.[37]

However, some Thy classifications are harder to interpret. Thy1, for instance, is defined as being "non-diagnostic for cytological diagnosis"[7] and occurs when the cells cannot be clearly visualised and technically evaluated. In this case, The British Thyroid Association (BTA) guideline recommends repeating a second US guided FNAC [37]

In other cases, the sample collected is sufficient for the evaluation, but the nature of the lesions cannot be determined solely by the FNA[7]. This is true for Thy3a and Thy3f that are generally referred as "indeterminate cytology", with Thy3f being more suggestive of follicular neoplasm than Thy3a. Current practice to manage Thy3a and Thy3f nodules vary across countries, and the BTA guidelines recommend a repeat FNAC for Thy3a and a hemithyroidectomy for Thy3f. It is worth mentioning that a repeat FNAC after an indeterminate cytology often give a further indeterminate result[24] which requires either a multi-disciplinary team (MDT) discussion or a diagnostic hemithyroidectomy to decide further management. Moreover, the rate of malignancy (ROM) of Thy3a and Thy3f are, respectively, 25% and 31% according to the latest meta-analysis conducted by the Royal College of Pathologists[38], implying that most of the hemithyroidectomies conducted on Thy3f are unnecessary as performed on people with benign nodule. Although a hemithyroidectomy is a relatively cheap and easy procedure, it can cause serious complications that can affect people quality of life (QoL) in the long-term.

The management of people with Thy1 and Thy3 have fuelled an ongoing debate around the world as different countries have opted for different management strategies. Although the terminology varies in the literature, in the course of this analysis a Thy1 cytology will be defined as "non-diagnostic", a Thy3a or Thy3f cytology as "indeterminate" whereas Thy1, Thy3a and Thy3f together are all "inadequate" cytologies. Sometimes Thy1 and Thy3a are referred in the literature as "inconclusive". Please see Appendix A: for a graphic explanation of the terminologies used in this analysis.

Several alternatives to deal with an inadequate cytology have been proposed. A Core-Needle Biopsy (CNB) is generally regarded as more precise than a FNAC[44 40, 45] as a second-line test as it is able to collect a larger sample from the nodule. However, it is more expensive than a FNAC and its cost effectiveness after a Thy1 or a Thy3a cytology is unknown.

Molecular diagnostic techniques have recently aroused great interest in the US as a non-invasive and very accurate tool to determine the malignancy of indeterminate nodules. There are several tests in the market, with the new generation of test known as GSC (gene sequencing classifier), and each use a particular gene expression to screen for genetic mutations of the nodules. Molecular testing has been widely used in the US and showed to be cost-effective[20] in that particular setting, where management of thyroid cancer is regarded as very expensive. However, in the UK, thyroid cancer management is considerably less expensive as, for instance, a hemithyroidectomy in E costs one third of the price in the US.

© NICE 2022. All rights reserved. Subject to Notice of rights.

No health economics analysis assessing molecular testing in the UK has been conducted so far, so its cost effectiveness in the NHS setting is unknown.

Published literature found that around 30% of FNA samples result in indeterminate cytology[6], and an additional 5 to 10% are non-diagnostic in the UK depending on the centre[39]. Therefore, any recommendation on further management of inadequate cytologies is expected to require a large use of NHS resource, thus strongly justifying the need of developing a health economic model. This analysis will look at the cost effectiveness of providing different diagnostic alternatives to people with an inadequate cytology after a FNAC in England, and it will be first one, to our knowledge, to look at molecular tests and core-needle biopsy as alternatives for hemithyroidectomy and repeat FNAC in the NHS setting. Three strategies will be assessed for Thy1: core-needle biopsy, repeat FNAC (which is current practice) and hemithyroidectomy. For Thy3a the following strategies are included in the analysis: molecular testing only, CNB and selective use of hemithyroidectomy or molecular testing, repeat FNAC and selective use of hemithyroidectomy (current practice for Thy3a) or molecular testing and diagnostic hemithyroidectomy. Finally, for Thy3f, the limited availability of data on repeat cytology and concerns that repeat sampling with either CNB or FNAC is particularly ineffective in this cytology allowed to include only molecular testing and hemithyroidectomy.

© NICE 2022. All rights reserved. Subject to Notice of rights.

2 Methods

2.1 Model overview

A cost-utility analysis was undertaken where lifetime quality-adjusted life years (QALYs) and costs from a current UK NHS and personal social services perspective were considered. The analysis followed the standard assumptions of the NICE reference case for interventions with health outcomes in an NHS setting including discounting at 3.5% for costs and health effects. An incremental analysis was undertaken.

2.1.1 Comparators

The following comparators were included in the analysis:

1. Routine use of molecular testing
   a. Afirma-GSC
   b. ThyroSeq V3
   c. ThyroSeq V1
   d. ThyGenX/ThyraMIR
2. Core-needle biopsy (CNB)
   a. With selective use of hemithyroidectomy after further indeterminate cytology
   b. With selective use of molecular testing after further indeterminate cytology
3. Repeat Fine-needle aspiration cytology (FNAC)
   a. With selective use of hemithyroidectomy after further indeterminate cytology
   b. With selective use of molecular testing after further indeterminate cytology
4. Routine use of diagnostic hemithyroidectomy

As there is no clinical reason for people with Thy1 to undergo a molecular test, only CNB with hemithyroidectomy (2a), FNAC with hemithyroidectomy (3a) and diagnostic hemithyroidectomy (4) were included as comparators for this subpopulation.

The committee were aware that repeat sampling with either CNB or FNAC after Thy3f was rarely useful and most of the data on the accuracy of repeat sampling were exclusively on Thy3a nodules. Hence, only strategies 1 and 4 were included for this subpopulation.

The following table summarizes the strategies included in each cohort analysis.

Table 1: Strategies included for each subgroup

	Thy1	Thy3a	Thy3f
Routine use of molecular testing	X	√	√
CNB & selective use hemithyroidectomy	√	√	X
CNB & selective use of molecular testing	X	√	X
FNAC & selective us of hemithyroidectomy	√	√	X
FNAC & selective use of molecular testing	X	√	X
Routine use of hemithyroidectomy	√	√	√

© NICE 2022. All rights reserved. Subject to Notice of rights.

The current practice in the UK for Thy1 and Thy3a is to repeat FNAC and for Thy3f to offer diagnostic hemithyroidectomy. The most recent BTA guidelines recommend a MDT discussion in case of a new Thy3a on repeat sample. MDT discussion could not be included in the analysis due to the lack of data on its effectiveness and accuracy in determining the malignancy of thyroid nodules and this limitation will be further discussed (see section 4.2).

2.1.2 Population

The population of the analysis was adults with thyroid nodule on ultrasound at initial presentation who underwent fine-needle aspiration cytology and received an inadequate result. The population was stratified according to the RCPath Thy score for non-diagnostic and indeterminate[7]:

1. Thy1: non-diagnostic
2. Thy3a: indeterminate
3. Thy3f: indeterminate

No other further stratification was deemed necessary by the committee.

2.2 Approach to modelling

The model consists of two decision trees for the outcomes of the diagnosis and surgery and a follow-up Markov model to estimate long-term health outcomes (QALYs) and cost of each strategy. The main outcomes of each strategy is the number of false positives and false negatives obtained with each diagnostic alternative. These have long-term consequences as false positive lead to an unnecessary hemithyroidectomy whereas false negative lead to a delayed diagnosis of malignancy. The follow-up Markov model was utilized to estimate the consequences of unnecessary hemithyroidectomies and delayed diagnosis on mortality, quality of life and public healthcare cost. This allows us to calculate costs and QALYs of each strategy and to rank all the strategies in order of cost effectiveness.

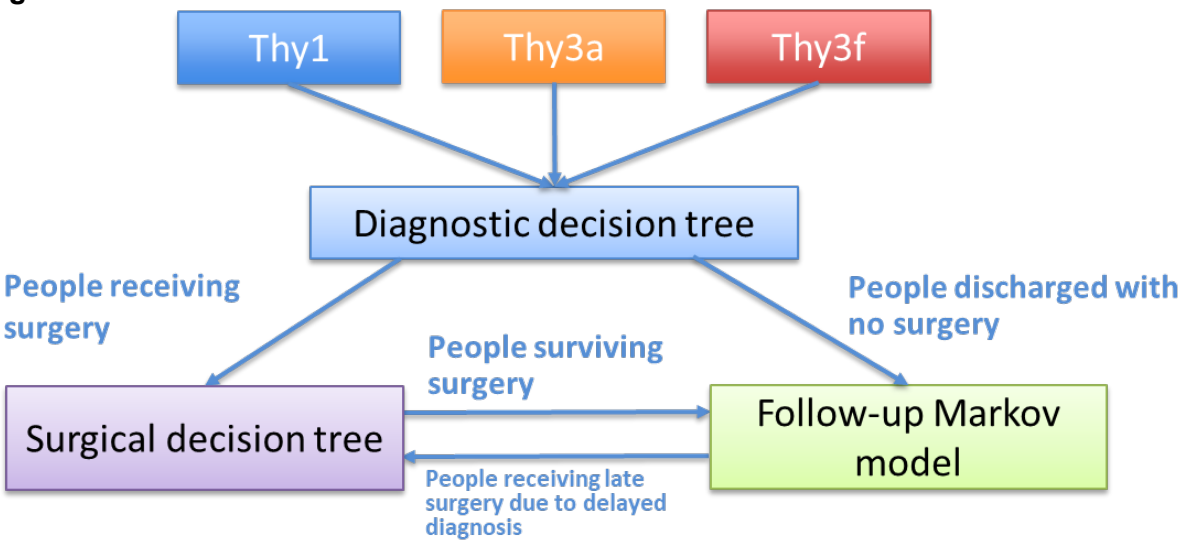
Further details of the model structure are described in section 2.2.1. To account for uncertainty, a probabilistic analysis was undertaken (see section 2.2.2).

2.2.1 Model structure

The overall model structure is described in Figure 1. People with Thy1, Thy3a or Thy3f enter the diagnostic decision tree where they receive the corresponding test. The Diagnostic decision tree incorporates both sensitivity and specificity of each testing strategy and ultimately determines how many people are discharged and how many are referred to a further diagnostic hemithyroidectomy. These latter will enter a second decision tree, the Surgical decision tree, where they will receive surgery, which is a hemithyroidectomy if benign, or a hemithyroidectomy followed by a completion thyroidectomy if malignant. This second decision tree is utilized to estimate the outcomes and costs of the surgery. At the end of the second decision tree, people surviving surgery and those who were discharged from the diagnostic decision tree enter the third and final model: The Follow-Up Markov model. Those with a malignant nodule who were mistakenly diagnosed as benign will move at some point to the surgical tree where they receive late surgery. This Markov model calculates lifelong costs, quality of life and mortality of people based on the outcomes estimated by the two previous decision trees. Quality of life and mortality are combined to obtain QALYs that, together with total costs, are used to determine the cost-effectiveness of each strategy.

© NICE 2022. All rights reserved. Subject to Notice of rights.

Figure 1: Full model structure

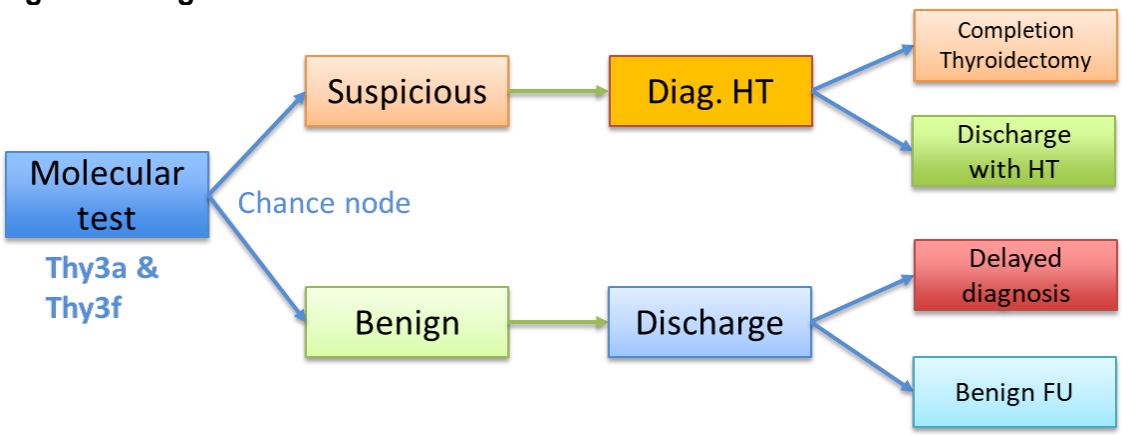


2.2.1.1 Diagnostic decision tree

The diagnostic decision tree was used to determine the outcomes of each diagnostic strategy and incorporates both the cost of each test and their accuracy.

Figure 2 illustrates the structure of the Diagnostic decision tree for a molecular test. Molecular tests are binary tests and therefore can give either a suspicious or benign result. This will depend both on the accuracy of the test as well as on the prevalence of cancer among the population considered. People with suspicious results are assumed to receive a diagnostic hemithyroidectomy whereas people with benign results are discharged. After the diagnostic hemithyroidectomy, all people with a malignant nodule will receive a completion thyroidectomy in the base case scenario, whereas people with benign nodule will be discharged. In the sensitivity analysis different proportions of people with malignancy needing a completion surgery after the first hemithyroidectomy are tested (see section 2.4). Those who were discharged can be either benign, if the diagnosis turns out to be correct, or delayed diagnosis, if the result of the molecular test was a false negative. This will have an implication on their costs and rate of recurrence.

Figure 2: Diagnostic decision tree – molecular test

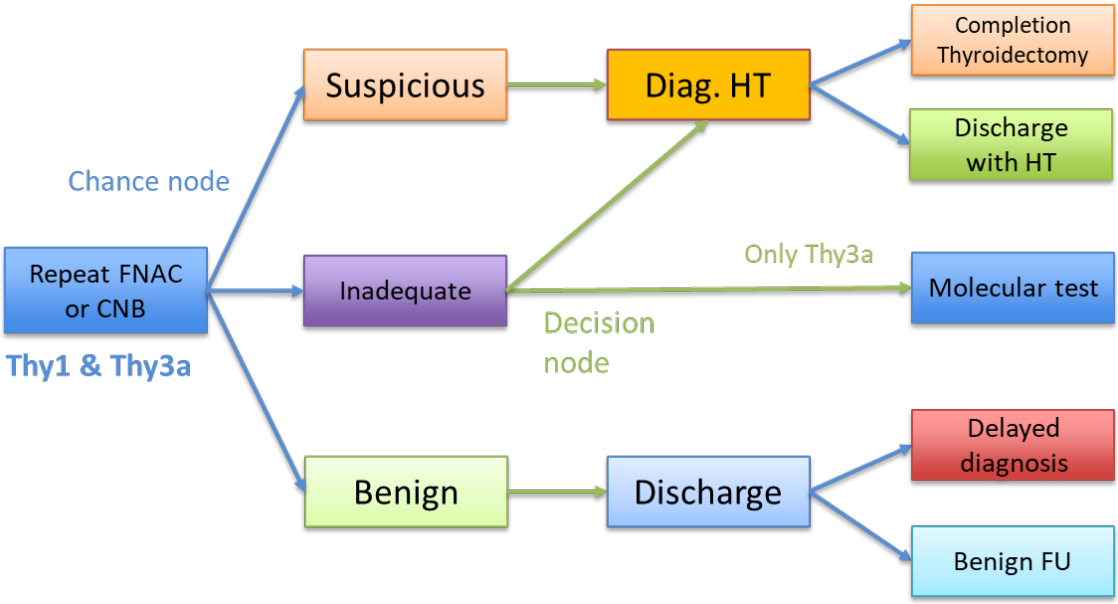


Abbreviations: HT = hemithyroidectomy; FU = follow up

© NICE 2022. All rights reserved. Subject to Notice of rights.

In the repeat FNAC and CNB strategies, included only for Thy1 and Thy3a, the model structure is slightly different as shown in Figure 3. A FNAC and a CNB are not binary tests as a percentage of people tested will receive an inconclusive result (a further Thy1 or Thy3a). After a second inconclusive result, further repeat CNB or FNAC tests are clinically meaningless, and the only options become diagnostic hemithyroidectomy or molecular test (only for Thy3a). Molecular testing is not an included strategy for Thy1, therefore people with Thy1 receiving a further non-diagnostic cytology can only undergo a diagnostic hemithyroidectomy. People with Thy3a receiving a molecular test after a further indeterminate result will enter the same decision tree model illustrated in the Figure 2. As mentioned before, people being discharged after a benign test will either be false negative (delayed diagnosis) or true negative (benign follow-up) whereas people undergoing hemithyroidectomy will receive completion thyroidectomy if malignant or discharged if benign.

Figure 3: Diagnostic decision tree – repeat FNAC and CNB



Abbreviations: CNB = core needle biopsy; FNAC = fine needle aspiration cytology; HT = hemithyroidectomy; FU = follow up

At the end of this first decision tree, all the people undergoing surgery, either a diagnostic hemithyroidectomy or a further completion thyroidectomy, enter the Surgical decision tree.

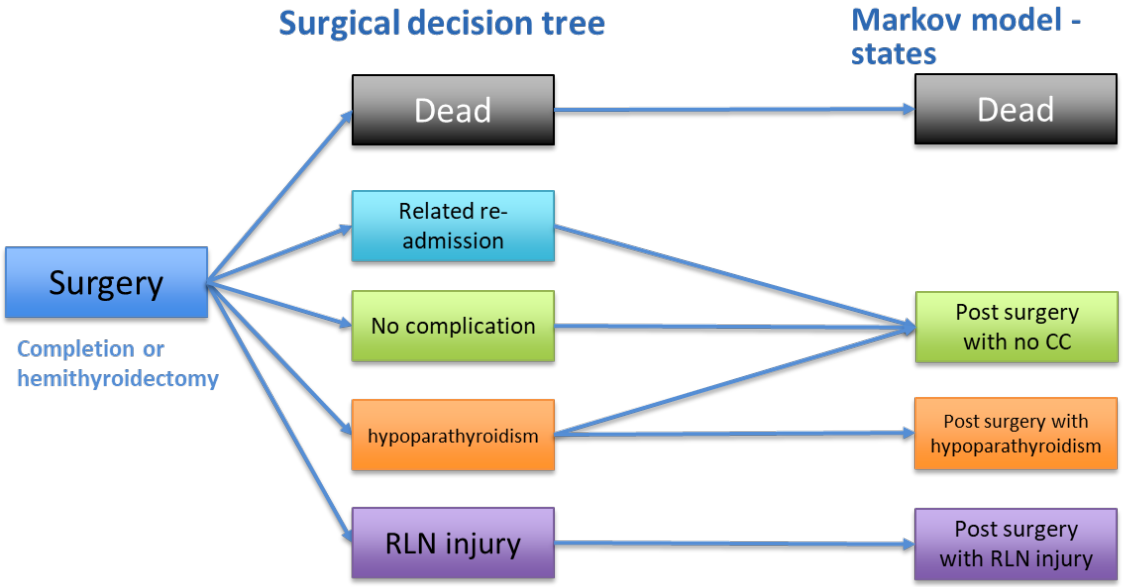
2.2.1.2 Surgical decision tree

People who receive a hemithyroidectomy or a completion thyroidectomy are a risk of several complications, including death. These are illustrated in Figure 4.

Most of the complications, such as related re-admission and most cases of hypoparathyroidism, resolve in the short term and cause only a one-off healthcare cost and quality of life harm. These people will enter the Follow Up Markov model in the post-hemithyroidectomy / post-completion thyroidectomy with no complications. Some complications, however, are known to be permanent, like recurrent laryngeal nerve (RLN) injury and a proportion of those experiencing hypoparathyroidism post-surgery. People with these complications will enter their own post-surgery with complication (hypoparathyroidism or RLN injury) state. People who died during surgery will enter the Markov model in the dead state.

© NICE 2022. All rights reserved. Subject to Notice of rights.

Figure 4: Diagnostic decision tree – repeat FNAC and CNB



Abbreviation: CC = complications, RLN = recurrent laryngeal nerve

2.2.1.2.1 Follow-up Markov model

The final Follow-Up Markov model is illustrated in Figure 5. People enter the Follow-up Markov model in the states defined by the two previous decision tree models. People who were discharged with a benign nodule and avoided surgery, enter the model in the "benign follow-up" state, which is part of the macro state "Benign" also including people with a benign nodule who underwent a diagnostic hemithyroidectomy. People who received a hemithyroidectomy or a completion thyroidectomy enter the corresponding states in the Markov model and those who have a permanent complication, namely RLN injury or hypoparathyroidism, enter the state with the corresponding complication. People whose diagnosis was delayed due to a false negative diagnostic result firstly enter the non-treated state but are assumed to seek medical help within 2 years due to clinical concern on the nodules thus receiving late surgery.

People in the benign, malignant and delayed diagnosis macro states can die with a state-specific risk although the risk does not vary across macro states. In other words, people who are well with malignant nodules have the same mortality of people who are well with a benign nodule as it is assumed that, if the nodule was successfully treated and there is no evidence of recurrent/persistent disease and no surgical complications, the condition does not increase mortality. By contrast both RLN injury and hypoparathyroidism increase mortality risk and reduce quality of life compared with no complications.

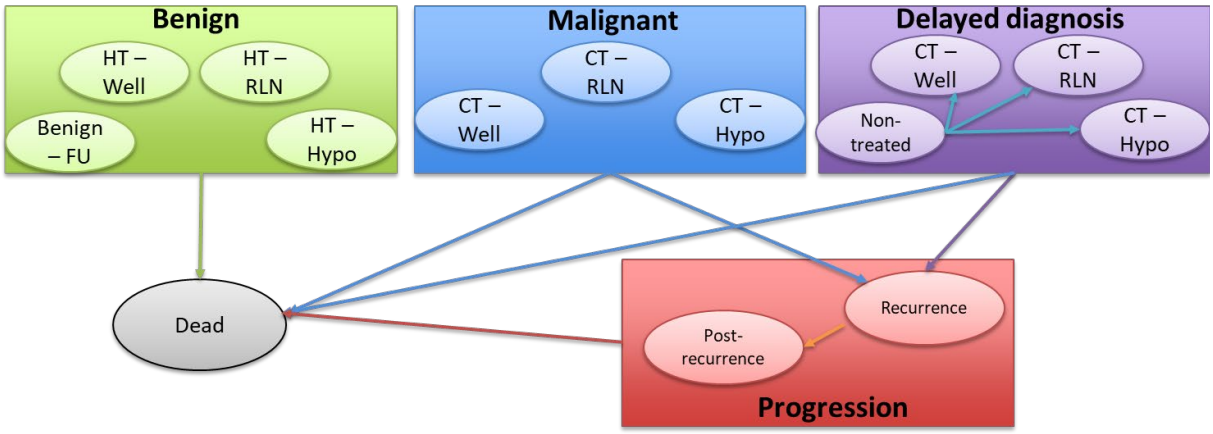
People with a delayed diagnosis enter the Markov model in the non-treated state where they share the same quality of life and mortality of the benign population. However, differently from people in the benign state, they are at risk of progression if their cancer worsens. Two years after entering the model, the model assumes that further clinical concerns (e.g. nodule enlargement) prompt non-treated people to seek medical help leading to the correct diagnosis of the cancer. Subsequently, they move to the Surgical decision tree which determines their post-surgical states. From that time onward, they will be a at a higher risk of progression due to the delayed surgery (see section 2.3.6).

All people with malignant nodules, namely people in the malignant or delayed diagnosis macro states, are at risk of developing recurrent or persistent disease moving to the progression macro state. Here, they first enter a tunnel state representing the first year of

© NICE 2022. All rights reserved. Subject to Notice of rights.

recurrence when they receive the treatment (radioactive iodine ablation in the base case scenario) and then move to a post-recurrence state characterized by higher long-term cost, lower quality of life and higher mortality.

Figure 5: Follow-up Markov model



Abbreviations: CT = Completion thyroidectomy; FU = Follow-up; HT = Hemithyroidectomy, RLN = Recurrent laryngeal nerve; Hypo = Hypoparathyroidism,

A yearly cycle length was chosen, and a half-cycle correction was applied to the Markov model which assumes that events occurred halfway through the cycle (at 6 months).

The following is a list of key simplifying assumptions:

- People can have only one complication. Although there may be cases where people have both RLN and permanent hypoparathyroidism, these are extremely rare and expected not to have a large impact to the analysis
- People moving to the recurrent state have the same mortality and quality of life regardless of whether they had any complication before. This was done for practical purposes as mortality in the recurrent state come from a real-world study that did not look at differences in mortality due to surgical complications. Moreover, as recurrence and complication rates are both very low, this assumption is not expected to have any significant impact on the final results.
- A delayed diagnosis is assumed to have no direct impact on mortality but to increase mortality through the indirect channel of recurrence. This was confirmed by the committee who explained that, when a person receives a wrong diagnosis, it is very unlikely that the cancer will cause the death of the person before being detected. However, a missed diagnosis can significantly delay surgery, which can have a direct impact on the stage of the cancer at the time of surgery. This, in turn, increases the probability of an unsuccessful total ablation of thyroid issue, which affects the risk of experiencing persistent/recurrent disease downstream.

2.2.2 Uncertainty

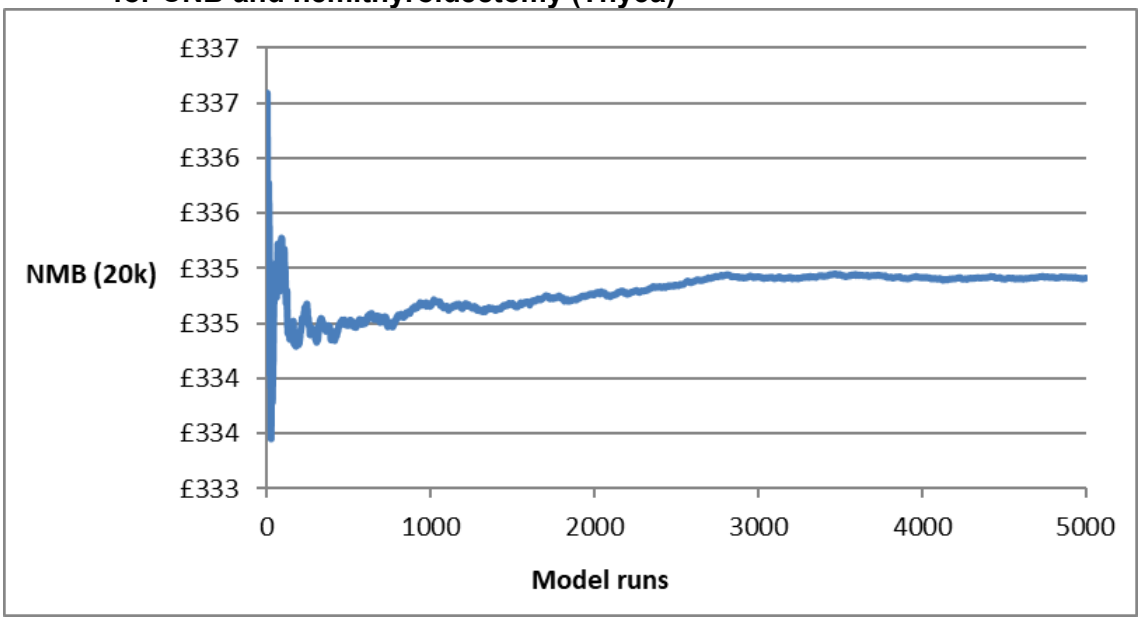
The model was built probabilistically to take account of the uncertainty around input parameter point estimates. A probability distribution was defined for each model input parameter. When the model was run, a value for each input was randomly selected simultaneously from its respective probability distribution; mean costs and mean QALYs were calculated using these values. The model was run repeatedly 5,000 times for each subpopulation and results were summarised.

When running the probabilistic analysis, multiple runs are required to take into account random variation in sampling. To ensure the number of model runs were sufficient in the

© NICE 2022. All rights reserved. Subject to Notice of rights.

probabilistic analysis we checked for convergence in the incremental costs, QALYs and net monetary benefit at a threshold of £20,000 per QALY gained. This was done by plotting the number of runs against the mean outcome at that point (see example in Figure 6) for the base-case analysis. Convergence was assessed visually and all had stabilised before 5000 runs.

Figure 6: Checking for convergence: Net monetary benefit at a threshold of £20,000 for CNB and hemithyroidectomy (Thy3a)



Abbreviations: NM B= Net Monetary Benefits

The way in which distributions are defined reflects the nature of the data, so for example event probabilities were given a beta distribution, which is bounded by 0 and 1, reflecting that the probability of an event occurring cannot be less than 0 or greater than 1. All of the variables that were probabilistic in the model and their distributional parameters are detailed in Table 2 and in the relevant input summary tables in section 2.3.1. Probability distributions in the analysis were parameterised using error estimates from data sources.

Table 2: Description of the type and properties of distributions used in the probabilistic sensitivity analysis

Parameter	Type of distribution	Properties of distribution
Risk of malignancy Surgery consequences Sensitivity of test Probability of recurrence Mortality Disutility factors	Beta	Bounded between 0 and 1. As the sample size and the number of events were specified alpha and beta values were calculated as follows: • Alpha = (number of patients hospitalised) • Beta = (number of patients) − (number of patients hospitalised)
Hazard ratios	Lognormal	The natural log of the mean and standard error were calculated as follows: • Mean = ln(mean cost) − $SE^2/2$ • SE = [ln(upper 95% CI) − ln(lower 95% CI)]/(1.96×2) $\sqrt{\ln \frac{SE^2 + mean^2}{mean^2}}$

© NICE 2022. All rights reserved. Subject to Notice of rights.

Parameter	Type of distribution	Properties of distribution
		This formula includes a correction to ensure the mean generated in the probabilistic analysis will be the same as the reported mean.
Diagnostic odds ratios Parameters of recurrence lognormal distribution Dosage	Normal	Symmetric from the peak of the curve with most of the observed data clustered near the mean. It is unbounded and it was used not to contain direction of change.
Utilities	Beta	Bounded between 0 and 1. Derived from mean and its standard error, using the method of moments. Alpha and Beta values were calculated as follows: Alpha = $mean^2 \times [(1-mean)/SE^2] - mean$ Beta = $alpha \times [(1-mean)/mean]$
Duration of acute events Length of stay Utility decrements	Gamma	Bounded at 0, positively skewed. Derived from mean and its standard error. Alpha and beta values were calculated as follows: • Alpha = $(mean/SE)^2$ • Beta = $SE^2/Mean$

Abbreviations: 95% CI = 95% confidence interval; SE = standard error; SMR = standardised mortality ratio.

The following variables were left deterministic (that is, they were not varied in the probabilistic analysis):

- the cost-effectiveness threshold (which was deemed to be fixed by NICE),
- incidence of asthma based on national audit of all NHS trusts
- drug prices based on BNF and NHS indicative price
- average weight of females and males in England based on ONS national data
- sensitivity and specificity of hemithyroidectomy which is always assumed to be 1 as recommended by the committee
- Health states costs based on analyses using unit costs from UK national sources
- Mortality in the general population based on English life tables
- Utility scores in the general population based on the latest Decision Support Unit (DSU) analysis

In addition, various deterministic sensitivity analyses were undertaken to test the robustness of model assumptions. In these, one or more inputs were changed and the analysis rerun to evaluate the impact on results and whether conclusions on which intervention should be recommended would change. Details of the sensitivity analyses undertaken can be found in methods section 2.4 Sensitivity analyses.

2.3 Model inputs

2.3.1 Summary table of model inputs

Model inputs were based on clinical evidence identified in the systematic review undertaken for the guideline, supplemented by additional data sources as required. Model inputs were validated with clinical members of the guideline committee. A summary of the model inputs used in the base-case (primary) analysis is provided in Table 3 below. More details about sources, calculations and rationale for selection can be found in the sections following this summary table.

© NICE 2022. All rights reserved. Subject to Notice of rights.

Table 3: Overview of parameters and parameter distributions used in the model

Input	Data	Source	Probability distribution
Comparators	• Molecular testing ○ Afirma-GSC ○ ThyroSeq V3 ○ ThyroSeq V1 ○ ThyGenX/ThyraMIR • Core-needle biopsy (CNB) ○ With molecular testing after further inadequate cytology ○ With hemithyroidectomy after further inadequate cytology • Repeat fine-needle aspiration cytology (FNAC) ○ With molecular testing after further inadequate cytology ○ With hemithyroidectomy after further inadequate cytology • Diagnostic hemithyroidectomy		n/a
Population	People with inadequate FNAC results (Thy1, Thy3a and Thy3f)		n/a
Perspective	UK NHS & PSS	NICE reference case[25]	n/a
Time horizon	Lifetime		n/a
Discount rate	Costs: 3.5% Outcomes: 3.5%	NICE reference case[25]	n/a
Cohort settings			
Age	50	BAETS 2021[46]	Fixed
Proportion of females	79%	BAETS 2021[46]	Beta
Weight	Female: 70.2 kg Male: 83.6 kg	ONS[33]	Fixed
Risk of malignancy (ROM)			

© NICE 2022. All rights reserved. Subject to Notice of rights.

Input	Data	Source	Probability distribution
Thy1	RCPath: 12% BAETS: 17% Bongiovanni: 17%	RCPath 2020[38] BAETS 2021[46] Bongiovanni 2012 [5]	Beta
Thy3a	RCPath: 25% BAETS: 27% Bongiovanni: 16%	RCPath 2020[38] BAETS 2021[46] Bongiovanni 2012 [5]	Beta
Thy3f	RCPath: 31% BAETS: 27% Bongiovanni: 26%	RCPath 2020[38] BAETS 2021[46] Bongiovanni 2012 [5]	Beta
Tests accuracy			
Afirma-GSC	**Thy3a** Sensitivity: 0.93 Specificity: 0.71 **Thy3f** Sensitivity: 0.88 Specificity: 0.64	Patel 2018 [36]	Sensitivity = Beta DOR = Normal
ThyroSeq V3	**Thy3a** Sensitivity: 0.91 Specificity: 0.85 **Thy3f** Sensitivity: 0.97 Specificity: 0.75	Nikiforov 2019[32]	Sensitivity = Beta DOR = Normal
ThyGenX + ThyraMIR	**Thy3a** Sensitivity: 0.94 Specificity: 0.79 **Thy3f** Sensitivity: 0.82 Specificity: 0.91	Labourier 2015[19]	Sensitivity = Beta DOR = Normal
ThyroSeq V1	**Thy3a and Thy3f** Sensitivity: 0.85 Specificity: 0.65	Shrestha 2016[43]	Sensitivity = Beta DOR = Normal
CNB inconclusive rate	Thy1: 7.3% Thy3a: 20.6%	Pyo 2016[40]	Beta
Repeat FNAC inconclusive rate	Thy1: 30.1% Thy3a: 32.7%	Pyo 2016[40]	Beta
CNB accuracy (excluding inconclusive)	**Thy1 and Thy3a** Sensitivity: 0.94 Specificity: 88	Pyo 2016[40]	Sensitivity = Beta DOR = Normal
FNAC accuracy (excluding inconclusive)	**Thy1 and Thy3a** Sensitivity: 0.94 Specificity: 0.86	Evidence review D FNAC or biopsy	Joint posterior distribution
Diagnostic hemithyroidectomy	Sensitivity: 1 Specificity: 1	Assumed, expert opinion	Fixed

© NICE 2022. All rights reserved. Subject to Notice of rights.

Input	Data	Source	Probability distribution
Surgery consequences			
Hemithyroidectomy			
30 days mortality	0.05%	HES-ONS[29]	Beta
Length of stay (days)	1.29	BAETS 2021[46]	Beta
Permanent RLN palsy	1.19%	BAETS 2021[46]	Beta
Transient hypoparathyroidism	0.65%	BAETS 2021[46]	Beta
Permanent hypoparathyroidism	0%	Assumed, expert opinion	Fixed
Need of TH replacement	15.50%	BAETS 2021[46]	Beta
Related readmission	1.09%	BAETS 2021[46]	Beta
Voice change	5.87%	BAETS 2021[46]	Beta
Completion thyroidectomy			
30 days mortality	0.1%	HES-ONS [29]	Beta
Length of stay (days)	1.48	BAETS 2021[46]	Beta
Permanent RLN palsy	1.98%	BAETS 2021[46]	Beta
Transient hypoparathyroidism	8.93%	BAETS 2021[46]	Beta
Permanent hypoparathyroidism	6.01%	BAETS 2021[46]	Beta
Need of TH replacement	100%	Assumed, expert opinion	Fixed
Related readmission	2.02%	BAETS 2021[46]	Beta
Voice change	7.66%	BAETS 2021[46]	Beta
Total thyroidectomy			
30 days mortality	0.1%	HES-ONS[29]	Beta
Length of stay (days)	2.09	BAETS 2021[46]	Beta
Permanent RLN palsy	2.75%	BAETS 2021[46]	Beta
Transient hypoparathyroidism	18.26%	BAETS 2021[46]	Beta
Permanent hypoparathyroidism	6.01%	BAETS 2021[46]	Beta
Need of TH replacement	100%	Assumed, expert opinion	Fixed
Related readmission	2.88%	BAETS 2021[46]	Beta
Voice change	8.78%	BAETS 2021[46]	Beta
Baseline probabilities and incidence			
Pneumonia incidence	51-61= 0.00295 61-71= 0.00461	BTS 2019[21]	Fixed

© NICE 2022. All rights reserved. Subject to Notice of rights.

Input	Data	Source	Probability distribution
	71-80= 0.00843 80 and over= 0.01838		
RLN on Pneumonia HR	6.21	Oshikiri 2021[35]	Lognormal
Progression if untreated	5 years: 5.26%	Ito 2014[12]	Beta
Recurrence	**Year** 0= 0% 1= 0.97% 2= 1.50% 3= 1.76% 4= 2.35% 5= 2.65% 6= 3.36% 7= 6.09% 8= 7.85%	HiLo trial [23]	Beta
Recurrence lognormal distribution	Mean: 4.72 Standard deviation: 1.70	Calculated using R studio and data from HiLo trial[23]	Normal
Recurrence with delayed diagnosis HR	2.28	Yeh 2004[51]	Lognormal
Duration of delay	2 years	Yeh 2004[51]	Fixed
Proportion of HT with malignancy needing CT	Base case: 100% Sensitivity analysis: • 50% • 25% • 0%	Assumed, expert opinion	Fixed
Proportion of people with recurrence needing salvage surgery	Base case: 100% Sensitivity analysis: • 50% • 25% • 0%	Assumed, expert opinion	Fixed
Mortality			
General population mortality	Gender and age-specific	ONS Life Tables 2016-2018[34]	Fixed
Persistent/recurrent disease mortality HR	3.01	Link 2015[22] calculated using the methodology described by Williamson[50]	Lognormal
Hypoparathyroidism mortality HR	2.09	Almquist 2015[1]	Lognormal
Pneumonia mortality	10.40%	BTS 2019[21]	Beta
Costs			
Total thyroidectomy	£2,515	NHS Reference Costs 2019-2020[31], NHS Reference Costs 2017-2018[30], BAETS 2021[46]	Fixed

© NICE 2022. All rights reserved. Subject to Notice of rights.

Input	Data	Source	Probability distribution
Hemithyroidectomy	£2,173	NHS Reference Costs 2019-2020[31], NHS Reference Costs 2017-2018[30], BAETS 2021[46]	Fixed
Completion thyroidectomy	£2,254	NHS Reference Costs 2019-2020[31], NHS Reference Costs 2017-2018[30], BAETS 2021[46]	Fixed
Afirma-GSC	**Manufacturer price** Retail price = £4,798 Medicare = £2,660 Cash pay = £2,918 **Published price** £4,454	Varacyte 19/01/2022 Kuo 2019[17]	Fixed
ThyroSeq V3	**Manufacturer price** International tariff = £1,407 **Published price** £2,227	University of Pittsburgh Medical Center, 11/01/2022 Kuo 2019[17]	Fixed
ThyGenX + ThyraMIR	**Manufacturer price** Unavailable **Published price** £2,245	Kuo 2019[17]	Fixed
ThyroSeq V1	Assumed to be same of ThyroSeq V3	n/a	Fixed
Shipping cost	£107	UPS Express from London to Pittsburgh[48]	Fixed
Box UN3373	£8	Royal Mail[42]	Fixed
US-guided FNAC	£299	NHS Reference Costs 2019-2020[31]	Fixed
US-guided CNB	£429	NHS Reference Costs 2019-2020[31]	Fixed
Endocrinology follow-up	£151	NHS Reference Costs 2019-2020[31]	Fixed
Re-admission after surgery	£1,184	NHS Reference Costs 2019-2020[31]	Fixed
Laryngoscopy	£189	NHS Reference Costs 2019-2020[31]	Fixed
Pneumonia	£1,909	NHS Reference Costs 2019-2020[31]	Fixed
Radioactive Iodine Ablation (RAI)	£535	NHS Reference Costs 2019-2020[31]	Fixed
Thyrotropin Alfa	£583	NHS Indicative Price[13]	Fixed
Healthcare cost with withdrawal	£142	See separated report "Cost-utility analysis: rhTSH vs thyroid Hormone withdrawal in	Fixed

© NICE 2022. All rights reserved. Subject to Notice of rights.

Input	Data	Source	Probability distribution
		people in preparation of Radioactive Iodine Ablation (RAI)"	
Calcium supplement (per year)	£39	BNF[13] PCA[11]	Fixed
Vitamin D supplement (per year)	£143	BNF[13] PCA[11]	Fixed
Thyroid hormone replacement (per year)	TSH suppression therapy = £87 Maintenance therapy = £41	BNF[13] PCA[11] Banovac 1990[3]	Gamma
Duration of suppression	10 years	Assumed, expert opinion	Fixed
Thyroglobulin test	£16	NICE thyroid disease guideline[28]	Fixed
Tg antibody test	£16	NICE thyroid disease guideline[28]	Fixed
TSH test	Non-stimulated = £2 Stimulated = £585	NICE thyroid disease guideline[28] NHS Indicative Price[13]	Fixed
Ultrasound scan without contrast	£52	NHS Reference Costs 2019-2020[31]	Fixed
Endocrinology attendance	£151	NHS Reference Costs 2019-2020[31]	Fixed
Quality of life			
General population	Gender and age-specific	NICE DSU unit[9]	Fixed
Disutility factor with RLN injury	0.627	Kebebew 2000[14]	Beta
Disutility factor with hypoparathyroidism	0.778	Kebebew 2000[14]	Beta
Disutility factor with recurrence	0.54	Kebebew 2000[14]	Beta
Utility decrement with THW	0.013	See separated report "Cost-utility analysis: rhTSH vs thyroid Hormone withdrawal in people in preparation of Radioactive Iodine Ablation (RAI)"	Gamma
Duration of transient hypoparathyroidism	2 months	Expert opinion	Gamma

Abbreviations: BATES = British Association of Endocrine and Thyroid Surgeons; BNF = British National Formulary; BTS = British Thoracic Society; CNB = Core needle biopsy; CT = Completion thyroidectomy; DOR = Diagnostic Odds Ratio; DSU = Decision Support Unit; FNAC = Fine needle aspiration cytology; GSC = Gene Sequencing Classifier; HT = Hemithyroidectomy; ONS = Office of National Statistics; PCA = Prescription Cost Analysis; RCPath = Royal College of Pathologists; RLN = Recurrent laryngeal nerve; TH = Thyroid Hormone; THW = Thyroid hormone withdrawal; TSH = Thyroid-stimulating hormone

© NICE 2022. All rights reserved. Subject to Notice of rights.

2.3.2 Cancer prevalence and cohort characteristics

To define the characteristics of a FNA cytology, this analysis will use the UK Thy classification[7], which was developed by the Royal College of Pathologist (RCPath) to address the need of a standardised classification across the NHS. There are 5 Thy levels in the RCPath classification although the indeterminate cytology Thy3 is further divided into Thy3a (neoplasm possible–atypia/nondiagnostic) and Thy3f (neoplasm possible). There is also a separate category for thyroid cysts: Thy1c and Thy2c. RCPath classification aligns with other international reporting systems for FNAC and can be easily compared (see Table 4). In the course of this analysis, studies based on other international classification systems were converted into the RCPath Thy classification system using the comparison provided by the RCPath[38] and illustrated in Table 4.

Table 4: International terminology system for FNAC

RCPath	Bethesda	Italian	Australian	Japanese
Thy1 – Non diagnostic	I – Nondiagnostic or unsatisfactory	TIR 1 – Nondiagnostic	1 – Nondiagnostic	1 – Inadequate
Thy1 – Nondiagnostic cystic	n/a	TIR 1c – Nondiagnostic cystic	n/a	n/a
Thy2 – Nonneoplastic	II – Benign	TIR2 – Nonmalignant	2 – Benign	2 – Normal or benign
Thy2c – Nonneoplastic cystic	n/a	n/a	n/a	n/a
Thy3a – Neoplasm possible - atypia	III – Atypia of undetermined significance (AUS/FLUS)	TIR 3A – Low-risk indeterminate lesion	3 – Indeterminate or follicular lesion of undetermined significance	3 – Indeterminate
Thy 3f – Neoplasm possible, suggesting follicular neoplasm	IV – Follicular neoplasm or suspicious for a follicular neoplasm	TIR 3B – High-risk indeterminate lesion	4 – Suggestive of a follicular neoplas	3 – Indeterminate
Thy4 – Suspicious of malignancy	V – Suspicious for malignancy	TIR 4 – Suspicious for malignancy	5 – Suspicious for malignancy	4 – Suspicious for malignancy
Thy5 – Malignant	VI – Malignant	TIR 5 – Malignant	6 – Malignant	5– Malignancy

Abbreviations: AUS = Atypia of Undetermined Significance; FLUS = Follicular Lesion of Undetermined Significance; RCPath = Royal College of Pathologists;
Source: Poller 2020[38]

Every cytology is associated with a different cancer prevalence or risk of malignancy (ROM) (see Table 5). Thy2 benign category has the lowest risk of malignancy, ranging from 3% to 9%, which justifies the current approach of discharging or repeat the test in the presence of other reasons of concern. The malignant categories Thy4 and Thy5 have the highest risk of malignancy and are often followed up with diagnostic or therapeutic surgery. The non-diagnostic category Thy1 and the indeterminate categories Thy3a and Thy3f show a risk of malignancy ranging from 12% (Thy1) and 25-30% (Thy3a and Thy3f) and their management is less certain (see Appendix A: for a graphic explanation of the terminologies that will be used in this report).

Currently, Thy1 and Thy3a receive repeat FNAC whereas Thy3f are followed up with diagnostic surgery. The low ROM of Thy3f implies that the majority of the surgeries are

© NICE 2022. All rights reserved. Subject to Notice of rights.

performed on people with benign nodules, and therefore could be avoided if better diagnostic techniques were available. Repeat FNAC is known to have poor performance on people who previously received a Thy1 or Thy3 cytology[40], thus it is likely that a large proportion of people whose FNAC is repeated would receive the same result. Given the uncertain cost-effectiveness of the current management of Thy1, Thy3a and Thy3f, this analysis seems to be strongly justified to shed light on the usefulness of alternative diagnostic techniques.

Table 5: ROM in the various RCPath categories and current practice

RCPath classification	Risk of Malignancy (ROM)	Current practice (BTA)
Thy1	12% (5% - 22%)	US assessment + repeat FNAC
Thy2	5% (3% - 9%)	Discharge or repeat FNAC
Thy3a	25% (20% - 31%)	Repeat FNAC
Thy3f	31% (24% - 39%)	Diagnostic hemithyroidectomy
Thy4	79% (70% - 87%)	Diagnostic hemithyroidectomy
Thy5	98% (97% - 99%)	Therapy appropriate to tumour

Sources: ROM: Poller 2020[38]. Current practice: British Thyroid Association (BTA)[37]

In the base case scenario ROM from Table 5 is used to estimate cancer prevalence in Thy1, Thy3a and Thy3f whereas two alternative sources for prevalence are tested in the sensitivity analysis[5, 46].

The characteristics of the cohort populating the model are illustrated in Table 6. People are assumed to have 50 years old as this is the age group more frequently receiving thyroid surgery in the UK[46]. The proportion of female in the cohort was taken from the latest audit from the British Association of Endocrine & Thyroid Surgeon (BAETS) whereas average weight in both genders was informed from the Office of National Statistics (ONS).

Table 6: Initial cohort characteristics

Characteristic	Value	Source
Age	50	BAETS 2021[46]
Proportion female	79%	BAETS 2021[46]
Average weight	Female: 70.2 kg Male: 83.6	ONS[33]

2.3.3 Diagnostic accuracy

Accuracy of the different diagnostic tests included in the model strategies was obtained from the "Evidence review D FNAC or biopsy" and, when needed, published literature, prioritising large meta-analyses when available. The quality of external studies was assessed through the QUADAS table, for single studies, and with the ROBIS method, for meta-analyses (see Appendix B:).

2.3.3.1 Molecular testing

The committee identified, four molecular tests to be included in the analysis, as they were either relevant for the UK or represent the state of the art currently available in the global market. These were Afirma-GSC, ThyroSeq V3 and V1 and ThyGenX/ThyraMIR. Table 7: Sources for molecular testing accuracyTable 7 summarises the studies used to estimate the accuracy of each test.

© NICE 2022. All rights reserved. Subject to Notice of rights.

Table 7: Sources for molecular testing accuracy

Test	Study	Study description	QUADAS 2 risk of bias[(a)]
Afirma-GSC	Patel 2018[36]	Prospective double-blinded multicentre study on 191 indeterminate nodules	Serious risk of bias
ThyroSeq V3	Nikiforov 2019[32]	Prospective double-blinded multicentre study on 257 indeterminate nodules	Serious risk of bias
ThyroSeq V1	Shrestha 2016[43]	Retrospective single centre study on 261 nodules	Very serious risk of bias
ThyGenX/ThyraMIR	Labourier 2015[18]	Retrospective multicentre study on 109 indeterminate nodules	Very serious risk of bias

(a) For the full QUADAS table please Appendix B:

Two molecular tests, Afirma-GSC and ThyroSeq V3, have been only recently introduced in the market, and their accuracy was estimated from recent prospective double-blinded multicentre analyses[32, 36]. A meta-analysis[49] reporting the accuracy of Afirma-GES and ThyroSeq v2 was identified, but these tests were considered outdated as they were found to be less accurate than their upgrade versions, Afirma-GSC and ThyroSeq V3 respectively, and they are presumably not in production anymore. Only one study[18] reporting the accuracy of ThyGenX/ThyraMIR was identified from the same systematic review[49] and used in the analysis although the study is a non-blinded retrospective study and with possibly lower quality than the two previous studies mentioned according to QUADAS 2. ThyroSeq V1 was the first ThjyroSeq NGS (Next Generation Sequency Assay) developed by the University of Pittsburgh Medical Center and is no longer in use as the upgraded version 3 includes additional significant genomic regions. Nevertheless, it was included in the analysis to address a specific request of the committee as it mimics the type of molecular test that can be currently offered from an NHS regional hub. Its accuracy was estimated from a retrospective single centre study on 261 nodules[43] and was found to be considerably lower than the accuracy of the other three molecular tests considered. These latter three, however, although more sophisticated, are produced only in the US and, if implemented, would likely require some time for the NHS to create a distribution network for routine use. Therefore, ThyroSeq V1 probably shows what the NHS can offer in the short term. The quality of each study was assessed using the QUADAS 2 methodology, finding the two retrospective studies on ThyGENx/ThyraMIR and ThyroSeq V1 with a higher risk of bias than the prospective studies (see Appendix B:). This will be further discussed later.

The sensitivity and specificity of the molecular tests included are illustrated in Table 8 and Table 9. As Thy3a and Thy3f nodules have different clinical features and different molecular markers, the accuracy of each test is expected to be different for each cytology, with some tests being more precise in Thy3f thanThy3a and vice versa. This is the reason accuracy of the test was extracted separately for Thy3a and Thy3f indeterminate nodules with the only exception of ThyroSeq V1, as the study did not distinguish between the two cytologies[43].

Table 8: Accuracy of molecular testing in Thy3a

Molecular test	Sensitivity (95%CI)	Specificity (95%CI)	Source
Afirma-GSC	0.93 (0.76 – 0.71)	0.71 (0.60 – 0.80)	Patel 2018[36]
ThyroSeq V3	0.91 (0.77 – 0.97)	0.85 (0.77 – 0.90)	Nikiforov 2019[32]
ThyroSeq V1	0.85 (0.55 – 0.98)	0.65 (0.45 – 0.81)	Shrestha 2016[43]
ThyGenX/ThyraMIR	0.94 (0.71 – 0.99)	0.79 (0.63 – 0.90)	Labourier 2015[18]

© NICE 2022. All rights reserved. Subject to Notice of rights.

Table 9: Accuracy of molecular testing in Thy3f

Molecular test	Sensitivity (95%CI)	Specificity (95%CI)	Source
Afirma-GSC	0.88 (0.64 – 0.99)	0.64 (0.51 – 0.76)	Patel 2018[36]
ThyroSeq V3	0.97 (0.85 – 1)	0.75 (0.63 – 0.84)	Nikiforov 2019[32]
ThyroSeq V1	0.85 (0.55 – 0.98)	0.65 (0.45 – 0.81)	Shrestha 2016[43]
ThyGenX/ThyraMIR	0.82 (0.56 – 0.95)	0.91 (0.76 – 0.98)	Labourier 2015[18]

Most of the tests show better accuracy in a Thy3a cytology than Thy3f. Afirma-GSC has a good sensitivity for both cytologies but falls behind in terms of specificity, especially when performed after a Thy3 cytology. ThyroSeq V3 is considered one of the most sophisticated tests on the market and achieves great sensitivity and a satisfactory specificity after both cytologies. ThyroSeq v1 which is the most outdated test included in the analysis showed, as expected, a relatively low accuracy.

2.3.3.2 CNB and repeat FNAC

The accuracy of CNB and repeat FNAC was sought from "Evidence review D FNAC or biopsy" and a published systematic review on the performance of the two tests after an inconclusive cytology.

A repeat FNAC or a CNB performed on a nodule that was previously found to be non-diagnostic or indeterminate are expected to be less accurate than first-time tests, due to the nature of the lesion of a Thy1 or Thy3 cytology. Specifically, a previous inadequate cytology is a good predictor of a further inadequate cytology as the new tissue collected through the additional biopsy does not always give additional information on the lesion. In particular, the Committee were aware that only rarely a repeat biopsy on a Thy3f nodule could give a different result and acknowledged that most studies on repeat sampling focused on Thy1 and Thy3a nodules only. For this reason, the repeat sampling strategies CNB and repeat FNAC were included only for Thy1 and Thy3a populations.

The clinical meta-analysis on the accuracy of first-time FNAC and CNB conducted for question 1.5 could not be solely used in this model, as using first-time accuracy would have greatly overestimated the accuracy of a repeat biopsy. Instead, external evidence on inconclusive rates of a repeat test[40] were sought and, when possible, combined with accuracy data from 1.5. Inconclusive results include both non diagnostic cytology Thy1 and indeterminate Thy3a. Thy3f, although indeterminate, is not considerate inconclusive and therefore is not included in the meta-analysis (See Appendix A:). Likewise, as mentioned, repeat sampling was not included as a strategy for Thy3f nodules.

The meta-analysis was assessed using the ROBIS methodology and was found to have a high risk of bias due to the lack of a proper pre-study protocol and unclear reporting of searching strategy. However, despite poor reporting, estimations of inconclusive rates were considered reliable and in line with most literature[24, 44, 45] and therefore included in the analysis. No other alternative source with comparable sample size for this information was identified.

Table 10: Proportion of inconclusive results of CNB and FNAC after prior FNAC with inadequate cytology

Test	Proportion of inconclusive results after Thy1	Proportion of inconclusive results after Thy3a
CNB	7.3% (3.4% to 15.3%)	20.6% (12.3% to 32.5%)
Repeat FNAC	30.1% (21.5% to 40.4%)	32.7% (26.8% to 39.2)

Inconclusive: Thy1 and Thy3a

Source: Pyo 2016[40]

© NICE 2022. All rights reserved. Subject to Notice of rights.

The meta-analysis of Pyo and colleagues found a much higher inconclusive rate with FNAC for both non-diagnostic and indeterminate cytologies (see Table 10). It is noteworthy that the inconclusive rate of repeat FNAC reported by Pyo 2016 is, generally, in line though in the lower range of the ones reported from other studies[45] that found rates up to 50% after a repeat FNAC. The literature generally agrees that the inconclusive rate is lower with CNB than with repeat FNAC, as CNB is able to collect a larger sample during the biopsy which can help the histopathologist or the radiologist to reach a final diagnosis.

Accuracy of a CNB on a population with who received an inconclusive result was estimated from the same meta-analysis Pyo 2016[40]. The accuracy of a repeat FNAC was not available from the same source and was estimated instead using the studies included for the review of 1.5. These were predominantly first-time FNAC and so, a meta-analysis of its accuracy was re-done using WinBugs which excluded the inconclusive categories Thy1 and Thy3a (see Appendix C:). Inconclusive rates of FNAC are instead estimated from Pyo 2016[40] (see Table 10) and, as expected, were much higher than the rates of the original review as a repeat test after an inconclusive result always return more inconclusive results than a first-time test. Accuracy of the remaining categories should not be significantly different between first-time and repeat test and therefore this approach was considered reasonable (see section 4.2). The final accuracy used in this analysis is presented in Table 11. The similarity between the accuracy of repeat FNAC and CNB is not unexpected, as the real difference between the two approaches is the number of inconclusive results (Thy1 and Thy3a) which are illustrated in Table 11.

Table 11: Sensitivity and specificity of FNAC and CNB of conclusive categories

Biopsy	Sensitivity (95% CI)	Specificity (95% CI)
FNAC[(a)]	0.94 (0.91 to 0.97)	0.86 (0.77 to 0.92)
CNB[(b)]	0.94 (0.88 to 0.97)	0.88 (0.84 to 91)

(a) 1.5 FNAC or biopsy evidence review
(b) Pyo 2016[40]

2.3.4 False positives and false negatives

2.3.4.1 False positives – unnecessary surgeries

Depending on the specificity of the tests used in each strategy, some people would be diagnosed with thyroid cancer even though their nodule was benign. These “false positives” are referred to a hemithyroidectomy they could have avoided were the tests more accurate. These surgeries are unnecessary as they will always result in a benign diagnosis and in the discharge of the person. People with benign nodules who undergo a diagnostic hemithyroidectomy enter their own state “HT well” or “HT with complications (either RLN palsy or hypoparathyroidism)” depending on the outcome of the surgery. They will have no lifetime risk of recurrence although a proportion of them will be under thyroid hormone replacement therapy.

2.3.4.2 False negatives – delayed diagnoses

Depending on the sensitivity of the tests used in each strategy, some people would be discharged with a benign diagnosis even though their nodule was malignant. People with a false negative diagnosis often return to the doctor when they note that the thyroid sweeling has increased in size. Yeh 2004[51] found that, on average, people who were initially discharge due to a negative cytology and whose surgery intervention was prompted on a later stage by clinical findings, such as an enlarging nodule, have to wait 2 years before receiving surgery. The model, therefore, assume that for the first 2 years after discharge, people with an undiagnosed thyroid cancer will continue living normally until, between year 2 and 3 of the Markov model, clinical concern such as the enlargement of the nodule will prompt them to

© NICE 2022. All rights reserved. Subject to Notice of rights.

attend a new endocrinological visit and receive a further FNAC. During this time, however, they will be at risk of the cancer progressing to clinical disease (see section 2.3.6.3).

After the FNAC, it is assumed that they will always undergo diagnostic hemithyroidectomy followed by completion thyroidectomy either as a consequence of a malignant cytology or due to clinical concerns.

The committee confirmed that a delayed diagnosis is less serious with thyroid cancer than with most of the other cancers due to its slowly (or even non-) progressive nature. It is fairly rare therefore, that a differentiated thyroid cancer will transform into a fatally aggressive anaplastic cancer before being diagnosed thus the same mortality of the general population is applied. However, a late diagnosis leads to a delayed surgery which, according to published findings[51], is associated with microscopic evidence of disease progression and predicts poorer outcome. This affects the probability of successful ablation and lifetime risk of developing recurrent or persistent disease (see section 2.3.6.3).

2.3.5 Thyroid surgery

As described in section 2.2.1, people receiving surgery enter a separate Decision Tree, where the outcomes are estimated depending on the type and number of surgeries received. There are two types of surgeries people can undergo: if people are found to be positive after a diagnostic test (either false or true), they receive a diagnostic hemithyroidectomy first. People can also receive hemithyroidectomy in some of the strategies included (i.e. CNB/Repeat FNAC & hemithyroidectomy) if the results at further test are still indeterminate. Hemithyroidectomy is assumed to have 100% sensitivity and specificity as it is the gold standard diagnostic technique for thyroid cancer. Hence, if the nodule is malignant, the hemithyroidectomy is always able to detect it, and the person is referred to a further completion thyroidectomy to remove the remaining part of thyroid tissue. In the scenario analyses, only a proportion of people with malignancy detected during hemithyroidectomy is assumed to receive completion thyroidectomy.

Hemithyroidectomy, completion thyroidectomy and total thyroidectomy have all different outcomes as illustrated in Table 12. Hemithyroidectomy is the simplest procedure and have a lower risk of mortality and complications and the shortest length of stay. In most cases people resume their normal life without needing any thyroid supplement although, in 15% of people who underwent hemithyroidectomy, the level of thyroid hormones produced by the remaining part of the thyroid is insufficient, and they are required to take adjustment doses of levothyroxine (T-4) to avoid hypoparathyroidism. The probability of experiencing permanent hypoparathyroidism after hemithyroidectomy is assumed to be 0, as recommended by the Committee.

Completion thyroidectomy, which is only performed after a hemithyroidectomy if the nodule was found to be malignant, has slightly higher complication rates and always require thyroid hormone replacement therapy afterwards as the entire of the thyroid tissue is assumed to be removed. Total thyroidectomy is the most complex procedure of the three and, although not explicitly modelled in the analysis, was tested in the scenario analysis.

Table 12: Surgery parameters

Parameters	Hemithyroidectomy	Completion thyroidectomy	Total thyroidectomy
Length of stay (days)	1.29	1.48	2.09
30 days mortality	0.05%	0.10%	0.10%
Permanent RLN palsy	1.19%	1.98%	2.75%
Transient hypoparathyroidism	0.65%	8.93%	18.26%

© NICE 2022. All rights reserved. Subject to Notice of rights.

Parameters	Hemithyroidectomy	Completion thyroidectomy	Total thyroidectomy
Permanent hypoparathyroidism	0%[(a)]	6.01%	6.01%
Need of TH replacement	15.5%	100%[(b)]	100%[(b)]
Related readmission	1.09%	2.02%	2.88%

(a) Assumed to be 0 based on expert opinion from the Committee
(b) Assumed to be 100% based on expert opinion from the Committee
Source: BAETS 2021[46] and HES-ONS mortality data[29]

At the end of the Surgical Decision Tree, people enter the long-term Markov Model in the state defined by the outcome of the surgery and the nature of their nodules. These affect their mortality, healthcare cost, quality of life and lifetime risk of recurrent/persistent disease. People who received a completion thyroidectomy are assumed to undergo radioactive iodine ablation (RAI) with recombinant human thyroid stimulating hormone (rhTSH) to increase radiation uptake.

2.3.6 Risk of recurrent/persistent disease

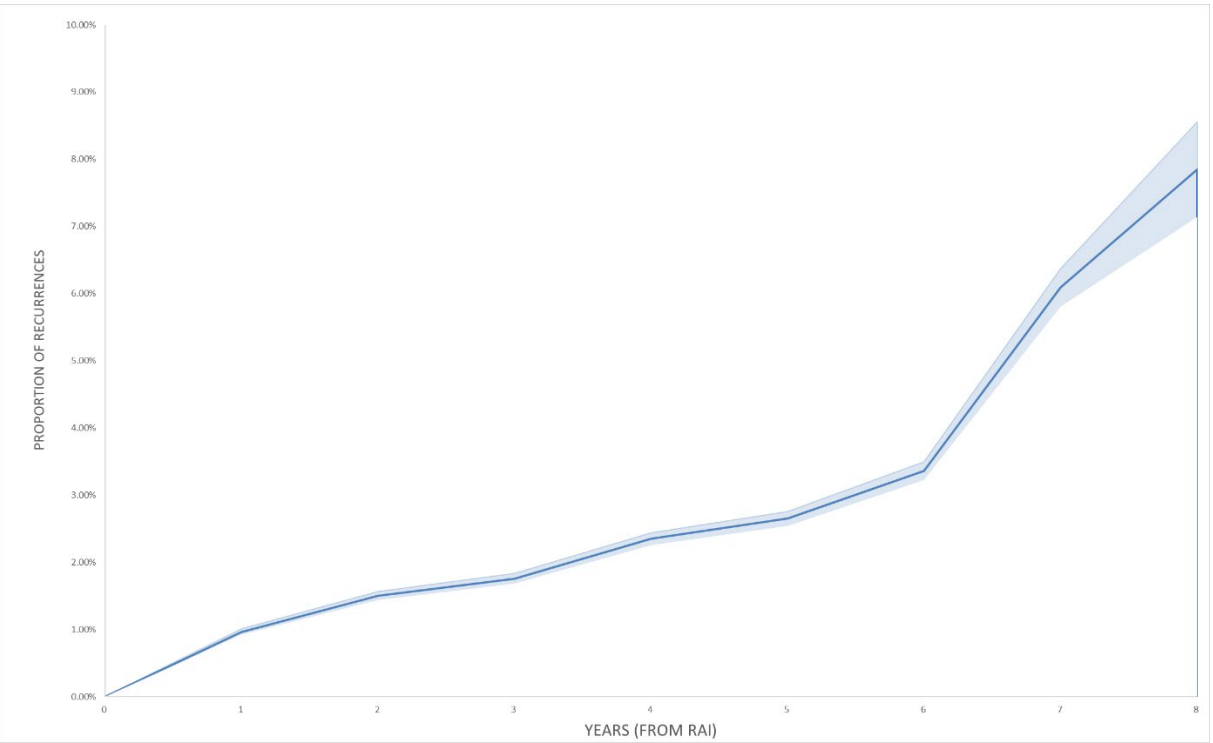
2.3.6.1 Baseline risk

Although uncommon, people treated for thyroid cancer are at risk of developing recurrent disease at some point during their lifetime. Published literature estimated that around 20% of the patients will develop either recurrent or persistent disease and the average time ranges anywhere from 6 months for persistent disease to decades later for recurrence[4].

Baseline rates of recurrence were estimated from the recent results of the English HiLo trial[10]. The Kaplan Meier (KM) curve of developing recurrent disease during the first 8 years after a total thyroidectomy is illustrated in Figure 7. The Kaplan Meier estimates a risk of 7.85% of having recurrence at 8 years although a large number of censoring occur between year 6 and year 8 implying that the last estimation points are less certain.

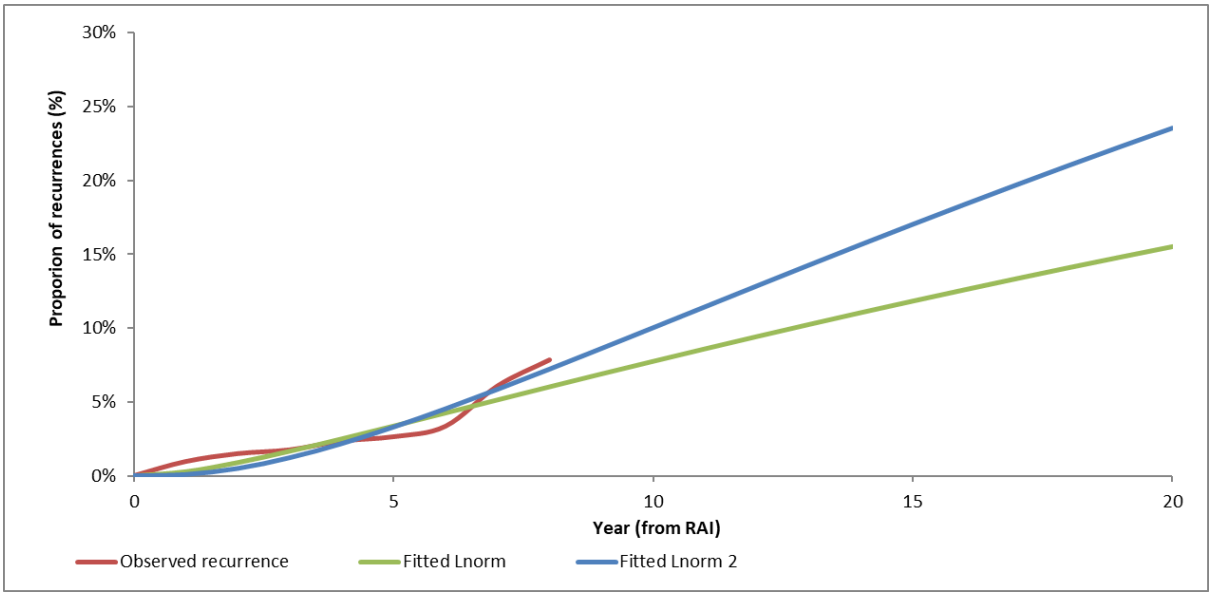
© NICE 2022. All rights reserved. Subject to Notice of rights.

Figure 7: Kaplan-Meier of risk of recurrence from HiLo trial



To extrapolate recurrence rate beyond the last follow-up, two different Log-normal curves were fitted using R studio. The blue curve was fitted using all the data points available from the Kaplan-Meier whereas, to construct the green curve, only the first 7 points were utilized to account for the high uncertainty beyond year 7 due to the low number of people at risk and high censoring.

Figure 8: Long-term extrapolation of recurrence with Log-Norm



The green curve achieved better face validity as it predicts a lifetime risk of developing recurrence equal to 24%, which is quite aligned with published literature[4]. On the other hand, the blue curve predicts a lifetime risk of 38% which is considerably higher than the estimation reported in the literature. For this reason, the green curve was chosen to predict the risk of

© NICE 2022. All rights reserved. Subject to Notice of rights.

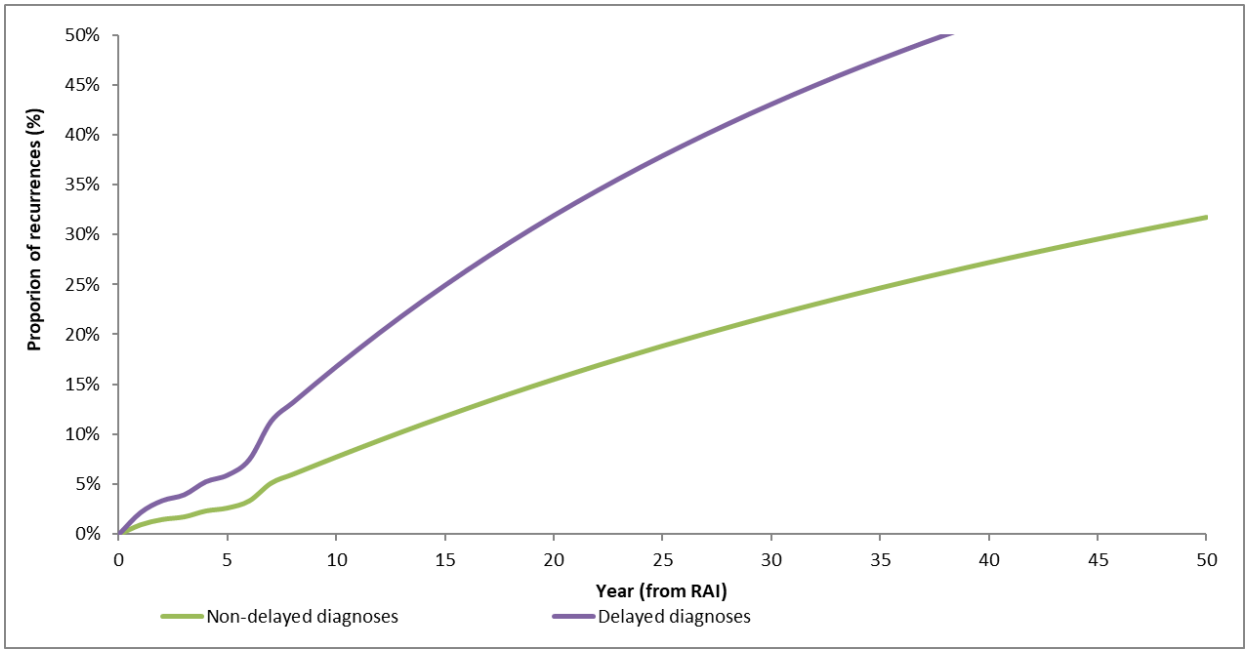
recurrence in the base case scenario whereas the blue curve was tested only in a sensitivity analysis.

2.3.6.2 Risk with delayed diagnosis

As anticipated in chapter 2.3.4.2, people with a delayed diagnosed are at a higher risk of developing recurrent or persistent disease during their lifetime after receiving surgery. This is because a delayed surgery is associated with microscopic evidence of disease progression. Yeh and colleagues[51] found in a multivariate analysis that a delayed diagnosis due to a false negative FNAC doubles the risk of persistent disease post treatment (HR 2.28, 1.07 to 4.85).

The committee agreed that persistent disease can be considered as a proxy of recurrent disease as the two conditions are very closely related, almost identical from a clinical perspective and mainly defined according to the time separating the event from surgery. Therefore, in the course of this analysis, recurrent and persistent diseases are considered as the same events and were modelled as a single Markov state. The hazard ratio from the multivariate analysis was applied to the baseline risk of recurrent disease to calculate the risk of recurrent/persistent disease in people whose diagnosed was delayed. This is shown in Figure **9**. People with a delayed diagnosis have a significantly higher risk of developing recurrent or persistent disease and, assuming a life expectancy of 84 years, the model predicts that 46% of them would develop the disease during their lifetime.

Figure 9: Risk of recurrence with delayed and non-delayed diagnosis



Abbreviations: RAI = Radioactive Iodine Ablation

© NICE 2022. All rights reserved. Subject to Notice of rights.

2.3.6.3 Progression to clinical disease for people with untreated malignant nodules

As explained in the previous section, people with a delayed diagnosis have a higher risk of developing recurrence once they receive surgery. However, prior to that, during the two years' time necessary for the cancer to be diagnosed, they are naturally not at risk of recurrence, as they never received surgery in the first place. They are instead at risk of natural progression of cancer to clinical disease if it becomes symptomatic. This is assumed to prompt immediate surgery that, as occurring after the cancer has shown clinical symptoms and possibly become metastatic, will leave people in a very high risk of having persistent disease thereafter. This was simulated in the model by assuming that people receiving surgery due to the clinical progression of an untreated cancer, move to the recurrence states where they share the same quality of life and costs of people who had a recurrence. This was deemed reasonable as they are likely to share a similar risk of having subsequent recurrences and therefore will be subject to the same quality of life, costs and mortality.

Natural progression of an average thyroid cancer is not known as most of the cancers are immediately treated once detected. However, especially in East Asian countries, it is becoming common practice to offer active surveillance instead of surgery to those showing features indicating very low risk of the cancer being aggressive. The largest observational study conducted on active surveillance, Ito 2014[12], followed a group of 1,235 with low-risk cancer and found a 5-year risk of progression to clinical disease ranging from 2.2% to 9.5% across three different age groups. The 5-year risk in the model was calculated by excluding the extremely low-risk group of people older than 60 giving an overall risk of 5.26%. However, this number reflects the risk of people with cancer showing low-risk features who were strictly monitored during the study time and, in a few cases, receive TSH suppression alongside the surveillance. Therefore, it likely underestimates the real risk of progression in an untreated and unmonitored population with some cancers having aggressive features that would prompt surgery if detected. As this risk is applied for the first two cycles of the model only, this is not expected to significantly affect the overall results of the analysis although this will be further discussed in section 4.1.

2.3.7 Mortality

2.3.7.1 Mortality with no complications or recurrence

People with a benign nodule and no complications or a successfully treated malignant nodule with no complications or sign of recurrence are assumed to share the same mortality of the general population. This was proposed by the Committee and confirmed by a published study[22] which found no difference between the general population and people with a treated malignant nodule and no evidence of persistent/recurrent disease. Their mortality was based on general population mortality data from life tables for England 2018-2019[34]. Cycle-specific mortality vectors were calculated taking into account the progression of age. At each cycle, the model uses gender split, which changes over time as males naturally die at a higher rate than women, to calculate the gender-weighted average probability of dying in each state, which is used to determine the overall mortality of the cohort in each cycle.

Mortality in the states with complications and persistent/recurrent disease was calculated by applying the related relative effects on baseline mortality (see next sections).

2.3.7.2 Mortality with surgical complications

The model includes two permanent surgical complication, hypoparathyroidism and recurrent laryngeal nerve (RLN) injury, that are known to affect people survival.

Although commonly found in the literature, the reason for the increased mortality in people with hypoparathyroidism remains still unclear. Some have proposed that the large use of

© NICE 2022. All rights reserved. Subject to Notice of rights.

supraphysiological doses of active vitamin D or lower PTH levels could play an important role[1]. One study[1] with 4899 patients using a multivariate cox regression found that permanent hypoparathyroidism doubles the risk of mortality (Hazard ratio = 2.09, 1.04 to 4.20). This number was validated by the Committee and used in the model to estimate mortality in people with hypoparathyroidism.

Although RLN is not expected to directly affect mortality, the Committee acknowledged that people with RLN may be at a higher risk of pneumonia during their lifetime which, in turn, may increase their mortality. A recent study found that postoperative recurrent laryngeal nerve palsy is strongly associated with pneumonia as it induces abnormal aspiration that can cause pneumonia[35]. The multivariate analysis showed that RLN significantly increases the incidence of pneumonia with a odds ratio of 6.21 (2.73 to 14.48). The increased incidence of pneumonia causes an excess risk of pneumonia in people with RLN that indirectly increases their mortality. The baseline incidence of pneumonia in the UK was calculated using the latest Pneumonia statistics from the British Lung Foundation[21]. These are shown in Table 13.

Table 13: Incidence of pneumonia in people with and without RLN injury

Age	Incidence in the general population	Incidence in people with RLN injury	Excess incidence rate
51-61	0.0029	0.018	0.015
61-71	0.0046	0.028	0.023
71-80	0.0084	0.050	0.042
80 and over	0.0184	0.104	0.086

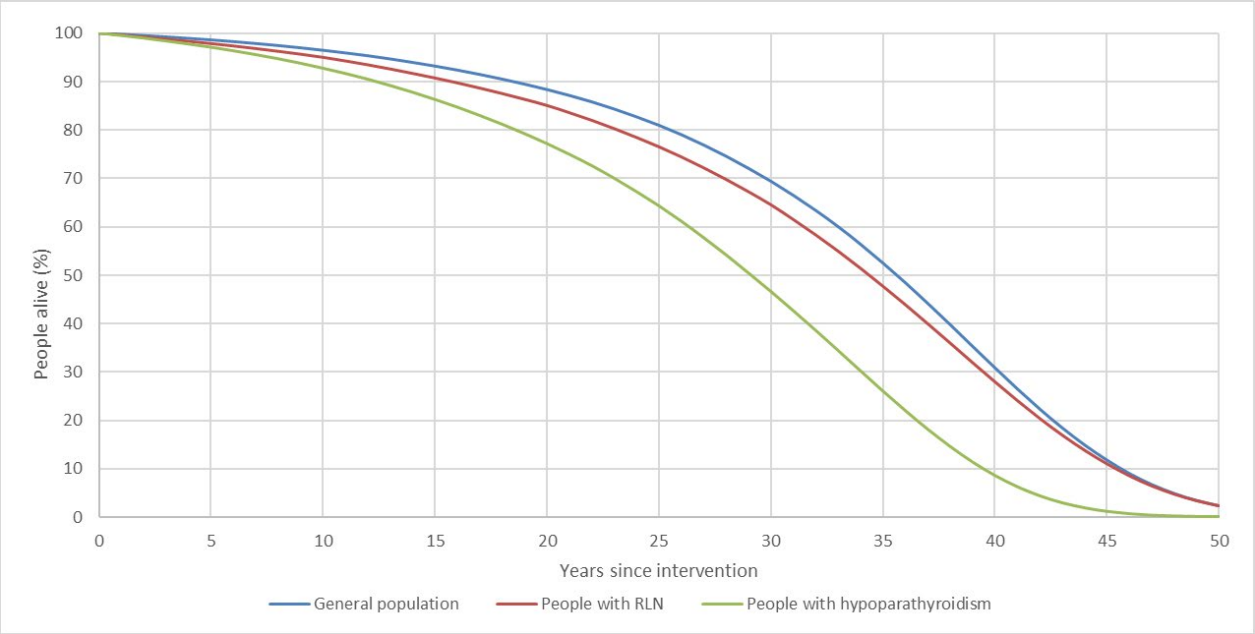
Abbreviations: RLN = Recurrent Laryngeal Nerve

The odds ratio of 6.21 recovered from the literature was used to calculate the incidence in people with RLN (third column of Table 11). However, only the excess incidence rate (fourth column), i.e. the additional number of cases of pneumonia caused by having RLN, is relevant for the model as the general population mortality estimated using English life tables already includes people who died of pneumonia. Using the actual incidence instead of the excess incidence would lead to double counting people dying of pneumonia in the RLN state.

Mortality with pneumonia was informed from the latest British Thyroid Society (BTS) audit on community acquired pneumonia[21] which estimated a 10.4% mortality risk in the UK in 2019. The mortality risk reported in the BTS audit was combined with the excess incidence rate previously calculated to create a vector of excess mortality due to pneumonia in people with RLN. This vector was therefore used to estimate mortality in people with RLN. In addition, excess incidence of pneumonia was also used to calculate excess pneumonia healthcare cost of people with RLN each year (see 2.3.9). Figure 10 illustrates and compare survival in people with complications and without complications.

© NICE 2022. All rights reserved. Subject to Notice of rights.

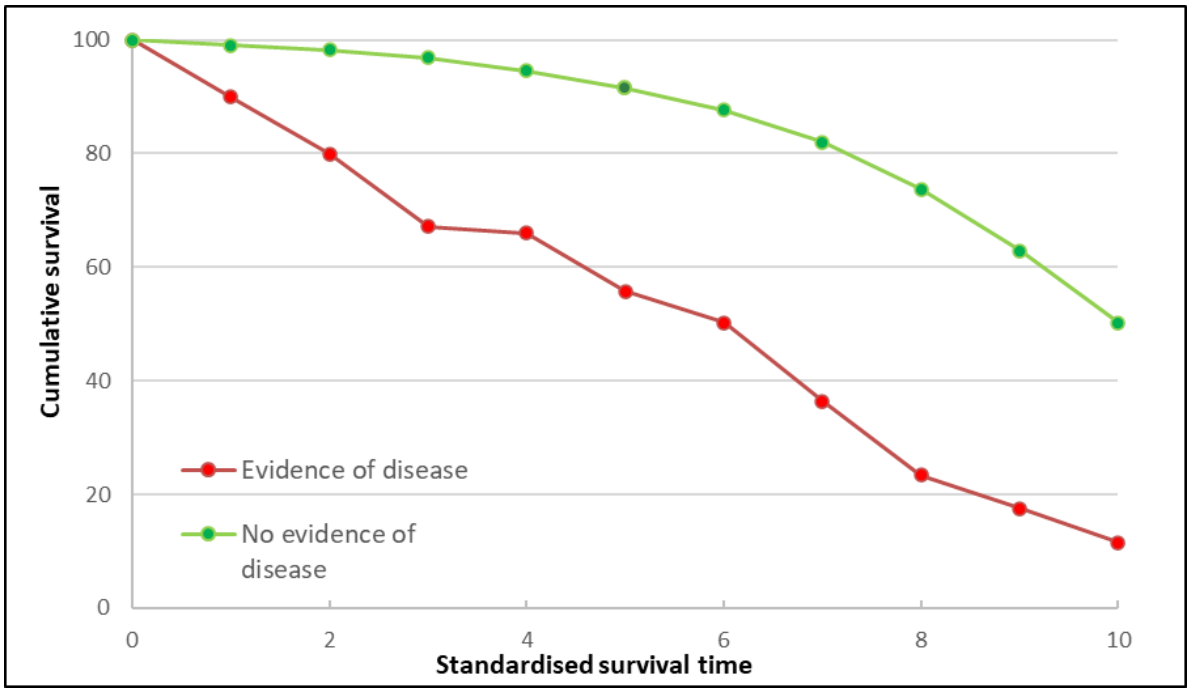
Figure 10: Survival in people with and without complications



2.3.7.3 Mortality with recurrent/persistent disease

Although survival with treated differentiated thyroid carcinoma is excellent and compared survival in the general population, recurrent or persistent disease is known to cause severe health complications that can reduce life expectancy. Link and colleagues[22] compared mortality in people who had cancer but no current evidence of disease and people showing either recurrent or persistent disease finding a significantly higher mortality in the latter group (see Figure 11).

Figure 11: Cumulative survival in people with and without evidence of disease



Source: Link 2005[22]

© NICE 2022. All rights reserved. Subject to Notice of rights.

The green curve representing survival in the disease-free group is fairly aligned to the survival of the general population, confirming that people with a treated cancer and no recurrence have no additional risk of mortality. On the other hand, the red curve representing survival in people with evidence or recurrent or persistent disease, is considerably steeper and, by the end of last follow-up, 88% of them are not alive compared with the 50% in the group with no disease.

The methodology described by Williamson and colleagues[50] to study time-to-event outcomes were used to estimate a hazard ratio (HR) of mortality in people with evidence of disease compared to people without evidence of disease. As the original study did not provide information on censoring, this was assumed to be constant between the beginning and the end of the study. The hazard ratio was estimated to be 3.01 (2.1 to 4.3) suggesting that people with evidence of disease die 3 times faster than people with no evidence. This hazard ratio was applied to the English general population mortality rates to estimate mortality on those developing recurrent or persistent disease.

There are two shortcomings in this approach that need to be mentioned. Firstly, the methodology presented by Williamson was developed to study Kaplan-Meier curves in randomized controlled trials (RCT) as it only allows to calculate unadjusted HRs. Link 2005[22], naturally, was an observational study where patients were not randomly assigned and may have potential confounders or differences in baseline causing a higher mortality in one of the groups. Yet, the authors used a time transformation (standardised survival time) to adjust for age at diagnosis, implying that the HR calculated from Figure 11 is not biased by different age in the two groups. Although other confounders, such as initial tumour status or time to surgery may still exist, this was not considered problematic for this analysis, which is trying to determine real-world mortality in people showing evidence of disease and it is assuming already that they would have different characteristics and confounding factors than the disease-free population (e.g. a delay diagnosis). A second limitation is caused by the shape of the red curve in figure Figure 11, which is not a linear transformation of the original curve as people showed to die at a much higher rates during the first time period. The obvious solution of dividing the curve into two or more curves to calculate different HRs could not be adopted as the standardised survival time could not be converted into Markov model cycles. This suggests that, although the model correctly predicts long-term survival of people with disease, short-term survival could be overestimated as in the original curve most of the deaths occur early (see section 4.2).

2.3.8 Utilities

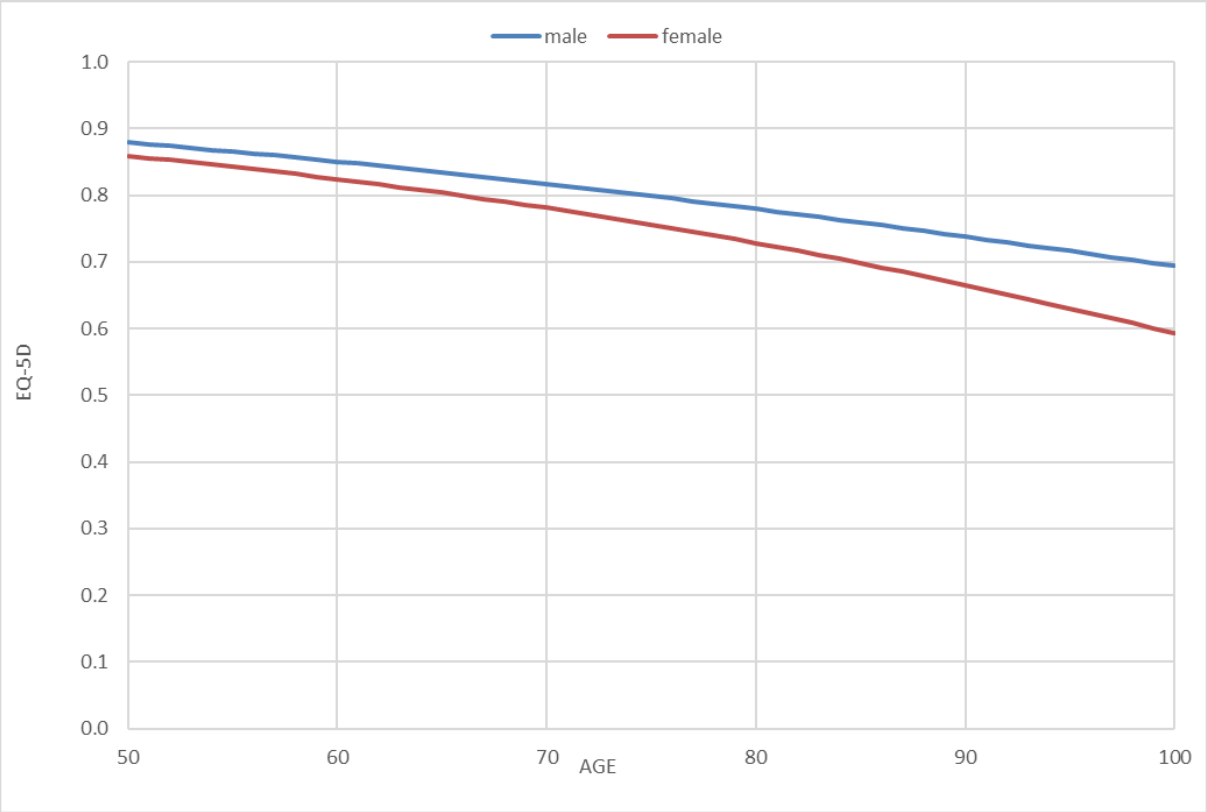
2.3.8.1 Utility with no complications or recurrence

People with benign nodules and people with malignant nodules who developed no complication or recurrent disease have a quality of life comparable to the general population. This was confirmed by the committee who acknowledged that even people with hypothyroidism managed with levothyroxine (LT4) usually compare well with the general population in terms of quality of life.

EQ-5D utility scores in the general population was estimated using the Adjusted Limited Dependent Variable Mixture Models (ALDVMM) developed by Hernández Alava and colleagues[9] and based on the Health Survey for England 2014. The model was recently developed by the Decision Support Unit (DSU) of NICE and it is expected to reflect more realistically quality of life of English population (see Figure 12).

© NICE 2022. All rights reserved. Subject to Notice of rights.

Figure 12: EQ-5D in males and females estimated with the ALDVMM model



As quality of life is different by gender, with women usually reporting a lower score than men, the gender split overtime was used to calculate the weighted-average quality of life of the population in the model considering how the split changes overtime.

2.3.8.2 Utility with surgical complications or recurrence

Utility scores in people with surgical complications were sought from the published literature and models obtained through a systematic review. Kebebew and colleagues[14] used a survey to ask to assign a numerical value for each health outcomes based on an anchor states of 1 for perfect health and 0 for being dead.

The values given to each health states are illustrated in Table 14.

Table 14: Utility of each health outcome state

Health state	Mean utility	Range of utilities
Well	1	-
Dead	0	-
Unilateral RLN injury	0.627	0.10 – 0.95
Hypoparathyroidism	0.778	0.20 – 0.98
Recurrence	0.54	0.03 – 0.60

Source: Kebebew 2000[14]. Note: in the sensitivity analysis a standard error equal to 20% of the mean was assumed

The values reported in the study were used as relative disutility factors and multiplied for the general population utility scores in Figure 12 to calculate utility in each health state. This represents a deviation from NICE reference case, as EQ-5D is generally preferred to other surveys when estimating disutility factors as well. However, no other alternative data was found from the systematic review on quality of life.

© NICE 2022. All rights reserved. Subject to Notice of rights.

RLN and permanent hypoparathyroidism were assumed to be permanent states and as such, their decrement in quality of life is assumed to last for the lifetime. Post-surgery transient hypoparathyroidism was assumed to last for 2 months only as recommended by the Committee.

The QoL discounting factor for recurrence was applied for lifetime in the base-case scenario, but this assumption was tested in the sensitivity analysis, where people were assumed to incur in a QoL harm only during the first year of recurrence, when they receive the treatment (RAI in the base case scenario). See section 2.4 for further discussion on the scenario analysis.

2.3.9 Resource use and costs

2.3.9.1 Prices of molecular tests

The cost of the four molecular tests included in the analysis were obtained directly from the manufactures and, when not possible, from an external source[17]. Prices were originally in US dollar and were converted in UK Pounds using 2022 exchange rates. Permission was asked to publish the prices in this report quoting the manufactures (see Table 15).

Table 15: Price of molecular tests

	Manufacturer	Kuo 2019[17]
Afirma-GSC	Retail price[(a)] = £4,798 Medicare[(a)] = £2,660 Cash pay[(a)] = £2,918	£4,454
ThyroSeq V3	International tariff[(b)] = £1,407	£2,227
ThyGenX + ThyraMIR	n/a	£2,245
ThyroSeq V1	n/a	n/a

(a) Varacyte 19/01/2022
(b) University of Pittsburgh Medical Center, 11/01/2022

Veracyte provided three different prices that were all tested in the sensitivity analysis with the Medicare price used in the base case scenario as it represents the most realistic price that would be negotiated with the NHS. It is worth mentioning that the price provided for Afirma includes the cost of the Xpression Atlas (XA), which is a further RNA sequencing test often offered in combination with Afirma to people with suspicious findings during Afirma-GSC. However, data on the accuracy of XA was not available and the price of Afirma-GSC without XA could not be calculated. Therefore, it is possible that the accuracy of Afirma-GSC with XA is underestimated by the model although this could not be verified with published evidence.

The cost of ThyroSeq V1 could not be found as this is an outdated test which is not in commerce anymore that was included only to mimic current tests available within the NHS. Therefore, the same cost of ThyroSeq V3 was applied. With the exception of ThyroSeq V1, all the other tests are assumed not to be locally available and, therefore, an additional cost associated with packaging and shipping the sample abroad was applied (see Table 16). The cost of a box complying with UN3373, biological substances shipped by air, was collected from the Royal Mail website[42] and the cost of shipping a light package with UPS express from London to the Pittsburgh laboratories performing ThyroSeq V3 was applied.

Table 16: Cost of packaging and

	Cost	Source
Packaging cost	£8	Box UN3373. Royal Mail[42]

© NICE 2022. All rights reserved. Subject to Notice of rights.

	Cost	Source
Shipping cost	£107	UPS Express® 2ibs from London to Pittsburgh laboratories[48]

2.3.9.2 FNAC and CNB

The cost of fine-needle aspiration (FNA) and core-needle biopsy (CNB) were both taken from the latest NHS Reference Costs 2019-2020[31] and illustrated in Table 17. The model assumes these are ultrasound-guided tests to reflect current practice in England and the available costs in the NHS Reference Costs reports.

Table 17: Cost of FNAC and CNB

Test	HRG code	Cost
Ultrasound-guided FNAC	YC02Z	£299
Ultrasound-guided CNB	YC01Z	£429

Source: NHS Reference Costs 2019-2020[31]

2.3.9.3 Thyroid surgery

The cost of a thyroidectomy was sought from the latest NHS Reference Costs 2019-2020[31]. The Reference Costs include three different HRGs for thyroid procedures, each related to a different CC score which captures the comorbidity and complications of patients (see Table 18).

Table 18: NHS Reference Costs for Thyroid Procedures

Currency Code	Currency Description	Average Unit Cost
KA09C	Thyroid Procedures with CC Score 4+	£6,418
KA09D	Thyroid Procedures with CC Score 2-3	£5,002
KA09E	Thyroid Procedures with CC Score 0-1	£2,954

Source: NHS Reference Costs 2019-2020[31]

The main difference in cost in hemithyroidectomy, completion thyroidectomy and total thyroidectomy is caused by length of stay, as people receiving hemithyroidectomy are usually dismissed the following day whereas people receiving completion and total thyroidectomy are often required to spend an additional day in the hospital ward. To calculate the cost of each procedure, NHS Reference Costs were firstly transformed into "procedural costs" by removing the cost associated with length of stay. This was achieved by using data on average length of stay and excess bed day cost (inflated to 2019-2020) from a previous version of the Reference Cost[30].

$$Procedural\ cost = NHS\ Ref\ Cost - aLOS \times Excess\ bed\ day\ cost$$

A weighted average across all the CC scores category was calculated as an estimation of the average cost of a thyroid procedure in England (see Table 19).

© NICE 2022. All rights reserved. Subject to Notice of rights.

Table 19: Calculation of procedural cost

Currency description	Average Unit Cost for HRG	Average length of stay	Excess bed day cost[(a)]	Procedural cost[(b)]
Thyroid Procedures with CC Score 4+	£6,418	10.24	£400	£2,323
Thyroid Procedures with CC Score 2-3	£5,002	3.76	£589	£2,789
Thyroid Procedures with CC Score 0-1	£2,954	3.26	£482	£1,384
Weighted average	-	-	-	£1,622

Source: NHS Reference Costs 2019-2020[31] and NHS Reference Costs 2017-2018[30].

(a) Excess bed day costs were inflated to 2019-2020 using the change in HRG cost between 2017 and 2019 as inflator
(b) Procedural cost = average unit cost – average LOS x excess bed day cost

Once the procedural cost was known, cost of each procedure was calculated by adding the corresponding LOS cost using average LOS reported in the BAETS audit[46] (see Table 12 in section Thyroid surgery2.3.5). The final estimations with relevant international comparison are presented in Table 20.

Table 20: Cost of each thyroid procedure

Procedure	UK (NICE)	Canada[20]	South Korea[15]
Total thyroidectomy	£2,515	£2,546	£2,662
Hemi thyroidectomy	£2,173	£2,388	£2,252
Completion thyroidectomy	£2,254	£2,476	-

There were some limitations with this methodology. Firstly, HRG procedure cost may not reflect the true cost as it likely includes more complex thyroid interventions as ipsilateral or bilateral neck dissection or mediastinal dissection. This explains why average length of stay for each CC score category was higher than the length of stay reported in the BAETS audit[46]. Furthermore, as excess bad day cost and LOS are not collected anymore in the NHS Reference Costs, an older version from 2018[30] had to be used which introduces bias due to costs collected in different years. Nevertheless, the estimation of the cost of each procedure aligns quite well with the estimations from other countries with comparable healthcare systems and cost.

2.3.9.4 Surgical complications

Surgery can cause a variety of complications but only some were identified as potential sources of additional cost. These were hospital readmission, hypoparathyroidism and RLN injury.

Hospital readmissions are assumed to happen shortly after the first admission and are usually caused by symptoms of other complications like hypoparathyroidism[46]. However, hypoparathyroidism and readmission were treated as separate events and hospitalisation costs were removed from the costs associated with the other complications to avoid double counting. The cost of a hospital admission was sought from the NHS Reference Costs 2019/2020[31] under the HRG KA08 as this code include people treated for symptoms of hypoparathyroidism. A weighted average of £1,184 was calculated and used as the unit cost for a readmission (see Table 21: Hospital re-admission costTable 21)

© NICE 2022. All rights reserved. Subject to Notice of rights.

Table 21: Hospital re-admission cost

HRG code	HRG description	Total HRG cost
KA08A	Other Endocrine Disorders with CC Score 4+	£1,621
KA08B	Other Endocrine Disorders with CC Score 2-3	£735
KA08C	Other Endocrine Disorders with CC Score 0-1	£535
Weighted average		£1,184

Source: NHS Reference Costs 2019/2020[31]

Although a hospital re-admission is a separated procedure, not included in the HRG used for surgery, this is less certain for laryngoscopy, which is often used to diagnose RLN on people who experience voice change after surgery. In some cases, laryngoscopy is performed pre-procedurally too. The cost of a laryngoscopy was taken from the NHS Reference Costs but added separately only in the sensitivity analysis, as assumed to be already included in the HRG in the base case scenario (see section 2.4). The cost of laryngoscopy was informed by the NHS Reference Cost 2019/2020 HRG CA69A relative to diagnostic laryngoscopy or pharyngoscopy in people older than 19.

Hypoparathyroidism can be either short and long-term consequence of surgery and require supplements of calcium and vitamin D to be maintained at asymptomatic level. Dosages and cost per mg were collected from the British National Formulary (BNF)[13] whereas the Prescription Cost Analysis (PCA)[11] database was used to calculate the average price across all preparations and drugs (see Table 22Table 24).

Table 22: Hypoparathyroidism pharmacological cost

Class	Drug	Dosage	Cost of drug (daily)	Cost of class (daily)
Calcium supplement	Calcium carbonate	1500 mg	£0.11	£0.11
Vitamin D supplement	Calcitrol	1 mcg	£0.65	£0.39
	Alfacalcidol	0.75 mcg	£0.32	

Source: BNF[13] *and PCA*[11]

Temporary hypoparathyroidism is assumed to last for 2 months only as recommended by the committee. Permanent hypoparathyroidism, instead, will require lifelong pharmacological treatment.

Finally, RLN is not associated with a specific cost but, as mentioned before, increases annual risk of pneumonia. The cost of treating pneumonia was sought from the NHS Reference Costs 2019/2020 under HRG DZ11 and, as before, a weighted average across all CC scores was calculated (see Table 23).

Table 23: Pneumonia cost

HRG code	HRG description	Total HRG cost
DZ11K	Lobar, Atypical or Viral Pneumonia, with Multiple Interventions, with CC Score 14+	£8,350
DZ11L	Lobar, Atypical or Viral Pneumonia, with Multiple Interventions, with CC Score 9-13	£6,269
DZ11M	Lobar, Atypical or Viral Pneumonia, with Multiple Interventions, with CC Score 0-8	£4,193

© NICE 2022. All rights reserved. Subject to Notice of rights.

HRG code	HRG description	Total HRG cost
DZ11N	Lobar, Atypical or Viral Pneumonia, with Single Intervention, with CC Score 13+	£4,943
DZ11P	Lobar, Atypical or Viral Pneumonia, with Single Intervention, with CC Score 8-12	£3,628
DZ11Q	Lobar, Atypical or Viral Pneumonia, with Single Intervention, with CC Score 0-7	£2,696
DZ11R	Lobar, Atypical or Viral Pneumonia, without Interventions, with CC Score 14+	£3,031
DZ11S	Lobar, Atypical or Viral Pneumonia, without Interventions, with CC Score 10-13	£2,141
DZ11T	Lobar, Atypical or Viral Pneumonia, without Interventions, with CC Score 7-9	£1,561
DZ11U	Lobar, Atypical or Viral Pneumonia, without Interventions, with CC Score 4-6	£1,252
DZ11V	Lobar, Atypical or Viral Pneumonia, without Interventions, with CC Score 0-3	£921
Weighted average		£1,909

Source: NHS Reference Costs 2019/2020[31]

2.3.9.5 Radioactive Iodine Ablation (RAI)

People who received a total thyroidectomy or a completion thyroidectomy usually undergo radioactive iodine ablation (RAI) to destroy all the remaining thyroid tissue. A radioactive substance called iodine-131 is taken as a capsule and, as thyroid cells use iodine to make thyroid hormones, they absorb the radioactive substance and are destroyed together with any remaining cancerous cells. This is expected to decrease lifetime risk of recurrent and persistent disease.

In the base case scenario, all people with a malignant nodule receive a completion thyroidectomy and, therefore, all of them undergo RAI. RAI cannot be given after a hemithyroidectomy as the rationale for a hemithyroidectomy is to preserve part of the thyroid gland to avoid hypothyroidism and the need of thyroid hormone replacement therapy. In the scenarios where a proportion of people with a malignant nodule is treated with a hemithyroidectomy only, these would not have to undergo RAI after.

The cost of RAI treatment is composed of two different cost. The first is the cost of the actual treatment (medical check-up, cost of Iodine-131 and hospital stay) that was sought from the NHS Reference Costs 2019/2020[31]. The second cost is associated with thyroid hormone stimulation. To increase Iodine-131 uptake, a high level of thyroid stimulating hormone (TSH) is required. This was historically achieved by suspending levothyroxine (T4) treatment and forcing a state of hypothyroidism but, in recent years, TSH has been artificially created in a laboratory and given to patients through an intramuscular injection. This artificial TSH is called recombinant human TSH (rhTSH) or Thyrotropin Alfa. The cost of the two doses needed was taken from the BNF[13] and added to the procedural RAI cost to calculate the total cost associated with RAI (see Table 24).

Table 24: Cost of RAI

Item	Cost	Source
Recombinant human TSH (2 doses)	£583	BNF[13]
Radiotherapy for Thyroid ablation	£535	HRG RN51Z NHS Reference Costs 2019/2020[31]

© NICE 2022. All rights reserved. Subject to Notice of rights.

RAI was assumed to be the treatment of choice in the base case scenario for any episode of recurrence, and therefore it is administered to those developing persistent or recurrent disease. In a scenario analysis, salvage surgery is offered alongside RAI after a recurrence.

RAI is not known to cause any impact on quality of life if given after rhTSH instead than withdrawal and therefore no QoL detriment was applied. In the sensitivity analysis, people were assumed to undergo thyroid hormone withdrawal instead and a disutility value estimated from the TSH model was applied (see section 2.4).

2.3.9.6 Thyroid Hormone Therapy

All people with a total/completion thyroidectomy and a proportion of people with hemithyroidectomy will need to take manmade thyroid hormone to raise abnormally low levels of natural thyroid hormone in the body. This is usually given in pill form and helps to avoid hypothyroidism which can cause severe quality-of-life impairments. In addition, people who underwent total or completion thyroidectomy are often treated with an excessive dose of thyroid hormone usually for a period of 10 years to suppress TSH and reduce the risk of recurrence after surgery.

The cost of thyroid hormone replacement and TSH suppression therapy was sought from the BNF[13]. There are two common drugs used to raise the level of thyroid hormone in the blood: levothyroxine (T4) and liothyronine (T3). Historically, both were provided by the NHS but, in recent years, the number of prescriptions of liothyronine has sharply decreased following an increase of more than 6,000% of the original price set by the manufacturer[2]. The Competition and Markets Authority (CMA) imposed a fine to the pharmaceutical company in 2021 for "excessive and unfair prices" and the cost per tablet is expected to decrease to international levels in the future. However, to avoid any complication and bias induced by a contested price, liothyronine (T3) was taken out from the analysis and only levothyroxine (T4) was included as a therapy for hormone replacement. This should not cause any bias as T4 is the most prescribed drug for thyroid hormone replacement anyway, and the prices of T3 and T4 are very similar in almost every country in the world except the UK.

Cost per mg of T4 was taken from the BNF and the PCA database was used to determine the average cost across all the formulations. TSH suppression and maintenance therapies require different dosages per kg that were sought from the available literature[3] and used to determine average dosage in men and women using English weight data from the Office of National Statistics (ONS) (see Table 25).

Table 25: Dosages and costs per day of suppression and maintenance therapy with T4

Therapy	TSH target	Dosage (mg/kg)	Dosage (mg)	Cost per day
Suppression therapy	mU/L= 0.1	0.0034	0.25	£0.24
Maintenance therapy	mU/L= 0.5 – 6.2	0.0016	0.12	£0.11

Source: BNF[13] and PCA[11]

2.3.9.7 Monitoring cost

After being discharged, people with a malignant nodule need to be monitored for a certain period of time to assess and monitor their risk of recurrences. The duration of monitoring depends on the risk of the person defined with the dynamic risk stratification (DRS) framework. People with a low risk are usually followed up for 2 years, people with medium risk for 5 risks and people with high risk for 10 years. People in this analysis are likely to be predominantly low risk as their cytology at first FNAC was not clearly suggestive of malignancy. Therefore, in the base case scenario, monitoring is assumed to occur for the first 2 years and the 5- and 10-years scenarios are both tested in a sensitivity analysis.

© NICE 2022. All rights reserved. Subject to Notice of rights.

There is a range of tests that are often used to monitor people who had surgery for thyroid cancer. Thyroglobulin (tg) and Thyroglobulin antibody tests are often performed twice a year immediately after surgery for two years, and once a year thereafter.

TSH test is often performed once a year and may be either offered after thyroid hormone stimulation or without it. In the first case, an injection of thyrotropin alpha (rhTSH) is required making this test extremely expensive. In the base case scenario TSH test is assumed to be non-stimulated and the version with rhTSH is tested in a sensitivity analysis.

Finally, it is assumed that, for the duration of monitoring, a yearly attendance to an endocrinologist and an ultrasound scan without contrast will be required to clinically assess and check for any remaining lump in the neck. Table 26 illustrates all the tests and costs of monitoring included in the model.

Table 26: Cost of monitoring after thyroid surgery

Test	Cost	Source	Frequency
Thyroglobulin	£16	NICE thyroid disease guideline[28]	Twice a year for the first 2 years, once a year thereafter
Tg antibody	£16	NICE thyroid disease guideline[28]	Twice a year for the first 2 years, once a year thereafter
TSH test	Non-stimulated: £2 Stimulated: £585	NICE thyroid disease guideline[28]	Once a year
Ultrasound scan without contrast	£52	NHS Reference Costs 2019/2020[31]	Once a year
Endocrinology attendance	£151	NHS Reference Costs 2019/2020[31]	Once a year
Total[(a)]	First 2 years: £268 >2 years: £237		

(a) Assuming a non-stimulated TSH test

This cost is fairly aligned to the estimation of monitoring cost in the US by Lee 2014 which is equal to £233[20].

People who had a recurrence are assumed to have the strictest monitoring (two years) and to be followed for the rest of their life. They are also assumed to receive a TSH suppression dosage of T4 for life.

2.4 Sensitivity analyses

In addition to the probabilistic sensitivity analysis, a range of one-way sensitivity analyses were undertaken. These are shown in Table 27, where the scenarios used in the base case scenario are highlighted in green:

Table 27: Scenario analyses

Feature	Scenarios	Description
Cancer prevalence	RCPath 2019	Based on RCPath meta-analysis
	BAETS audit	Based on BAETS audit
	Bongiovanni 2012	Based on Bongiovanni meta-analysis

© NICE 2022. All rights reserved. Subject to Notice of rights.

Feature	Scenarios	Description
Proportion of hemithyroidectomy with malignancy needing completion	100%	All malignant HTs require CT
	50%	Half malignant HTs require CT
	0%	None of the malignant HTs require CT
BAETS surgery parameters	Use repeat lobectomy for CT	Use “repeat lobectomy” outcomes from BAETS 2021 for CT
	Use total lobectomy for CT	Use “total thyroidectomy” outcomes from BAETS 2021 for CT
Disutility with recurrence	Life-long	Loss of QoL due to recurrence disease is lifelong
	Only first year	Loss of QoL due to recurrence disease lasts for 1 year only
Molecular test in the selective use strategies	ThyroSeq V3	ThyroSeq V3 is the test offered in the strategies with selective use of molecular test
	ThyroSeq V1	ThyroSeq V1 is the test offered in the strategies with selective use of molecular test
Price of molecular tests	From manufactures	When available, prices given to use from the same manufactures are used
	2nd FNAC necessary	Add the cost of a 2nd FNAC to the price
	From Kuo 2019	Use Kuo 2019 study for costs
Price of Afirma-GSC	Retail price	Use retail price of Afirma GSC
	Medicare price	Use Medicare price of Afirma GSC
	Cash price	Use cash price of Afirma GSC
RAI preparation	rhTSH	People preparing from RAI receive rhTSH
	Thyroid hormone withdrawal	People preparing from RAI undergo hormone withdrawal
TSH suppression duration	Life-long	TSH suppression for malignant nodules is lifelong
	10 years	TSH suppression lasts for 10 years
TSH test	Non-stimulated	TSH test is performed without stimulation (rhTSH)
	Stimulated	TSH is performed after stimulation (rhTSH)
Length of monitoring	2 years (low risk)	Monitoring lasts for the first 2 years
	5 years (medium risk)	Monitoring lasts for the first 5 years
	10 years (High risk)	Monitoring lasts for the first 10 years
Proportion of people with recurrence needing surgery	0%	Nobody needs salvage surgery after recurrence

© NICE 2022. All rights reserved. Subject to Notice of rights.

Feature	Scenarios	Description
	50%	50% need salvage surgery after recurrence
	100%	Everyone needs salvage surgery after recurrence

2.4.1 Prevalence

In the base scenario, cancer prevalence in the different RCPath categories was estimated using a meta-analysis from the Royal College of Pathologists[38]. Two alternative sources[5, 46] were available and were tested in the scenario analysis.

2.4.2 Completion thyroidectomy

All people with a malignant nodule whose diagnosis is confirmed by a hemithyroidectomy are assumed to need a completion thyroidectomy to fully remove the tumour, followed by a RAI later. This was considered a reasonable assumption as, within the model pathway, people always undergo a diagnostic hemithyroidectomy before receiving a therapeutic thyroidectomy or being discharged. However, it is possible to fully remove the entire cancerous mass during the first hemithyroidectomy, thus avoiding the need of a completion thyroidectomy. This is quite important from a health economics perspective, as people with hemithyroidectomy would not need to undergo further expensive therapies, such as RAI, and in the majority of cases do not need thyroid hormone replacement therapy either. Hence, two scenarios where 50% and 0% of people with malignant nodules were treated with a hemithyroidectomy only were tested to assess the importance of this assumption on the results. The same recurrence rate was applied regardless of the type of surgery received: although this is expected for people with low-risk cancer, evidence is still lacking and an upcoming trial (HoT) will hopefully help shed light on this in the future.

2.4.3 BAETS parameters

BAETS audit reported parameters for the procedure “repeat lobectomy” and “total thyroidectomy”[46]. “Repeat lobectomy” was identified as the category more likely including completion thyroidectomy and so its parameters were used in the base case scenario. However, as this was not completely certain, a further scenario analysis was conducted using the parameters of “total thyroidectomy” instead.

2.4.4 Disutility with recurrence

In the base case scenario, the disutility for recurrence estimated by Kebebew 2000[14] was applied for the entire lifetime, as evidence found that complications related to persistent or recurrent disease are often long-term. However, this assumption was tested in the sensitivity analysis where the disutility was applied only during the first year until people received treatment for recurrence (RAI in the base case scenario). Beyond the first year, people with recurrence are assumed to return to general population quality of life even though their mortality remain higher reflecting the findings of Link and colleagues[22].

2.4.5 Molecular test in the selective use of molecular testing strategies

In the base case scenario ThyroSeq V3 is the test offered in the two strategies with selective use of molecular testing in the Thy3a population: FNAC and selective use of molecular testing and CNB and selective use of molecular testing. This was decided because ThyroSeq V3 was found to the most effective and cheapest test on the market. However, ThyroSeq V3 is only available abroad and would require shipping the material overseas, which may be logistically complex at least in the short-term. ThyroSeq V1, although outdated and

© NICE 2022. All rights reserved. Subject to Notice of rights.

outclassed by modern tests, mimics what the NHS can offer locally and could represent the only available test before more modern equipment is purchased or agreement between NHS and manufacturers is finalized. Therefore, a scenario analysis where this test is offered in the two strategies with selective use of molecular testing was conducted.

2.4.6 Cost of molecular tests

Although a full threshold analysis on the price of each molecular test was conducted, different prices were tested in the scenario analyses as test prices were identified as a critical aspect of the analysis.

In the base case scenario, the prices provided by the manufactures were utilized. For ThyroSeq V3, the manufacturer's price is the international tariff applied to all patients living outside the US which likely reflects the price NHS patients would be charged. For Afirma, three different prices were available: in the base case scenario the price applied to the main public payer in US, Medicare, was used as this may reasonably represent the price the company would charge another public payer, like the NHS. The other two prices, the retail price, which is generally charged by distributors, and the cash price, which is charged by the company itself to private buyers, were both tested in the scenarios analysis.

In a further scenario analysis, costs obtained by a published analysis, Kuo 2019[17], were used instead of manufacturers' price. This analysis estimated the price of molecular tests in US dollars likely from the manufacturers as well. The company producing ThyGenX + ThyraMIR did not provide their price, so Kuo cost estimation of this test was used in the base case analysis.

Finally, although in most cases the same sample collected from previous tests can be used for further molecular testing, a scenario where an additional FNAC is necessary before the molecular test was tested. This was expected to be unlikely to occur if people previously received a CNB, as a CNB is able to extract a sample large enough for further analysis. However, it is not uncommon that a single FNAC does not have enough material for a molecular test, prompting the test to be repeated for further molecular analyses.

2.4.7 RAI preparation

Reflecting current and established practice in the UK, the base case scenario assumes that people would receive recombinant human TSH (rhTSH) prior to ablation with iodine-131.

In the scenario analysis we tested the assumption that people undergo thyroid hormone withdrawal instead. These two strategies were analysed in the TSH model published alongside the guideline, and all the relevant inputs, such as costs and disutility associated with rhTSH or withdrawal, were taken from the TSH model.

2.4.8 TSH suppression duration

In the base case analysis TSH suppression on people with a treated malignant nodule was expected to last for 10 years, reflecting the recommendation of the committee. In a further analysis we tested the assumption of a lifetime TSH suppression, which is only rarely recommended for people with an extremely high risk of recurrent. Lifetime TSH suppression was included as the base case monitoring approach for people who had a recurrent/persistent disease episode.

2.4.9 TSH test

TSH test, which is a relatively cheap test, can be either done without any TSH stimulation or after TSH stimulation using rhTSH (thyrotropin alpha). In the base case scenario, it was

© NICE 2022. All rights reserved. Subject to Notice of rights.

assumed that TSH test was done without prior stimulation, as it seems to reflect current practice. A scenario where stimulation is needed prior the test was tested instead in the sensitivity analysis and the cost of 2 doses of thyrotropin alpha was added to the cost of the test.

2.4.10 Length of monitoring

In the base case scenario, monitoring was assumed to last for 2 years, reflecting Committee's recommendation for people with low-risk cancer, who are the majority. Other durations of monitoring were tested in the scenario analysis: 5 years reflecting monitoring for people with medium risk and 10 years for people with high risk. The risk is generally calculated using the DRS (dynamic risk stratification) framework.

2.4.11 Salvage surgery for people with recurrence

In the base case scenario, people with recurrence are assumed to be treated through radioactive iodine ablation with iodine-131. In the sensitivity analysis we tested two scenarios where, instead, they need a further intervention (which is assumed to be a new completion thyroidectomy) to surgically remove the cancerous mass.

2.5 Model validation

The model was developed in consultation with the committee; model structure, inputs and results were presented to and discussed with the committee for clinical validation and interpretation.

The model was systematically checked by the health economist undertaking the analysis; this included inputting null and extreme values and checking that results were plausible given inputs. The model was peer reviewed by a second experienced health economist from the NGC; this included systematic checking of the model calculations.

2.6 Estimation of cost effectiveness

The widely used cost-effectiveness metric is the incremental cost-effectiveness ratio (ICER). This is calculated by dividing the difference in costs associated with 2 alternatives by the difference in QALYs. The decision rule then applied is that if the ICER falls below a given cost per QALY threshold the result is considered to be cost effective. If both costs are lower and QALYs are higher the option is said to dominate and an ICER is not calculated.

$$ICER = \frac{Costs(B) - Costs(A)}{QALYs(B) - QALYs(A)}$$

Where: Costs(A) = total costs for option A; QALYs(A) = total QALYs for option A

Cost effective if:
- ICER < Threshold

When there are more than 2 comparators, as in this analysis, options must be ranked in order of increasing cost then options ruled out by dominance or extended dominance before calculating ICERs excluding these options. An option is said to be dominated, and ruled out, if another intervention is less costly and more effective. An option is said to be extendedly dominated if a combination of 2 other options would prove to be less costly and more effective.

It is also possible, for a particular cost-effectiveness threshold, to re-express cost-effectiveness results in term of net monetary benefit (NMB). This is calculated by multiplying the total QALYs for a comparator by the threshold cost per QALY value (for example, £20,000) and then subtracting the total costs (formula below). The decision rule then applied is that the comparator with the highest NMB is the cost-effective option at the specified

© NICE 2022. All rights reserved. Subject to Notice of rights.

threshold. That is the option that provides the highest number of QALYs at an acceptable cost.

$Net\ Monetary\ Benefit(X) = (QALYs(X) \times \lambda) - Costs(X)$ *Where: λ = threshold (£20,000 per QALY gained)*	Cost effective if: • Highest net benefit

Both methods of determining cost effectiveness will identify exactly the same optimal strategy. For ease of computation NMB is used in this analysis to identify the optimal strategy.

2.7 Interpreting results

NICE sets out the principles that committees should consider when judging whether an intervention offers good value for money.[262725] In general, an intervention was considered to be cost effective if either of the following criteria applied (given that the estimate was considered plausible):

- The intervention dominated other relevant strategies (that is, it was both less costly in terms of resource use and more clinically effective compared with all the other relevant alternative strategies), or
- The intervention costs less than £20,000 per quality-adjusted life-year (QALY) gained compared with the next best strategy.

As we have several interventions, we use the NMB to rank the strategies on the basis of their relative cost effectiveness. The highest NMB identifies the optimal strategy at a willingness to pay of £20,000 per QALY gained.

© NICE 2022. All rights reserved. Subject to Notice of rights.

3 Results

The analysis was repeated three times to simulate three cohorts each with a different cytology at the first FNAC: Thy1, Thy3a and Thy3f. Not every strategy is included in all cohorts. Molecular testing is hardly justifiable after Thy1, so only repeat sampling and selective use of hemithyroidectomy were included for this cohort. Likewise, repeat sampling is rarely useful after a Thy3f, therefore only routine use of molecular testing and hemithyroidectomy were tested for Thy3f (see Table 1 in section 2.1.1).

In the following sections the mean probabilistic and deterministic results of each cohort are presented.

3.1 Non-diagnostic Thy1

For the category Thy1, three diagnostic strategies were included: repeat FNAC and selective use of hemithyroidectomy (current practice), CNB and selective use of hemithyroidectomy and routine hemithyroidectomy.

3.1.1 Base case analysis

Table 28 illustrates the probabilistic diagnostic outcomes in the base case scenario of each included strategy for a cohort of 1,000 people with Thy1 and an average ROM of 12%[38].

Table 28: Diagnostic performance – Thy1 (probabilistic)

Strategy	True positive	False positive	True negative	False negative
Repeat FNAC & HT	115 (47 to 207)	353 (298 to 418)	527 (452 to 592)	5 (2 to 10)
CNB & HT	113 (46 to 204)	177 (107 to 300)	703 (568 to 802)	7 (2 to 16)
Routine HT	120 (49 to 216)	880 (784 to 951)	0	0

Note: People with benign nodules undergoing hemithyroidectomy are considered false positives (avoidable surgery) even though the hemithyroidectomy would correctly identify the nodule as benign

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

False positives represent people who have received a diagnostic hemithyroidectomy even though they had a benign nodule and who were discharged consequently. Naturally, all people in the hemithyroidectomy arm received a diagnostic hemithyroidectomy so this arm has the largest number of false positives. Likewise, it is not possible for people in the hemithyroidectomy arm to be false negative (missed diagnosis) as both sensitivity and specificity were assumed to be 100% with surgery.

As the tables shows, a CNB effectively halves the number of false positives compared to the current practice of repeat FNAC, due to the lower number of inconclusive results required to undergo surgery. This, in turn, significantly reduces the number of hemithyroidectomies carried out on benign nodules. Table 29 and Figure 13Figure 13: illustrate the cost-effectiveness results of the analysis.

Table 29: Cost-effectiveness results – Thy1 (probabilistic)

Strategy	Cost per patient	QALYs per patient	NMB (£20,000 per QALY)	Probability ranked 1st (£20,000 per QALY)
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%

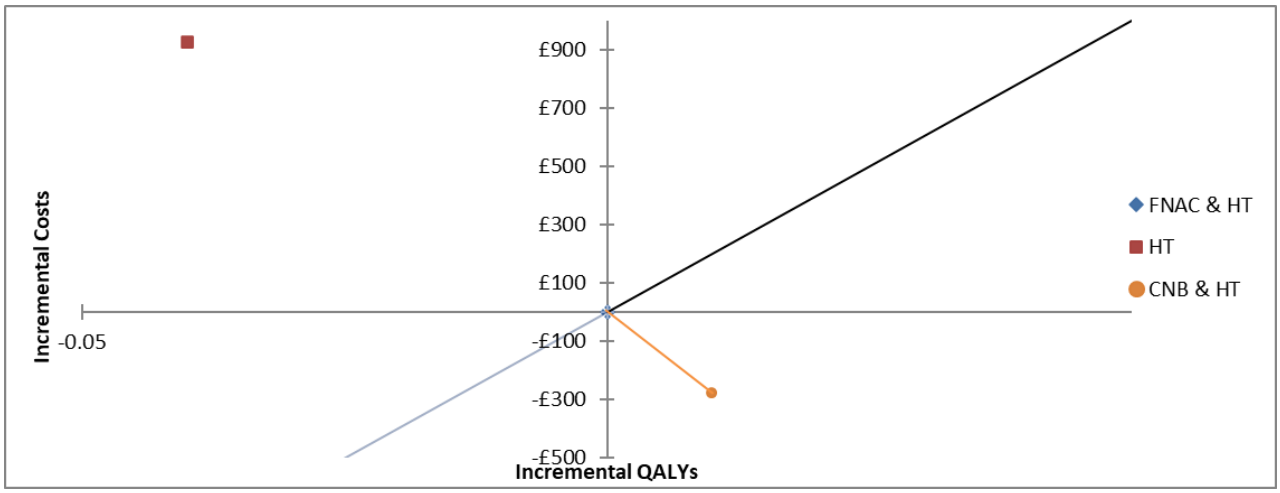
© NICE 2022. All rights reserved. Subject to Notice of rights.

Strategy	Cost per patient	QALYs per patient	NMB (£20,000 per QALY)	Probability ranked 1st (£20,000 per QALY)
CNB & HT	£1,815 (£1,223 to £2,600)	17.18 (16.81 to 17.36)	£341,874 (£333,729 to £345,965)	98%
Routine 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

CNB was found to dominate the other two strategies, being both cheaper and more effective. The reduced number of hemithyroidectomies needed for the management of further inconclusive cytology, reduced the cost per patient of around £300 compared to repeat FNAC and improves quality of life by reducing surgery complication. The probability that CNB is the most cost-effective strategy among those included is 99%, with current practice being the most cost-effective in only 2% of the Monte Carlo simulations. Hemithyroidectomy is undoubtedly not cost-effective due to the very low prevalence of cancer among people with Thy1.

Figure 13: Cost-effectiveness plane (Thy1)



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

Note: All the strategies are compared with current practice FNAC & HT which lies in the origin of the axes

3.1.2 Scenario analyses

All the scenarios illustrated in Table 27 in section 2.4 were run deterministically to see if the rank of the strategies changes by when one or more parameters is varied. The ranking did not change in any of these scenarios: CNB remained the most cost-effective strategy followed by repeat FNAC and, lastly, hemithyroidectomy.

3.2 Indeterminate Thy3a

For the indeterminate category Thy3a, all the diagnostic strategies were included:

- routine molecular testing,
- repeat FNAC and selective use of hemithyroidectomy (current practice),
- repeat FNAC and selective use of molecular testing,

© NICE 2022. All rights reserved. Subject to Notice of rights.

- CNB and selective use of hemithyroidectomy,
- CNB and selective use of molecular testing and
- routine hemithyroidectomy.

3.2.1 Base case analysis

Table 30 illustrates the probabilistic diagnostic outcomes in the base case scenario of each included strategy for a cohort of 1,000 people with Thy3a and an average risk of malignancy of 25%[38].

Table 30: Diagnostic performance – Thy3a (probabilistic)

Strategy	True positive	False positive	True negative	False negative
Afirma-GSC	228 (168 to 289)	218 (147 to 299)	532 (445 to 616)	23 (4 to 68)
ThyroSeq V3	223 (162 to 284)	113 (68 to 166)	637 (569 to 702)	27 (6 to 74)
ThyGenx/ThyraMIR	225 (138 to 293)	156 (69 to 269)	593 (476 to 697)	25 (2 to 111)
ThyroSeq V1	204 (113 to 275)	262 (137 to 404)	488 (347 to 618)	46 (7 to 134)
FNAC & HT	241 (191 to 297)	318 (274 to 367)	432 (378 to 479)	9 (5 to 16)
FNAC & MT[(a)]	232 (182 to 289)	110 (74 to 156)	640 (575 to 699)	18 (9 to 35)
CNB & HT	238 (188 to 294)	237 (179 to 330)	513 (417 to 583)	12 (5 to 23)
CNB & MT[(a)]	233 (183 to 289)	105 (55 to 197)	645 (547 to 719)	18 (8 to 32)
HT	250 (198 to 309)	750 (691 to 802)	0	0

Note: People with benign nodules undergoing hemithyroidectomy are considered false positives (avoidable surgery) even though the hemithyroidectomy would correctly identify the nodule as benign

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

Hemithyroidectomy for all, FNAC and selective use of hemithyroidectomy and CNB and selective use of hemithyroidectomy are the strategies that are more effectively able to identify a malignant nodule. Likewise, all the molecular tests, except the outdated ThyroSeq V1, show a good performance with the number of missed diagnoses ranging from 20 to 30 out of 250 cases of cancer. In terms of avoiding surgery on people with a benign nodule, the most effective strategies are FNAC and selective use of molecular testing, CNB and selective use of molecular testing and the molecular test ThyroSeq V3 alone. This is explained by the very high specificity of ThyroSeq V3 which is also the test assumed to be offered in the two histological strategies with selective use of molecular test. Compared to the two histological strategies with hemithyroidectomy, offering ThyroSeq V3 instead can prevent around 100 to 200 surgeries on people with a benign nodule.

Table 31 and Figure 14 illustrate the cost-effectiveness results for this population.

Table 31: Cost-effectiveness results – Thy3a (probabilistic)

Strategy	Cost per patient	QALYs per patient	NMB (£20,000 per QALY)	Probability ranked 1st (£20,000 per QALY)
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%

© NICE 2022. All rights reserved. Subject to Notice of rights.

Strategy	Cost per patient	QALYs per patient	NMB (£20,000 per QALY)	Probability ranked 1st (£20,000 per QALY)
ThyGenx/ThyraMIR	£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	£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	£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; NMB = net monetary benefit; QALY = quality-adjusted life year

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

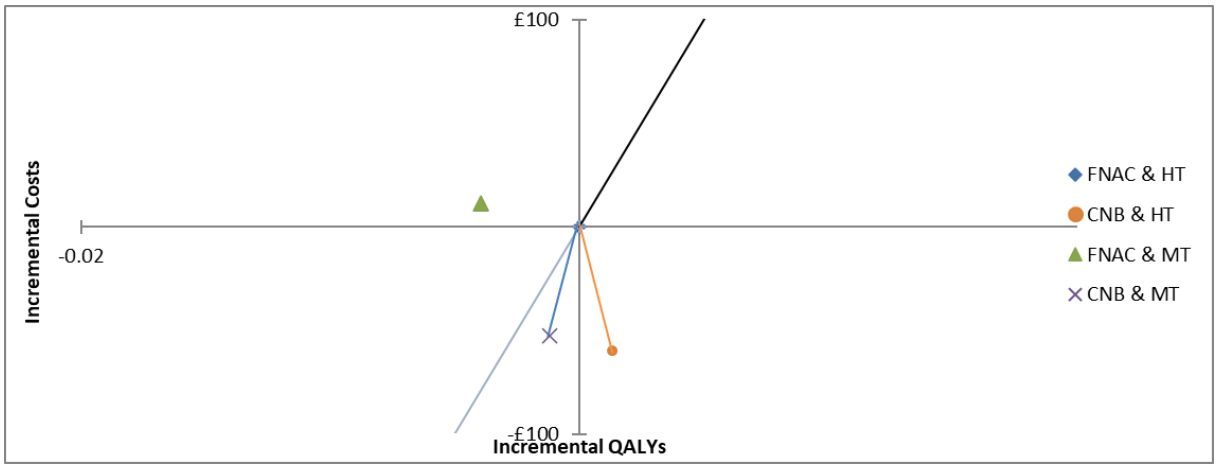
Among all the strategies, CNB and selective use of hemithyroidectomy is the cheapest strategy followed by CNB with selective use of molecular testing, which is slightly more expensive (£7). The two CNB strategies also yield the highest QALYs per patient making them most cost-effective compared to all the other alternatives. CNB and selective use of hemithyroidectomy has a higher NMB and a higher probability of being cost-effective (34%) compared to CNB and selective use of molecular testing (30%) although the difference is extremely small. It is worth noting the difference in uncertainty between CNB and molecular testing and CNB and hemithyroidectomy, which is driven by the fact that, while the accuracy of the molecular tests varies in each simulation, it is always 1 for hemithyroidectomy.

Comparing CNB and FNAC, in 64% of the simulations CNB strategies were more cost-effective than FNAC ones. This is because the lower inconclusiveness rate of CNB effectively prevents a large number of people from receiving hemithyroidectomy or molecular testing. Neither hemithyroidectomy alone nor molecular testing alone were cost-effective in any of the simulations.

The cost-effectiveness plane in Figure 14 is very eloquent. Compared to current practice FNAC and selective use of hemithyroidectomy, only two strategies lie below the £20,000 threshold line: CNB and molecular testing and CNB and hemithyroidectomy. However, while CNB and molecular testing lies in the south-west quadrant, CNB and hemithyroidectomy is in the south-east quadrant, suggesting that the latter is not only a cheaper strategy but slightly more effective as well. Hence, CNB and selective us of hemithyroidectomy dominates CNB and selective use of molecular testing.

© NICE 2022. All rights reserved. Subject to Notice of rights.

Figure 14: Cost-effectiveness plane (Thy3a)



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

Note: All the strategies are compared with current practice FNAC & HT which lies in the origin of the axes

3.2.2 Deterministic scenario analysis

All the scenarios illustrated in Table 27 in section 2.4 were run deterministically to see if the rank of the strategies changes when varying one or more parameters. Table 32 reports all the scenarios where the rank changed compared to the base case analysis.

Table 32: Scenario analysis – Thy3a (deterministic)

Scenario	1st ranked	2nd ranked	3rd ranked
Base case	CNB & HT	CNB & MT	FNAC & HT
Bongiovanni 2012 prevalence	CNB & MT	FNAC & MT	CNB & HT
Disutility with recurrence lasts for 1 year only	CNB & MT	CNB & HT	FNAC & MT
ThyroSeq V1	CNB & HT	FNAC & HT	CNB & MT
Kuo 2018[16] prices of molecular tests	CNB & HT	FNAC & HT	CNB & MT

(a) Abbreviations: CNB = core needle biopsy; CT = completion thyroidectomy; FNAC = fine needle aspiration cytology; HT = hemithyroidectomy; MT = molecular testing

The prevalence of cancer or risk of malignancy (ROM) in people with Thy3a was found to be a critical assumption able to change the overall conclusion of the analysis. When Bongiovanni 2012[5] meta-analysis was used, CNB and selective use of molecular test became the most cost-effective strategy. Bongiovanni 2012 estimated a much lower ROM in Thy3a equal to 16% compared to the 25% of the base case scenario. With such a low malignancy prevalence, the use of hemithyroidectomy is hardly justifiable and the first two ranked strategies become, respectively, CNB with selective use of molecular testing and FNAC with selective use of molecular testing.

The assumption on the duration of disutility was also found to be particularly relevant. If disutility of recurrence last for 1 cycle only, CNB and selective use of molecular test becomes

© NICE 2022. All rights reserved. Subject to Notice of rights.

the most cost-effective strategies as it is associated with more cases of missed diagnosis than CNB and hemithyroidectomy.

As expected, all the assumptions about the performance and cost of the molecular tests were found to influence the rankings in the scenario analysis. In the scenario where no other test except the NHS mimicking ThyroSeq V1 was available, both strategies with selective use of hemithyroidectomy become more cost-effective than selective use of molecular testing. Likewise, if we use a higher estimation for the price of the test[16] CNB with selective use of molecular testing becomes the third most cost-effective strategy.

In none of the scenario analyses FNAC was found to be more cost-effective than a CNB strategy, which strengthens the main analysis findings of CNB being superior to FNAC for people with Thy3a nodules.

3.3 Indeterminate Thy3f

As mentioned earlier, repeat sampling with FNAC or CNB is rarely of any use after a Thy3f cytology, and therefore these strategies were excluded when the analysis was run for a cohort of people with Thy3f. Hence, the strategies included for Thy3f are routine hemithyroidectomy reflecting current practice and the routine use of the four molecular tests included.

3.3.1 Base case analysis

Table 33 illustrates the probabilistic diagnostic outcomes in the base case scenario of each included strategy for a cohort of 1,000 people with Thy3f and an average risk of malignancy of 31%[38].

Table 33: Diagnostic performance – Thy3f (probabilistic)

Strategy	True positive	False positive	True negative	False negative
Afirma-GSC	264 (163 to 352)	246 (162 to 341)	444 (342 to 544)	46 (7 to 137)
ThyroSeq V3	295 (213 to 375)	172 (106 to 250)	518 (426 to 606)	16 (1 to 69)
ThyGenx/ThyraMIR	225 (55 to 347)	62 (10 to 152)	628 (518 to 722)	85 (6 to 257)
ThyroSeq V1	252 (133 to 347)	241 (126 to 377)	449 (310 to 579)	58 (8 to 170)
HT	310 (236 to 388)	690 (612 to 764)	0	0

Note: People with benign nodules undergoing hemithyroidectomy are considered false positives (avoidable surgery) even though the hemithyroidectomy would correctly identify the nodule as benign

Abbreviation: HT = hemithyroidectomy

As the table shows, two molecular tests performed quite well on a population with Thy3f cytology. ThyroSeq V3 had the best sensitivity and therefore reduces the number of missed diagnoses to 15 (out of 310 people with cancer). ThyGenX/ThyraMIR, on the contrary, had the best specificity among all the tests included, leading to only 63 unnecessary surgeries.

Table 34 and Figure 15 show the cost-effectiveness results.

Table 34: Cost-effectiveness results – Thy3f (probabilistic)

Strategy	Cost per patient	QALYs per patient	NMB (£20,000 per QALY)	Probability ranked 1st (£20,000 per QALY)
Afirma-GSC	£5,867 (£5,223 to £6,609)	16.69 (15.87 to 17.12)	£327,882 (£310,894 to £336,889)	0%

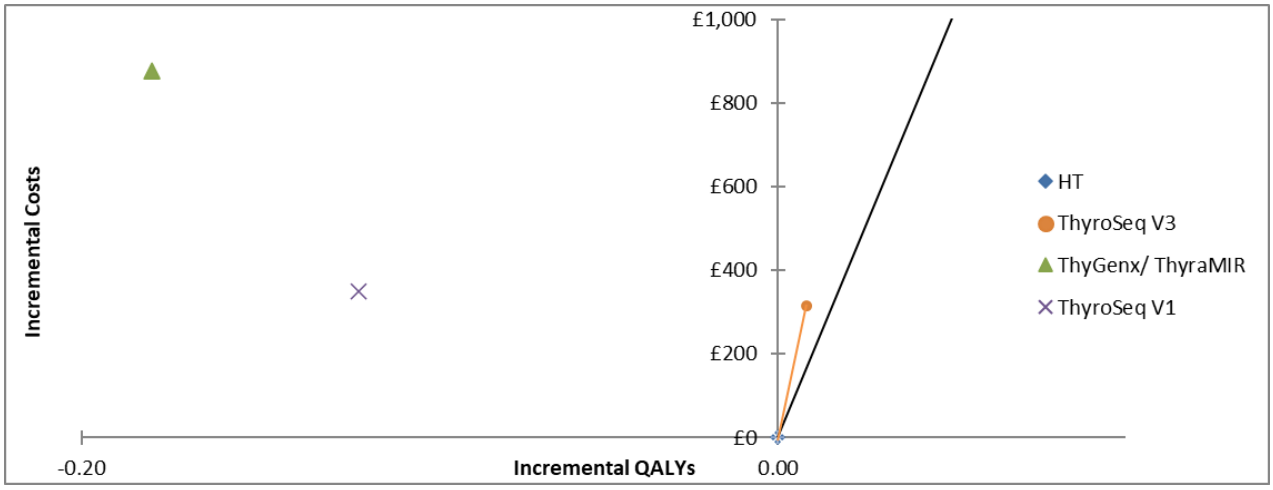
© NICE 2022. All rights reserved. Subject to Notice of rights.

Strategy	Cost per patient	QALYs per patient	NMB (£20,000 per QALY)	Probability ranked 1st (£20,000 per QALY)
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

Unsurprisingly, due to the high prevalence of cancer among people with Thy3f, hemithyroidectomy was the cheapest strategy of all, dominating most of the molecular testing strategies. However, ThyroSeq V3 was slightly more effective than hemithyroidectomy due to the large number of unnecessary surgeries avoided and reduction of surgery-related complications. ThyroSeq V3 was found to be the most cost-effective strategy in 57% of the simulations compared with 40% of hemithyroidectomy. Nevertheless, average NMB remains higher with hemithyroidectomy highlighting, as in the case of Thy3a, the difference in uncertainty between molecular testing and hemithyroidectomy, whose accuracy is always assumed to be 1. As Figure 15 shows, although more effective, the higher incremental cost of ThyroSeq V3 means that the molecular testing strategy lies above the £20,000 threshold line.

Figure 15: Cost-effectiveness plane (Thy3f)



Abbreviations: HT = hemithyroidectomy.

Note: All the strategies are compared with current practice hemithyroidectomy which lies in the origin of the axes

3.3.2 Scenario analyses

All the scenarios illustrated in Table 27 in section 2.4 were run deterministically to look whether the rank of the strategies would change by varying one or more parameters. In Table 35 reports all the scenarios where the rank changed compared to the base case analysis.

© NICE 2022. All rights reserved. Subject to Notice of rights.

Table 35: Scenario analysis – Thy3f (deterministic)

Scenario	1st ranked	2nd ranked	3rd ranked
Base case	ThyroSeq V3	Diagnostic HT	ThyroSeq V1
Kuo 2018[16] prices of molecular tests	Diagnostic HT	ThyroSeq V3	ThyGenX/ThyraMIR
Additional FNAC needed for MT	Diagnostic HT	ThyroSeq V3	ThyroSeq V1

Abbreviation: FNAC = fine needle aspiration cytology; HT = hemithyroidectomy; MT = molecular testing

Unsurprisingly, the scenario analysis found that the results were sensitive to the cost of molecular testing. When the higher prices estimated from Kuo 2018[16] were used, or when an additional FNAC was deemed necessary to extract materials for molecular testing, diagnostic hemithyroidectomy becomes the most cost-effective strategy. It should be emphasized that, in people with Thy3f, the scenario where an additional FNAC is needed for the molecular test is more likely to occur than in people with Thy3a. With the latter, a CNB is assumed to be performed in the most cost-effective scenario, which extracts a larger sample compared to FNAC, which should be sufficient for further molecular analyses. On the other hand, people with Thy3f only received a first FNAC, which might not have extracted enough material.

3.4 Threshold analyses

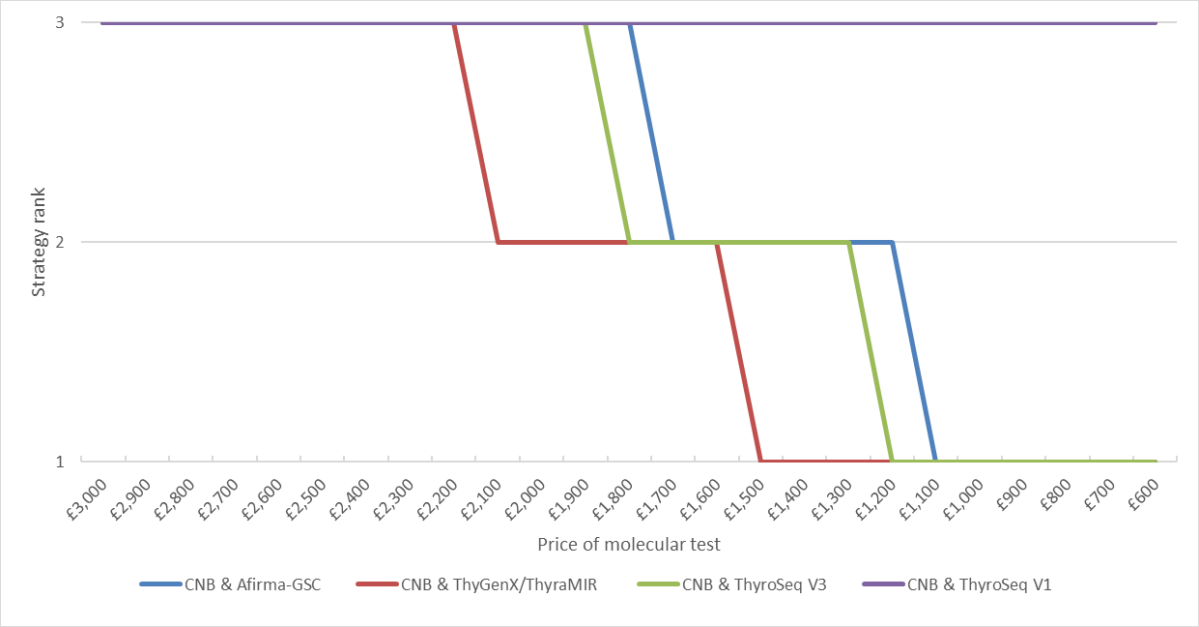
3.4.1 Threshold price of molecular testing

A threshold analysis on the price of each molecular test was performed for both subgroups where molecular test was an included strategy: Thy1 and Thy3a. The price of each test was varied (keeping constant the prices of the other tests) from a maximum value of £3,000 to a minimum of £600. The threshold price was that at which that specific molecular test switched to being cost-effective at £20,000 per QALY gained.

The results for Thy3a are illustrated in **Error! Reference source not found.** below. None of the strategies with the routine use of molecular testing were found to be cost-effective at any price, as CNB with selective use of molecular test was always found to be superior. Therefore, **Error! Reference source not found.** shows the threshold price making CNB and selective use of each test cost-effective.

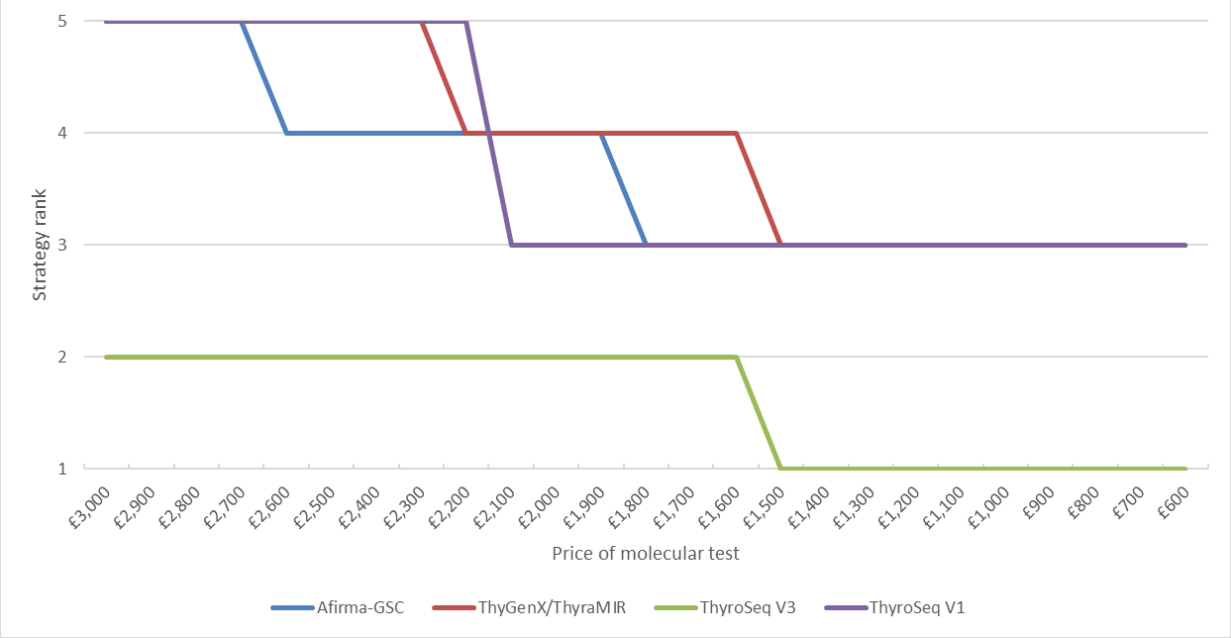
© NICE 2022. All rights reserved. Subject to Notice of rights.

Figure 16: Molecular tests price threshold analysis – CNB and selective use of molecular test (Thy3a)



CNB biopsy and selective use of most of the molecular tests becomes cost-effective compared with the alternative of selective use of hemithyroidectomy when the cost of the tests ranges between £1,500 and £1,100. Due to its poor performance, selective use of ThyroSeqV1 is not preferable to selective use of hemithyroidectomy at any price. None of the molecular tests, except ThyroSeq V3, are known to have a price included in this range although the estimated price of ThyroSeq V3 of £1,407 lies above the threshold price of this test (£200), which explains the close similarity between the NMB of CNB and selective use of ThyroSeq V3 and the NMB of CNB and selective use of hemithyroidectomy.

Figure 17: Molecular tests price threshold analysis – routine use of molecular test (Thy3f)



Due to the poor sensitivity of most of the tests, only ThyroSeq V3 test becomes cost-effective when the price lies below £1,500. This threshold is just above the current estimated price of

© NICE 2022. All rights reserved. Subject to Notice of rights.

£1,402 although it is worth recognising that the actual cost of molecular test may be higher for Thy3f, as an additional FNAC may be required if the original FNAC failed to extract enough material for further analyses.

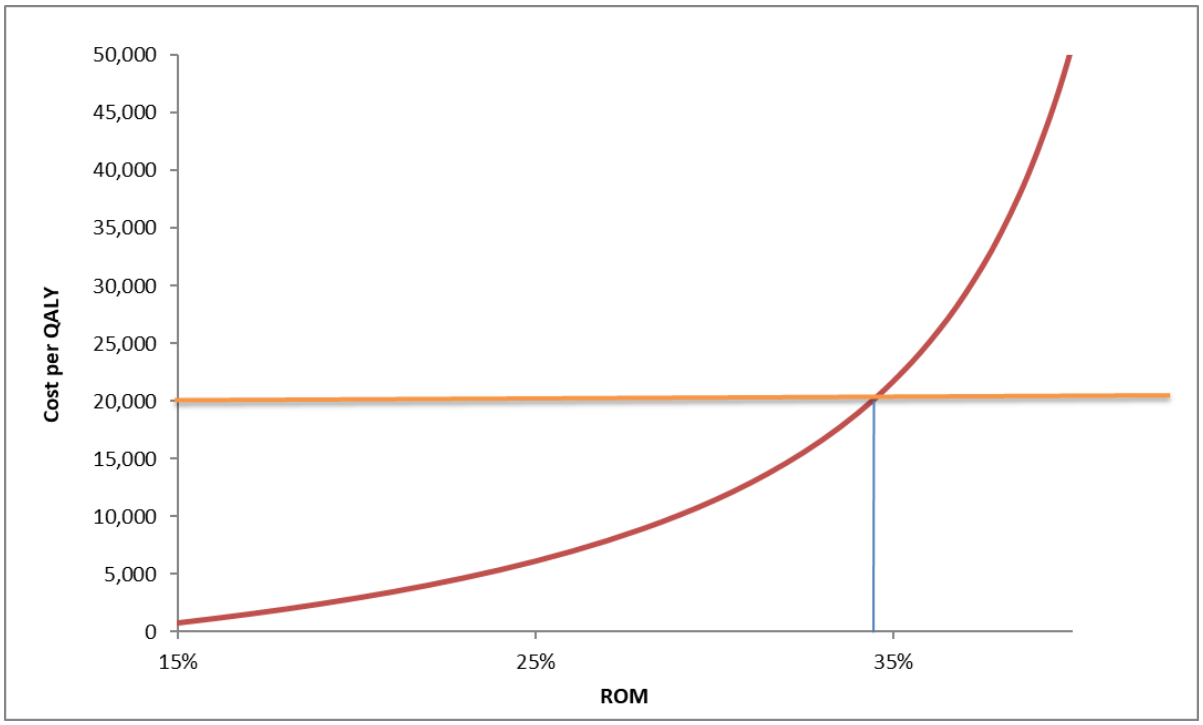
3.4.2 Threshold risk of malignancy

A threshold analysis on the risk of malignancy was conducted for both Thy3a and Thy3f where molecular testing was compared against the most cost-effective alternative: for Thy3a, CNB and molecular testing was compared against CNB and hemithyroidectomy; for Thy3f, routine use of molecular testing with ThyroSeq V3 was compared with hemithyroidectomy.

The threshold analysis for Thy3a indicates that CNB and molecular testing dominates CNB and hemithyroidectomy for any risk of maligancy below 23%. Above this threshold value, CNB and hemithyroidectomy dominates CNB and molecular testing. This highlights the importance of understanding the real risk of malignancy in each cytology. If we assume that Thy3a has a lower ROM than the one used in the base case scenario (25%), even a slight reduction would make selective use of molecular testing cost-effective.

The threshold analysis on Thy3f could be represented visually (see Figure 18). Molecular testing is cost-effective if the true prevalence of cancer in Thy3 cytologies lies below 35%. If the risk of malignancy is higher than 35%, then diagnostic hemithyroidectomy for all becomes the most cost-effective strategy. With a prevalence of 31%, as assumed in the base case scenario, molecular testing seems to be cost-effective, although the probabilistic analysis showed that this result is not particularly robust.

Figure 18: Threshold analysis on ROM (Thy3f) – Molecular testing (ThyroSeq V3) vs hemithyroidectomy



Abbreviation: ROM = risk of malignancy.

© NICE 2022. All rights reserved. Subject to Notice of rights.

4 Discussion

4.1 Summary of results

This original cost-utility analysis found that for people with suspicious nodules who received an inadequate result after a FNAC:

- If they received a non-diagnostic cytology Thy1, repeat sampling with core needle biopsy dominates both repeat FNAC and hemithyroidectomy
- If they receive an indeterminate cytology Thy3a, repeat sampling with core needle biopsy dominates both repeat FNAC, routine use of molecular testing and hemithyroidectomy. If a further inconclusive result is obtained, selective use of hemithyroidectomy dominates selective use of molecular testing with ThyroSeq V3.
- If they receive an indeterminate cytology 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. It is noteworthy that molecular testing strategies in both Thy3a and Thy3f populations had a high degree of uncertainty compared to the alternative of hemithyroidectomy. Even if selective use of molecular testing and routine use of molecular testing were more likely to be cost-effective in, respectively, a Thy3a and a Thy3f population, their average NMB was below the average NMB of hemithyroidectomy.

4.2 Limitations and interpretation

This analysis demonstrated that for people who received a non-diagnostic cytology Thy1 or an indeterminate cytology Thy3a, CNB is cost-effective against all the alternative diagnostic strategies.

The cost-effectiveness of molecular testing in England remains less certain. In both Thy3a and Thy3f cytologies, there was a significant probability of a molecular testing strategy to be cost-effective (30% in Thy3a and 57% in Thy3f) although, in both cases, the alternative hemithyroidectomy strategy had a higher average NMB. This highlights the higher degree of uncertainty with molecular testing compared to hemithyroidectomy. Only one test was found to be potentially cost-effective, ThyroSeq V3, which is not currently available in the NHS and would need to be purchased abroad. Although shipping and packaging costs were included in the analysis, final costs may change if, for instance, shipping time or laboratory waiting time would require additional expenditures: e.g. freezing or storing of the sample. Moreover, shipping a sample abroad instead of performing diagnostic surgery immediately may delay the final diagnosis. This is of concern as an important proportion of people with Thy3a and Thy3f, 25% and 31% respectively, are expected to have cancer and further delays in the diagnosis may lead to several centres failing to meet the NHS cancer waiting time targets of 31 or 62 days. This may lead to severe adverse outcomes for people with cancer as time to surgery is an important predictor of lifelong risk of recurrence. Finally, if some of the assumptions used in the base case scenario do not hold, for instance if an additional FNAC is required for molecular testing, this latter would become even less cost-effective..

The analysis had some limitations. Firstly, due to lack of data, the role of multi-disciplinary team (MDT) could not be included in the analysis. The importance of the MDT in diagnosing a variety of cancers is well known in clinical practice[8], however it was not possible to determine the accuracy of including an MDT discussion after a FNAC and CNB. It is likely that, before offering a hemithyroidectomy, an MDT meeting will be held to discuss clinical features, ultrasound findings, FNAC and, when applicable, CNB cytopathologic findings. This likely reduces the number of diagnostic hemithyroidectomies performed on people with a low

© NICE 2022. All rights reserved. Subject to Notice of rights.

likelihood of having cancer, making strategies with selective use of hemithyroidectomy more cost-effective.

Secondly, there were some concerns about data used for accuracy, in particular for molecular testing. The new generation of GSC tests have been introduced relatively recently and a systematic review or meta-analysis were not available. The quality of each single study was individually assessed using QUADAS 2, which found significant concerns for two tests in particular: ThyroSeq V1 and ThyGenX/ThyraMIR (see Appendix B:). Neither of these two tests were found to be cost-effective in the analysis. Inconclusive rates as well as CNB accuracy after a Thy1 or Thy3a were collected from a meta-analysis[40] that was found to lack a proper protocol and having an unclear reporting of search strategy. However, the results of the study were in line with most of the published data and with the accuracy data collected for the review of "Evidence review D FNAC or biopsy" and therefore deemed reliable. Accuracy of repeat FNAC was not available from the same source.

Accuracy of repeat FNAC was estimated using the studies included for "Evidence review D FNAC or biopsy". As these studies were predominantly assessing first-time FNAC, inconclusive rates were estimated through the more appropriate meta-analysis on repeat FNAC[40]; thus only the accuracy of conclusive categories Thy2, Thy3f, Thy4 and Thy5 was calculated from the studies included in the clinical review. This approach was considered reasonable as the main limitation of a repeat FNAC is the higher rate of inconclusive results, but it might possibly overestimate accuracy of FNAC nevertheless. If this is the case, the comparative advantage of CNB found in this analysis would be even greater.

Some simplifications were made when developing the model. Firstly, real data on natural progression were unavailable as thyroid cancers are generally treated immediately once detected and active surveillance of malignant nodules is still rarely practiced. However, the model included cancers which were erroneously diagnosed as benign nodules and received a very late treatment. To model the natural progression of an untreated cancer, a large observational study of active surveillance of people that had cancers with low-risk features was used. Some adjustments were made to exclude very low-risk cancers from the study although it is still expected that the model underestimates the real-world probability of progression to clinical disease. This may lead to an overestimation of the cost-effectiveness of all the strategies with a larger number of missed diagnoses. However, it might not affect the results significantly, as missed cancers are assumed to be detected in two years as show by one of the evidence[51].

Finally, mortality of people with recurrent disease was calculated by applying a hazard ratio estimated from a Kaplan-Meier curve of a published observational study[22]. A first limitation was caused by the methodology used to estimate the hazard ratio[50], which is generally applied to Kaplan-Meier curves of clinical trials as it only allows to calculate unadjusted HRs that are not controlled for potential confounding factors. Were this study aimed at estimating the real causal effect of recurrence on mortality, the use of an unadjusted hazard ratio would be hard to justify. However, the goal of this analysis is to estimate real-world mortality of people who had recurrent disease, assuming already that they would be inevitably different from people who had no recurrence. One of the most important differences between the two groups, time to surgery, was in fact already incorporated in the model as people with recurrence are more likely to have had a missed diagnosis before. The only difference between people with and without recurrence that is not controlled for would be age as the model simulated a cohort of people with same age. However, the use of standardised survival time used by Links 2005[22] should avoid biases in mortality arising from differences in age. A second limitation is caused by the fact that mortality in people with recurrence is not a linear transformation of general population mortality and the use of a constant hazard ratio may underestimate mortality due to recurrence in the short-term, even though long-term mortality should be adequately captured. This is not expected, however, to affect the results in a significant way.

© NICE 2022. All rights reserved. Subject to Notice of rights.

4.3 Generalisability to other populations or settings

This analysis is based on people who received a non-diagnostic Thy1 or an indeterminate Thy3a and Thy3f in England.

Unit costs were collected from UK national sources and, as such, the analysis might not be generalisable to other settings. It is worth noting that management of thyroid cancer in the UK is relatively low cost compared to other countries, where molecular testing may represent a more cost-effective alternative.

4.4 Comparisons with published studies

This is the first economic study, to our knowledge, comparing CNB with repeat FNAC on people who had a Thy1 and Thy3a results.

Some economic evaluations comparing molecular testing with hemithyroidectomy are available in the literature. Lee 2014[20] compared molecular test with diagnostic hemithyroidectomy in Canada finding routine GEC (gene expression classifier) followed by selective use of GMP (gene mutation panel) potentially cost-effective. However, similarly to this analysis, a high degree of uncertainty was observed as, in the probabilistic analysis, diagnostic hemithyroidectomy had the highest probability of being cost-effective. Moreover, the molecular tests included in the analysis represent the old generation of tests as the newly developed GSC tests were not yet available.

Ronen and colleagues recently published a UK cost-comparison analysis on the new generation of molecular test GSC[41]. The analysis found hemithyroidectomy to be cost saving compared to GSC when a risk of malignancy of 30% (Thy3f) was assumed. The authors conducted a threshold analysis finding a threshold price for GSC in England equal to £2,177. Although this is considerably higher than the threshold of £1,500 estimated in the analysis for this guideline (GSC test: ThyroSeq V3 – see section 3.4), the estimated price of this test of £1,400 lies below both thresholds.

4.5 Conclusions

This economic evaluation demonstrated that repeat sampling with core-needle biopsy is cost-effective in people with non-diagnostic cytology Thy1 and indeterminate cytology Thy3a compared with repeat FNAC, hemithyroidectomy or routine use of molecular testing (for Thy3a only). The results were found to be robust both in the probabilistic analysis and in the scenario analyses. If a further inconclusive result is obtained with CNB, it is cost-effective to offer hemithyroidectomy to people with Thy3a instead of molecular testing.

For people who received a Thy3f cytology, it is uncertain whether routine use of ThyroSeq V3 is cost-effective compared to hemithyroidectomy as, even though molecular testing was more cost-effective in a higher number Monte Carlo simulation, its average NMB was lower than the average NMB of hemithyroidectomy.

© NICE 2022. All rights reserved. Subject to Notice of rights.

References

1. Almquist M, Ivarsson K, Nordenstrom E, Bergenfelz A. Mortality in patients with permanent hypoparathyroidism after total thyroidectomy. British Journal of Surgery. 2018; 105(10):1313-1318

2. Ambrose J, Davies R. Pharma firm Advanz fined after thyroid drug price hike of 6,000%. 2021. Available from: https://www.theguardian.com/business/2021/jul/29/pharma-firm-advanz-fined-after-thyroid-tablets-price-inflation-of-6000 Last accessed: 24/02/2022.

3. Banovac K, Carrington SA, Levis S, Fill MD, Bilsker MS. Determination of replacement and suppressive doses of thyroxine. Journal of International Medical Research. 1990; 18(3):210-218

4. Bates MF, Lamas MR, Randle RW, Long KL, Pitt SC, Schneider DF et al. Back so soon? Is early recurrence of papillary thyroid cancer really just persistent disease? Surgery. 2018; 163(1):118-123

5. Bongiovanni M, Spitale A, Faquin WC, Mazzucchelli L, Baloch ZW. The Bethesda System for Reporting Thyroid Cytopathology: a meta-analysis. Acta Cytologica. 2012; 56(4):333-339

6. Cibas ES, Ali SZ, N. C. I. Thyroid FNA State of the Science Conference. The bethesda system for reporting thyroid cytopathology. American Journal of Clinical Pathology. 2009; 132(5):658-665

7. Cross P, Chandra A, Giles T, Johnson S, Kocjan G, Poller D et al. Guidance on the reporting of thyroid cytology specimens. London. The Royal College of Pathologists, 2016. Available from: https://www.rcpath.org/uploads/assets/7d693ce4-0091-4621-97f79e2a0d1034d6/g089_guidance_on_reporting_of_thyroid_cytology_specimens.pdf

8. Davies AR, Deans DAC, Penman I, Plevris JN, Fletcher J, Wall L et al. The multidisciplinary team meeting improves staging accuracy and treatment selection for gastro-esophageal cancer. Diseases of the Esophagus. 2006; 19(6):496-503

9. Decision Support Unit, Alava MH, Pudeny S, Wailoo A. Estimating eq-5d by age and sex for the UK. Sheffield. 2022. Available from: https://www.sheffield.ac.uk/sites/default/files/2022-02/DSU%20Age%20based%20utility%20-%20Final%20for%20website.pdf

10. Dehbi HM, Mallick U, Wadsley J, Newbold K, Harmer C, Hackshaw A. Recurrence after low-dose radioiodine ablation and recombinant human thyroid-stimulating hormone for differentiated thyroid cancer (HiLo): long-term results of an open-label, non-inferiority randomised controlled trial. The Lancet Diabetes & Endocrinology. 2019; 7(1):44-51

11. 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:

12. Ito Y, Miyauchi A, Kihara M, Higashiyama T, Kobayashi K, Miya A. Patient age is significantly related to the progression of papillary microcarcinoma of the thyroid under observation. Thyroid. 2014; 24(1):27-34

© NICE 2022. All rights reserved. Subject to Notice of rights.

13. 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

14. 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

15. Kim SY, Kim SM, Chang H, Kim BW, Lee YS, Kwon SS et al. Cost for treatment and follow-up of thyroid cancer increases according to the severity of disease. Head and Neck. 2019; 41(7):2376-2379

16. Kuo EJ, Wu JX, Zanocco KA. Cost effectiveness of immediate biopsy versus surveillance of intermediate-suspicion thyroid nodules. Surgery. 2018; 164(6):1330-1335

17. 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

18. Labourier E. Utility and cost-effectiveness of molecular testing in thyroid nodules with indeterminate cytology. Clinical Endocrinology. 2016; 85(4):624-631

19. 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

20. 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

21. Lim WS, Lawrence H. National Adult Community Acquired Pneumonia Audit 2018/19. British Thoracic Society Reports,. 2019; 10(4)

22. Links TP, van Tol KM, Jager PL, Plukker JTM, Piers DA, Boezen HM et al. Life expectancy in differentiated thyroid cancer: a novel approach to survival analysis. Endocrine-Related Cancer Endocr Relat Cancer. 2005; 12(2):273-280

23. 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

24. Na DG, Kim JH, Sung JY, Baek JH, Jung KC, Lee H et al. Core-needle biopsy is more useful than repeat fine-needle aspiration in thyroid nodules read as nondiagnostic or atypia of undetermined significance by the Bethesda system for reporting thyroid cytopathology. Thyroid. 2012; 22(5):468-475

25. National Institute for Health and Care Excellence. Developing NICE guidelines: the manual [updated October 2020]. London. National Institute for Health and Care Excellence, 2014. Available from: http://www.nice.org.uk/article/PMG20/chapter/1%20Introduction%20and%20overview

26. National Institute for Health and Care Excellence. The NICE Charter. 2020. Available from: https://www.nice.org.uk/about/who-we-are/our-charter Last accessed: 24/03/2022.

27. National Institute for Health and Clinical Excellence. Social value judgements: principles for the development of NICE guidance. London. National Institute for

© NICE 2022. All rights reserved. Subject to Notice of rights.

Health and Clinical Excellence, 2008. Available from: https://www.nice.org.uk/media/default/about/what-we-do/research-and-development/social-value-judgements-principles-for-the-development-of-nice-guidance.pdf

28. 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

29. NHS Digital. Linked HES-ONS mortality data. Available from: https://digital.nhs.uk/data-and-information/data-tools-and-services/data-services/linked-hes-ons-mortality-data#:~:text=linked%20mortality%20data-,Using%20HES%20data%20alone%20to%20analyse%20deaths,the%20underlying%20cause%20of%20death Last accessed: 07/03/2022.

30. NHS England. NHS England Annual Report 2017/18. London. 2018. Available from: https://www.england.nhs.uk/wp-content/uploads/2018/07/Annual-Report-Full-201718.pdf

31. 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

32. 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

33. Office for National Statistics. 'Average' Briton highlighted on UN World Statistics Day. 2010.

34. Office for National Statistics. National life tables, UK: 2016 to 2018. 2019. Available from: https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/lifeexpectancies/bulletins/nationallifetablesunitedkingdom/2016to2018 Last accessed: 07/03/2022.

35. Oshikiri T, Takiguchi G, Hasegawa H, Yamamoto M, Kanaji S, Yamashita K et al. Postoperative recurrent laryngeal nerve palsy is associated with pneumonia in minimally invasive esophagectomy for esophageal cancer. Surgical Endoscopy. 2021; 35(2):837-844

36. Patel KN, Angell TE, Babiarz J, Barth NM, Blevins T, Duh Q-Y et al. Performance of a Genomic Sequencing Classifier for the Preoperative Diagnosis of Cytologically Indeterminate Thyroid Nodules. JAMA Surgery. 2018; 153(9):817-824

37. Perros P, Boelaert K, Colley S, Evans C, Evans RM, Gerrard Ba G et al. Guidelines for the management of thyroid cancer. Clinical Endocrinology. 2014; 81(Suppl 1):1-122

38. Poller DN, Bongiovanni M, Trimboli P. Risk of malignancy in the various categories of the UK Royal College of Pathologists Thy terminology for thyroid FNA cytology: A systematic review and meta-analysis. Cancer Cytopathology. 2020; 128(1):36-42

39. Poller DN, Doyle V, Trimboli P, Bongiovanni M. Rates of Thy 1-non-diagnostic thyroid fine needle aspiration using the UK Royal College of Pathologists Thy Terminology. A systematic review of the literature comparing patients who undergo rapid on-site evaluation and those who do not. Cytopathology. 2020; 31(6):502-508

© NICE 2022. All rights reserved. Subject to Notice of rights.

40. 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

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

42. Royal Mail. Safebox 1st Class box of 48. Available from: https://shop.royalmail.com/safebox-1st-class-box-of-48 Last accessed:

43. 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

44. 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

45. 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

46. 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

47. University Hospitals Sussex NHS Foundation Trust. Thyroid lumps. 2021. Available from: https://www.bsuh.nhs.uk/wp-content/uploads/sites/5/2016/09/thyroid-lumps.pdf

48. UPS. UPS Shipping cost and time calculator. 2022. Available from: https://wwwapps.ups.com/ctc/request?loc=en_GB Last accessed: 24/03/2022.

49. 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

50. Williamson PR, Smith CT, Hutton JL, Marson AG. Aggregate data meta-analysis with time-to-event outcomes. Statistics in Medicine. 2002; 21(22):3337-3351

51. Yeh MW, Demircan O, Ituarte P, Clark OH. False-negative fine-needle aspiration cytology results delay treatment and adversely affect outcome in patients with thyroid carcinoma. Thyroid. 2004; 14(3):207-215

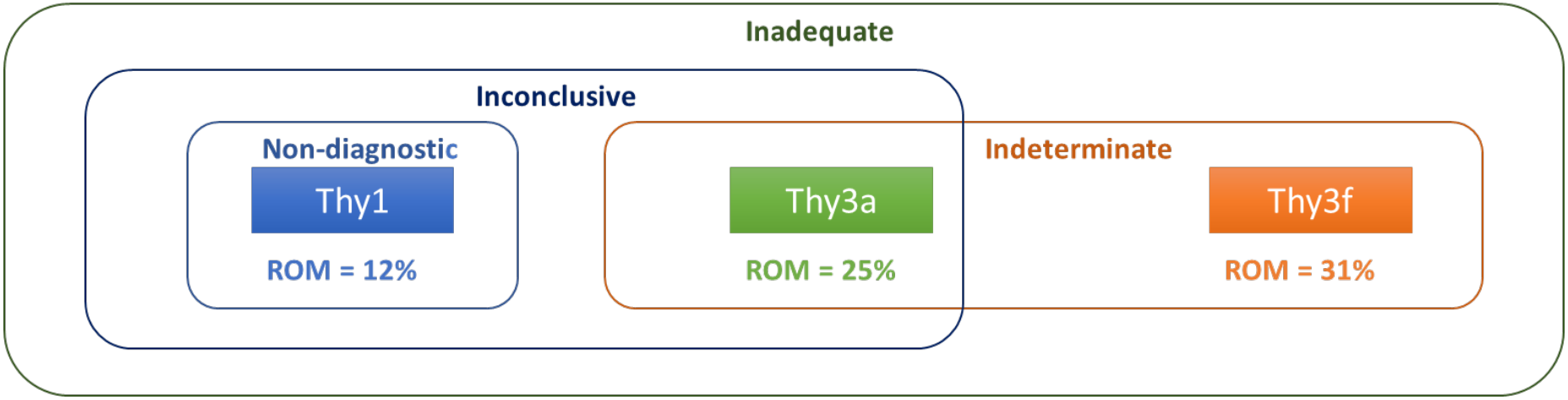
© NICE 2022. All rights reserved. Subject to Notice of rights.

Appendices

© NICE 2022. All rights reserved. Subject to Notice of rights.

Appendix A: Cytology definition

Figure 19: Visual representation of cytology terminologies



Abbreviations: ROM = risk of malignancy

© NICE 2022. All rights reserved. Subject to Notice of rights.

Appendix B: Quality appraisal

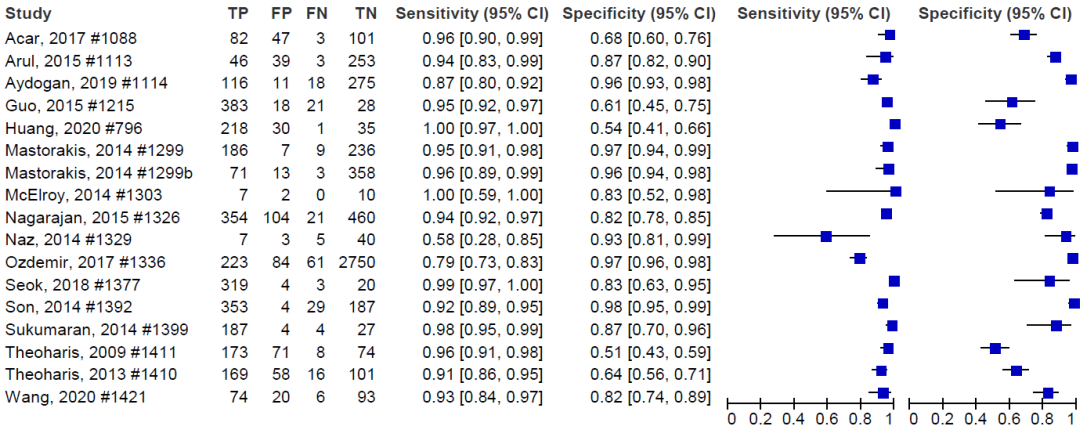
Table 36: QUADAS 2 table

	PATIENT SELECTION				INDEX TESTS			REFERENCE TEST			FLOW AND TIMING					OVERALL
Study	Was a consecutive or random sample of patients enrolled? (Y/N/U)	Was a case-control design avoided? (Y/N/U)	Did the study avoid inappropriate exclusions? (Y/N/U)	**Could the selection of patients have introduced bias? LOW/HIGH/ UNCLEAR RISK**	Were the index test results interpreted without knowledge of the results of the reference standard? (Y/N/U)	If a threshold was used, was it pre-specified? (Y,N,U)	**Could the conduct or interpretation of the index test have introduced bias? LOW/ HIGH/ UNCLEAR RISK**	Is the reference standard likely to correctly classify the target condition? (Y/N/U)	Were the reference standard results interpreted without knowledge of the results of the index test? (Y/N/U)	**Could the reference standard, its conduct, or its interpretation have introduced bias? LOW/ HIGH/ UNCLEAR RISK**	Was there an appropriate interval between index test(s) and reference standard? (Y/N/U)	Did all patients receive a reference standard? (Y/N/U)	Did patients receive the same reference standard? (Y/N/U)	Were all patients included in the analysis? (Y/N/U)	**Could the patient flow have introduced bias? LOW/ HIGH/ UNCLEAR RISK**	**Overall risk of bias: Very serious, serious or no serious risk of bias (This is an extension to make QUADAS compatible with GRADE)**
Patel, 2018	Y	Y	U	LOW RISK	Y	Y, based on prior study with n=634 samples	LOW RISK	Y	Y	LOW RISK	U	Y	Y	Y	UNCLEAR RISK	Serious risk of bias
Nikiforov, 2018: (Steward et al. 2019)	Y	Y	Y	LOW RISK	Y	Y	LOW RISK	Y	Y	LOW RISK	U	Y	Y	N, but appropriate exclusions	UNCLEAR RISK	Serious risk of bias
Shrestha, 2016	U	Y	U	UNCLEAR RISK	U	U	HIGH RISK	Y	U	HIGH RISK	U	Y	Y	U	UNCLEAR	Very serious risk of bias
Labourier, 2015	Y	Y	Y	LOW RISK	U	Y	HIGH RISK	Y	U	HIGH RISK	U	Y	Y	N	UNCLEAR RISK	Very serious risk of bias

© NICE 2022. All rights reserved. Subject to Notice of rights.

Appendix C: FNAC and CNB accuracy

Figure 20: Forest plot of sensitivity and specificity for FNAC after removing inconclusive categories



© NICE 2022. All rights reserved. Subject to Notice of rights.