National Institute for Health and Care Excellence

Draft for consultation

Thyroid disease: Diagnostic model

Cost consequence analysis: Ultrasound guidance for Fine Needle Aspiration

Economic analysis

Methods, results, conclusions

May 2019

Draft for consultation

Commissioned by the National Institute for Health and Care Excellence

Funding
National Institute for Health and Care Excellence

Appendices

Appendix A: Cost consequence analysis

A.1 Introduction

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

The priority for original economic analysis identified by the committee was to determine the most cost-effective diagnostic strategy when testing with Fine-needle aspiration cytology (FNAC) to detect thyroid malignancy and treat patients.

A.2 Methods

A.2.1 Model overview

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

A.2.1.1 Population

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

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

A.2.1.2 Comparators

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

The following diagnostic strategies were chosen as comparators:

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

A.2.1.3 Time horizon and discounting

The committee felt that a cost-utility analysis was difficult to do without major assumptions, and cancer (except preliminary investigation) is outside the guideline scope. Therefore, short-term costs and diagnostic accuracy was estimated, which the committee felt was sufficient to inform their recommendations.

Discounting was therefore not applicable, as the model did not capture future costs and benefits.

A.2.2 Approach to modelling

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

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

A.2.2.1 Model structure

When patients enter the model, they would have had a positive US test. The positive predictive value (PPV) of the US scan determines the proportion of the model population who actually have cancer.

The committee noted that the British thyroid association (BTA) numerical diagnostic categories are as follows:

- Thy1 (non-diagnostic)
- Thy2 (non-neoplastic)
- Thy3A (neoplasm possible with atypical features)
- Thy3F (follicular neoplasm)
- Thy4 (suspicious)
- Thy5 (diagnostic of malignancy)

For simplicity, some outcomes were grouped together as they follow the same diagnostic pathway. The Thy3F and Thy5 were grouped as malignant as they go directly to surgery rather than a repeat FNAC due to the high level of US suspicion. This is consistent with how studies have been treated in the guideline's clinical review. If Thy3F and Thy5 were reported separately in a study then they would have been added together in the review. Similarly, Thy4 and Thy3A were grouped as indeterminate and would have a repeat FNAC.

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

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

Figure 1 and Figure 2 illustrate the diagnostic pathways compared in this cost consequence analysis. For UGFNAC the pathway is the same for PGFNAC except that all FNACs are ultrasound guided.

The committee felt it would be accurate to assume that patients with indeterminate or inadequate results after a first FNAC would receive a repeat FNAC. Therefore, the indeterminate and the inadequate would follow the same diagnostic pathway.

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

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

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

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

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

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

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

Figure 1: UGFNAC without benign repeat

Figure 2: UGFNAC with benign repeat

A.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 outcomes (TP and FP) were calculated using these values. The model was run repeatedly – 10,000 times (Monte Carlo simulation).

When running 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 probabilistic analysis, we checked for convergence in the incremental total cost and incremental true positives for 'UGFNAC without benign repeat' versus 'PGFNAC without benign repeat' and also for 'UGFNAC with benign repeat' versus 'UGFNAC without benign repeat' by plotting the number of runs against the mean outcome at that point on a graph. The results had converged by the 4,000th iteration.

The way in which distributions are defined reflects the nature of the data so for example, probabilities were given a beta distribution, which is bounded by 0 and 1, reflecting that a probability cannot be outside this range. All of the variables that were probabilistic in the model and their distributional parameters are detailed in Table 1 below and in Table 6. Probability distributions in the analysis were parameterised using error estimates from data sources.

Table 1: Description of the type and properties of distributions used in the probabilistic sensitivity analysis

Parameter	Type of distribution	Properties of distribution
Prevalence before ultrasound	Beta	Bounded between 0 and 1. Derived from mean of a domain or total quality of life score and its standard error, using the method of moments. Alpha and Beta values were calculated as follows: Alpha = $(mean/SE)^2$ Beta = $SE^2/Mean$
Specificity of ultrasound	Beta	Bounded between 0 and 1. As the sample size and the number of events from the source study are known, alpha and Beta values were calculated as follows: Alpha = (number of cancer-free patients testing negative) Beta = (number of cancer-free patients testing positive)
Diagnostic odds ratio of ultrasound	Log Normal	Bounded at 0. Derived from mean and standard deviation.
Sensitivity and specificity of fine needle aspiration cytology	Bespoke	The 60,000 paired estimates that form the joint posterior distribution for sensitivity and specificity were extracted from the WinBUGS output of the diagnostic meta-analysis. In the PSA we sampled at random a pair of sensitivity and specificity, thus preserving the inverse correlation.
Test outcomes (Malignant, benign, indeterminate and inadequate)	Dirichlet	Represents a series of conditional distributions, bounded on 0–1 interval. Parameters are the number of patients in each category.
Test costs and FN cost	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 = SE2/Mean Where SE was estimated by assuming the 95% CI to be plus

Parameter	Type of distribution	Properties of distribution
		or minus 25% of the mean and SE=(UCI-LCI)/2Z0.975
Surgery costs	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 = $SE^2/Mean$ Beta = $SE^2/Mean$ Where SE was estimated from the upper quartile (UQ) and lower quartile (LQ) as follows: $SE=(UQ-LQ)/2Z_{0.75}$

Abbreviations: SE = standard error, UQ = upper quartile, LQ = lower quartile.

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 the results and whether conclusions on which diagnostic strategy should be recommended would change.

A.2.3 Model inputs

Model inputs were based on clinical evidence identified in the systematic review undertaken for the guideline, supplemented by additional data sources. Model inputs were validated by the clinical members of the guideline committee throughout model development. Please see summary Table 6 below for final inputs included in the model.

A.2.3.1 Prevalence

The prevalence of cancer in the population with a large but normally functioning thyroid gland was estimated to be 5% prior to US[1] *(Borget et al 2018)*.

That prevalence, along with ultrasound sensitivity and specificity data, was used to calculate the positive predictive value (PPV) of ultrasound, which was in turn the prevalence entering our model.

The prevalence post-ultrasound is the positive predictive value of ultrasound, which is defined as:

$$\mathrm{PPV} = \frac{\text{sensitivity} \times \text{prevalence}}{\text{sensitivity} \times \text{prevalence} + (1 - \text{specificity}) \times (1 - \text{prevalence})}$$

A.2.3.2 Diagnostic accuracy

A.2.3.2.1 Sensitivity and specificity

The diagnostic accuracy data were identified from the systematic review undertaken for this guideline and presented to the committee for discussion [2, 4, 5, 8, 9]

The diagnostic accuracy data was available from two review questions in this guideline.

Firstly 'imaging and who to FNAC' review question provided the US sensitivity and specificity data. Thirty-four studies that evaluated ultrasound under different criteria were included in the review. Of these, two studies were conducted using the British Thyroid Association (BTA) criteria. In the UK, the most commonly used ultrasound criteria are those of the BTA. For this reason the committee considered the study by *Persichetti et al. 2018*[7] was the most appropriate, for the base-case analysis

as it was the most recent study (2018) and had a larger cohort than the other study. It was considered to represent the best available evidence. Please see Table 2 below for the US diagnostic data used in the model from this study.

Secondly, the UG/ PG FNAC sensitivity and specificity data was taken from the 'FNAC with or without ultrasound guidance' review. This included five studies, [2, 4, 5, 8, 9]. The pooled estimate for both the sensitivity and specificity for UG and PG FNAC were taken from the clinical review and used in the model. Please see Table 2 below for the pooled FNAC diagnostic accuracy data used in the model.

Table 2: Diagnostic accuracy (%)

Index test	Sensitivity (95% CI)	Specificity (95% CI)
US scan	90 (85-95)	63 (60-67)
UGFNAC	90 (76 to 98)	86 (72 to 96)
PGFNAC	71 (48 to 87)	82 (59 to 96)

A.2.3.2.2 Calculating the proportion of patients that received malignant, benign, indeterminate and inadequate test results

Cesur et al 2006[2] was used to determine the proportion of patients that received a malignant, benign, indeterminate, or inadequate test result- see Table 3 below.

In the model, the sensitivity from the guideline's diagnostic meta-analysis times by the prevalence was used to identify the true positives after FNAC.

The missed cases (prevalence x 1-sensitivity) were then assigned benign, inadequate, or indeterminate status in proportion to the results seen in *Cesur et al 2006 cohort* (Table 3).

Similarly, in the no cancer cohort, the benign group was identified by the specificity.

Those cancer-free patients that were not immediately cleared were then assigned either malignant, inadequate, or indeterminate status in proportion to the results seen in *Cesur et al 2006* cohort. See Table 4 below, with worked calculations for each probability.

Table 3: Cytological findings of UG and PG FNAC from Cesur et al. 2006

Outcome	UGFNAC (n)	PGFNAC (n)
Malignant	6	5
Benign	216	186
Indeterminate	2	2
Inadequate	61	92

Table 4: The probability of each outcome in the UGFNAC and PGFNAC

Outcomes	Probability	Calculation
UGFNAC probabilities		
Cancer - Malignant (TP)	0.900	Sensitivity of UGFNAC
Cancer - Benign (FN)	0.077	(1-0.9)* 216/ (279) =(1-sensitivity) x benign over (benign, indeterminate and inadequate)
Cancer - Indeterminate	0.001	(1-0.9)* 2/ (279)
Cancer - Inadequate	0.022	(1-0.9)* 61/ (279)
No Cancer - Malignant (FP)	0.012	(1-0.86)* 6/ (69) =(1-specificity) x malignant over (malignant,

		indeterminate and inadequate)
No Cancer - Benign (TN)	0.860	Specificity of UGFNAC
No Cancer - Indeterminate	0.004	(1-0.86)* 2/ (69)
No Cancer - Inadequate	0.124	(1-0.86)* 61/ (69)
PGFNAC probabilities		
Cancer - Malignant (TP)	0.710	Sensitivity of PGFNAC
Cancer - Benign (FN)	0.193	(1-0.71)* 186/ (280)Weighted average of 1-sensitivity over the benign, indeterminate and inadequate
Cancer - Indeterminate	0.002	(1-0.71)* 2/ (280)
Cancer - Inadequate	0.095	(1-0.71)* 92/ (280)
No Cancer - Malignant (FP)	0.009	(1-0.82)* 5/ (99)Weighted average of 1-specificity over the benign, indeterminate and inadequate
No Cancer - Benign (TN)	0.820	Specificity of PGFNAC
No Cancer - Indeterminate	0.004	(1-0.82)* 2/ (99)
No Cancer - Inadequate	0.167	(1-0.82)* 92/ (99)

A.2.3.3 Resource use and costs

A.2.3.3.1 Diagnostic costs

The diagnostic test costs were not available from standard NHS sources so to estimate the cost of a PG and UG FNAC test, the committee members were contacted. The costs were obtained from two hospital trusts known to test using FNAC. The committee found these costs to be plausible and they were used in the model see Table 6.

A.2.3.3.2 Surgery cost

NHS reference costs 2016/17 were used to determine the cost of thyroid surgery. The codes for thyroid surgery were obtained from the committee. Three HRG codes for 'thyroid procedures' with different complications and co morbidities (KA09C, KA09D and KA09E) were selected and then weighted by activity to calculate a weighted average cost of for thyroid surgery.

It was noted that the codes do not differentiate between different thyroid surgeries i.e. total thyroidectomy (removing all of the thyroid gland) lobectomy or hemithyroidectomy (removing half of the thyroid gland. The committee were happy to use the calculated weighted average cost for the thyroid procedures and test the ranges in the sensitivity analysis.

A.2.3.3.3 Delayed diagnosis cost

The model focused on the short-term cost of the diagnostic pathway. Patients with a false negative result would be discharged from the diagnostic pathway after their FNAC. However, given that they have cancer, it seems unreasonable not to have a cost penalty in the model for a FN. At a minimum we would expect that they would re-present at some point and go through the diagnostic pathway again, this time going on to have surgery. Therefore, we added the cost of delayed diagnosis to the patients with a FN diagnosis.

The costs included the procedure cost as calculated above in addition to an extra FNAC, an endocrinology follow up appointment, and radiology follow up appointment.

Furthermore, HRG codes for radiology and endocrinology follow-up appointments, 'non-admitted face to face attendance, follow-up', for both consultant led and non-consultant led, were selected and the weighted average was calculated and used in the model.

Table 5: False negative (FN) delayed diagnosis costs

Parameter description	Costs	Source
Consultant and non-consultant led radiology follow-up appointment [weighted average of HRG codes WF01A(service code 811)]	£66	NHS Reference costs 2016/17
Consultant and non-consultant led endocrinology follow-up appointment [weighted average of HRG codes WF01A(service code 811)]	£147	NHS Reference costs 2016/17
UGFNAC	£295	Committee
Surgery	£3,689	NHS Reference costs 2016/17
Total for delayed diagnosis	**£4,197**	

Abbreviations: UGFNAC: Ultrasound guided fine-needle aspiration cytology; NHS: National Health Service

A.2.3.4 Summary table of model inputs

A summary of the model inputs used in the base-case analysis is provided in Table 6 below.

Table 6: Summary of base-case model inputs and parameter distributions used in the model

Parameter description	Point estimate	Source	Distribution
Diagnosis parameters			
Prevalence of cancer among patients with a normally functioning but enlarged thyroid	0.05	Borget, et al 2018[1]	Beta
Positive predictive value (PPV) of US	0.115	Calculation	Function of the prevalence of cancer above and the Sensitivity and Specificity of ultrasound (see A.2.4)
Sensitivity of US	0.904	Persichetti 2018 [7]	Function of the prevalence and DOR of US (see A.2.4)
Specificity of US	0.634	Persichetti 2018 [7]	Beta
Diagnostic odds ratio (DOR) of US	16.295	Function of sensitivity and specificity	Log Normal
Sensitivity of UGFNAC	0.900	Pooled estimate	Sampled from the joint distribution from WinBUGS
Specificity of UGFNAC	0.865	Pooled estimate	Sampled from the joint distribution from WinBUGS
Sensitivity of PGFNAC	0.71	Pooled estimate	Sampled from the joint distribution from WinBUGS
Specificity of PGFNAC	0.82	Pooled estimate	Sampled from the joint distribution from WinBUGS
Cost (£)			
UGFNAC	£295	Committee member	Gamma

Parameter description	Point estimate	Source	Distribution
PGFNAC	£242	Committee member	Gamma
Surgery	£3,689	NHS reference costs 2016/17	Gamma
FN cost (delayed diagnosis)	£4,197	NHS reference costs 2016/17, see Table 5	Gamma

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

A.2.4 Computations

The model was constructed in Microsoft Excel 2010.

To account for uncertainty around diagnostic accuracies and correlation between sensitivity and specificity, a joint distribution needs to be used when making diagnostic accuracies probabilistic.

To ensure the inverse correlation between the sensitivity and specificity is maintained we need to make the sensitivity a function of the specificity (or vice versa). This way we will only run one variable probabilistically and the other will vary accordingly.

The diagnostic odds ratio (DOR) was calculated as follows:

$$DOR = sensitivity/\ (1 - sensitivity) * specificity/\ (1 - specificity)$$

The standard error of the log DOR is calculated using the absolute values for the number of TP, TN, FP, and FN:

$$SE\ (ln\ (DOR))\ = \sqrt{}\ (1/TP + \ 1/FN + \ 1/TN + \ 1/FP)$$

The sensitivity can be expressed as a function of the specificity and DOR

$$Sensitivity = 1 - \ specificity/\ (specificity + \ (1 - \ specificity)\ * \ DOR)$$

To make this probabilistic, a normal distribution is fitted to the log of the DOR and a beta distribution is fitted around the specificity of the test.

In the deterministic and probabilistic analyses, for a cohort of 1000 patients, the number of patients and costs accrued by each subgroup (TP, FP, TN, and FN) was recorded. The total cost accrued by the whole cohort was divided by the number of TPs to calculate a cost per cancer detected.

A.2.5 Sensitivity analyses

A.2.5.1 One- way sensitivity analyses

A.2.5.1.1 Cancer prevalence

A sensitivity analysis was conducted assessing the effect of higher and lower prevalence of thyroid cancer. This was varied in the range found in the clinical papers between 4%-7%.

A.2.5.1.2 Cost of UGFNAC

Due to the assumption the committee made around the cost of UGFNAC costs, a one-way sensitivity analysis was conducted to explore the uncertainty. A minimum cost of £221 and a maximum cost of £369 were explored in one-way sensitivity analyses. This range was derived by plus/ minus 25% in the base case cost, which was believed to be a plausible range.

A.2.5.1.3 Cost of PGFNAC

A one-way sensitivity analysis was conducted to explore the uncertainty around the assumption made for the cost of PGFNAC. A minimum cost of £182 and a maximum cost of £303 were explored in one-way sensitivity analyses. Again, this range was assumed to be plus/ minus 25% of the base case cost.

A.2.5.1.4 Cost of surgery

A one-way sensitivity analysis was conducted to assess the effect of altering the cost of the surgery. A minimum cost of £2, 767 and a maximum cost of £4,611 were explored in one-way sensitivity analyses. Again, the range was assumed to be plus/ minus 25% of the base case cost..

The committee was aware that hemithyroidectomy might be cheaper than the total thyroidectomy but as the cost codes do not differentiate between different thyroid surgeries, it was difficult to calculate the costs of the different thyroid surgeries and therefore the committee were particularly interested in a cost reduction of the surgery. As people undergoing hemithyroidectomy are likely to stay in hospital for a shorter time and have a quicker recovery and would not need long term thyroxine treatment.

A.2.5.1.5 Cost of FN (delayed diagnosis)

One-way sensitivity analyses were conducted varying the cost of delayed diagnosis. It was varied by plus/ minus 25% of the base case cost (lower cost £3,148- higher cost £5,246).

A.2.5.2 Two-way sensitivity analyses

The committee considered it important that sensitivity analyses were conducted to test the robustness of the model results to changes in sensitivity and specificity data. Essentially these sensitivity analyses show how these parameters affect the model results (Impact on cost per cancer detected).

As sensitivity and specificity have an inverse correlation a two-way sensitivity (changing two parameters at a time) analysis was considered for the US, UG and PG FNAC.

A.2.5.2.1 Ultrasound: sensitivity and specificity

The sensitivity and specificity values from the *Persichetti 2018* [7] study, used in the deterministic sensitivity analysis are reported in Table 7 below. We conducted one sensitivity analysis with a high estimate of sensitivity and a low estimate of specificity. A second sensitivity analysis used a low estimate of sensitivity and a high estimate of specificity

A.2.5.2.2 UGFNAC: sensitivity and specificity

The heterogeneity found between the diagnostic accuracy studies implies uncertainty about test accuracy. We conducted one sensitivity analysis with a high estimate of sensitivity and a low estimate of specificity. A second sensitivity analysis used a low estimate of sensitivity and a high estimate of specificity.

The high and low estimates were the 95% confidence limits were from the diagnostic meta-analysis of the included studies in the review. They are reported in Table 7**.**

A.2.5.2.3 PGFNAC: sensitivity and specificity

As for UGFNAC, we conducted one sensitivity analysis with a high estimate of sensitivity and a low estimate of specificity. Similarly, a second sensitivity analysis used a low estimate of sensitivity and a high estimate of specificity **- see** Table 7 below.

Table 7: Diagnostic accuracy data used in the two way sensitivity analysis

Diagnostic accuracy	Sensitivity	Specificity
US: Low sensitivity and high specificity	0.85	0.67
US: High sensitivity and low specificity	0.95	0.60
UGFNAC: Low sensitivity and high specificity	0.76	0.96
UGFNAC: High sensitivity and low specificity	0.98	0.72
PGFNAC: Low sensitivity and high specificity	0.48	0.96
PGFNAC: High sensitivity and low specificity	0.87	0.59

A.2.6 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 National Guideline Centre (NGC); this included systematic checking of the model calculations.

A.2.7 Estimation of cost-effectiveness

In a secondary analysis, the cost per cancer detected was estimated. 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 effects (in this case the number of cancers detected or TPs). If both costs are lower and TPs are higher, the option is said to dominate its comparator and an ICER is not calculated.

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

Where: Costs(A) = total costs for option A; TP(A) = total number of true positives for option A (for a given cohort size)

When there are more than two comparators, as in this analysis, options must be ranked in order of decreasing 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 two other options would prove to be less costly and more effective.

A.2.8 Interpreting Results

NICE's report 'Social value judgements: principles for the development of NICE guidance' [6] sets out the principles that Committees should consider when judging whether an intervention offers good value for money.

- An intervention was considered unambiguously cost-effective if (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).

- Otherwise it was deemed likely to be cost effective if the incremental cost was deemed small relative to the number of extra cases detected.

A.3 Results

The diagnostic accuracy results for each strategy from the probabilistic analysis are shown in Table 8 below. It shows that UGFNAC without benign repeat produces the least FP and TN whereas UGFNAC with benign repeat resulted in the highest TP and the lowest FN. PGFNAC with benign repeat resulted in the most FPs.

Table 8: Breakdown of diagnostic accuracy (probabilistic analysis)

Diagnostic Accuracy	UGFNAC without benign repeat	UGFNAC+ benign repeat	PGFNAC without benign repeat	PGFNAC+ benign repeat
True Positives	105	113	90	107
False Negatives	9	1	25	7
False Positives	28	128	44	169
True Negatives	857	757	841	717

A.3.1 Base case

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

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

However, UGFNAC with benign repeat was dominant compared to PGFNAC with benign repeat, as PGFNAC with benign repeat was more costly and less effective in detecting cancer.

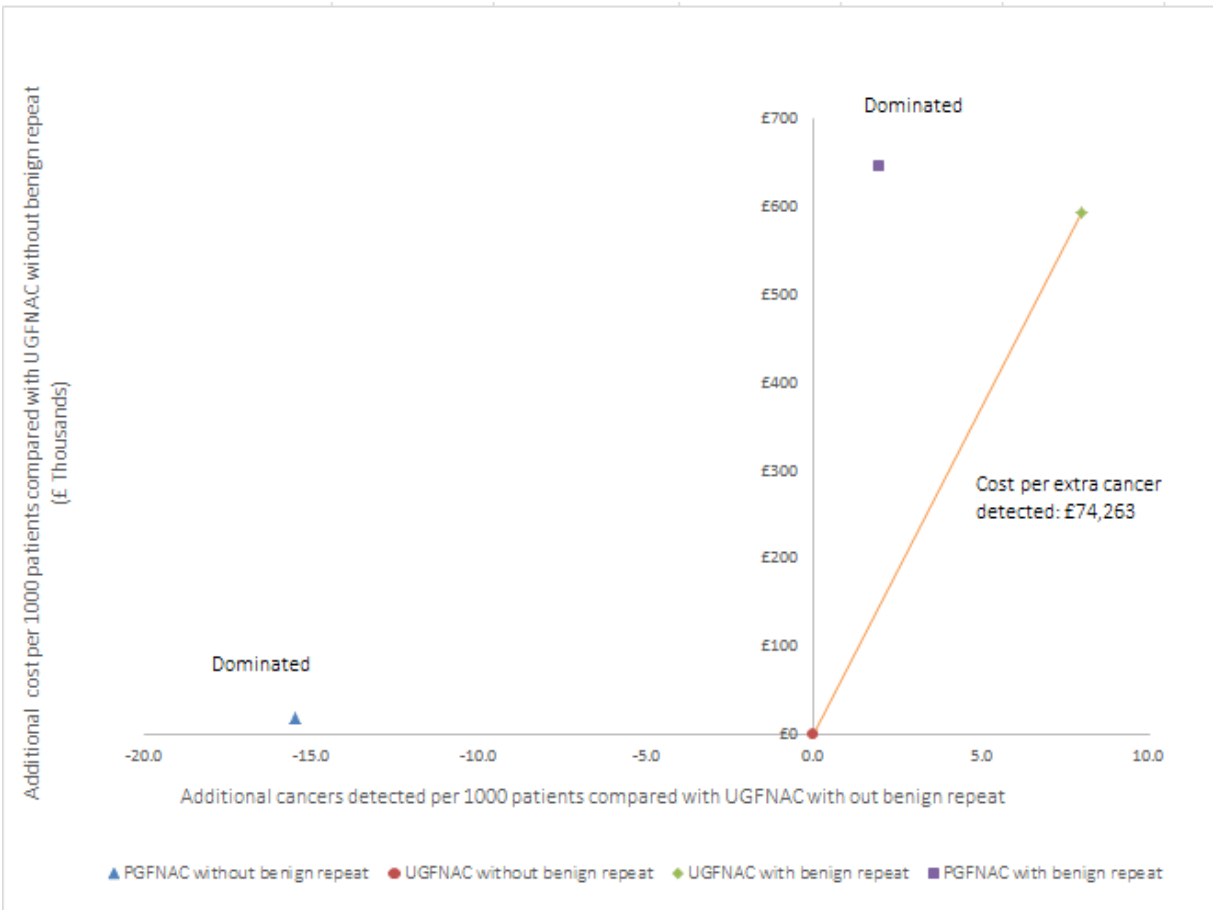
Results are summarised below in Table 9. The incremental costs and true positives from the probabilistic analysis have also been presented graphically on the cost-effectiveness plane, Figure 3.

Table 9: Base case analysis results per 1000 patients (probabilistic analysis)

Strategy	Costs	Cancers detected (True Positives)	Additional Cost (compared with row above)	Additional cancers detected (compared with row above)	Additional cost per extra cancer detected	Probability of the cheapest strategy	Probability of the strategy with the least FN

UGFNAC without benign repeat	£858,462	106	-	-	-	48%	0%
PGFNAC without benign repeat	£877,820	90	£19,359	-15.5	Dominated	52%	0%
UGFNAC with benign repeat	£1,451,699	114	£573,878	23.5	£74,263 (vs UGFNAC without benign repeat)	0%	95%
PGFNAC with benign repeat	£1,505,353	107	£53,655	-6.1	Dominated	0%	5%

Figure 3: Base case cost-effectiveness plane showing the different diagnostic strategies (probabilistic)



A.3.2 Sensitivity analyses

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

The results of the one- way sensitivity analyses were run deterministically and are summarised in Table 10 below. These showed that in general, changes in the cost of test or treatment do not result in very different estimates of the cost per cancer detected.

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

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

In each case, PGFNAC was no longer dominated but UGFNAC had a low cost per extra cancer detected.

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

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

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

Table 10: One-way sensitivity analysis

Analysis	Cost per cancer detected		
	UGFNAC no benign repeat vs PGFNAC no benign repeat	UGFNAC+ benign repeat vs UGFNAC no benign repeat	UGFNAC+ benign repeat vs PGFNAC+ benign repeat
Base Case (Deterministic)	Dominant	£73,492	Dominant
Base case (Probabilistic)	Dominant	£74,263	Dominant
Cancer Prevalence			
Low	Dominant	£92,881	Dominant
High	Dominant	£51,333	Dominant
Cost UGFNAC			
Low	Dominant	£66,567	Dominant
High	£5,204	£80,416	£15,164
Cost PGFNAC			
Low	£4,457	£73,492	£11,022
High	Dominant	£73,492	Dominant
Cost Surgery			
Low	Dominant	£60,994	Dominant
High	£183	£85,989	Dominant
FN Cost			
Low	£954	£74,541	Dominant
High	Dominant	£72,443	Dominant

A two-way sensitivity analysis was run and the results are summarised in the Table 11 below. The results were sensitive to changes in the sensitivity and specificity of the FNAC and less sensitive to the sensitivity and specificity of US.

Table 11: Two-way sensitivity analysis

Analysis	Cost per cancer detected		
	UGFNAC no benign repeat vs PGFNAC no benign repeat	UGFNAC+ benign repeat vs UGFNAC no benign repeat	UGFNAC+ benign repeat vs PGFNAC+ benign repeat vs
Base Case (Deterministic)	Dominant	£73,492	Dominant
Base case (Probabilistic)	Dominant	£74,263	Dominant
UGFNAC			
Low sensitivity high	Dominant	£ 21,592	Dominant

specificity			
High sensitivity low specificity	£ 11,119	£ 417,020	£ 67,372
PGFNAC			
Low sensitivity high specificity	£ 3,578	£ 73,492	£ 20,718
High sensitivity low specificity	Dominant	£ 73,492	Dominant
US			
Low sensitivity high specificity	Dominant	£ 70,487	Dominant
High sensitivity low specificity	Dominant	£ 76,461	Dominant

A.4 Discussion

A.4.1 Summary of results

This analysis found that UGFNAC without benign repeat was the lowest cost option compared to all the other strategies and resulted in the least false positives. UGFNAC without benign repeat detected more cancers then PGFNAC without benign repeat. The PGFNAC without benign repeat was dominated in most sensitivity analyses as shown in the tables above compared to UGFNAC without benign repeat.

A.4.2 Limitations and interpretation

This analysis suggests that UGFNAC without benign repeat dominated PGFNAC without benign repeat for diagnosing thyroid cancer in patients with a positive US scans results. Uncertainties in the model structure and assumptions were explored in sensitivity analyses.

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

The second limitation of this model is that the **diagnostic accuracy data for the US scan** was taken from one diagnostic accuracy study. A meta-analysis was discussed but it was decided that for a meaningful meta-analysis, five or more studies were needed. The committee agreed on choosing one study to represent best available evidence, which was the study by *Persichetti 2018*[7] that was more representative of UK current practice.

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

A fourth limitation of this model is that some structural assumptions were required with little clinical evidence to allow direct estimates to be made. In particular, it is difficult to test the assumptions made about the **suspicious results that were grouped together with the indeterminate (Thy3A) results.** The committee had a lengthy discussion to split the group into indeterminate and suspicious

but there was no consensus and the clinical evidence did not help quantify this issue. It was therefore agreed that for simplicity of the model, they are to be grouped together.

A.4.3 Comparisons with published studies

One cost-accuracy analysis (Cesur et al. 2006) was identified that compared PGFNAC with UGFNAC, in adults admitted to the outpatient thyroid clinic with nodular goitre.[2] The sensitivity and specificity of the strategies were calculated based on the cytologic and histologic results obtained from the 26 nodules out of 285 nodules, which underwent surgery. There were no clear inclusion criteria for surgery and therefore the study was downgraded for introducing bias.

Cesur et al. 2006 established the rates of inadequate material. The study only repeated FNAC in the inadequate sample and the repeat FNAC was always by UGFNAC. The costs included in this study were obtained from several hospitals in Turkey, and only included costs of the procedures. The mean cost per patient were £64 for the UGFNAC and £51 for the PGFNAC. This analysis found that in adults with nodular goiter, UGFNAC was more costly and more effective than PGFNAC for detecting malignancy (ICER: £1,361 per extra cancer detected). UGFNAC was found to be superior to PGFNAC both in obtaining adequate material and in terms of diagnostic accuracy.

The study had potentially serious limitations as the data was taken from a single study, cost year were not stated, and costs were taken from private and state hospitals in Turkey. In addition, no sensitivity analysis was undertaken in this study.

The committee felt that in UK practice, a percentage of the benign would get a repeat FNAC especially if it was initially done by PG rather than UG. Therefore, our analysis compares UG with PG with and without benign being repeated, to reflect clinical practice. This analysis demonstrated similar results to the Cesur et al. 2006 study, finding UGFNAC without benign repeat to be more costly and more effective compared to PGFNAC without benign repeat.

Overall, the committee agreed that both approaches to modelling yielded suggested that the optimal strategy was to preform FNAC under UG.

A.4.4 Conclusions

An original cost-consequence analysis found that UGFNAC without benign repeat was the lowest cost option and was dominant compared to the PGFNAC without benign repeat (less costly and more effective in detecting cancer). PGFNAC with benign repeat was dominated by UGFNAC with benign repeat, as it is less costly and more effective at detecting cancer. The committee noted that the UGFNAC with benign repeat is unlikely to be cost effective compared to UGFNAC without benign repeat as the cost per extra cancer detected £74,263, was considered relatively high. The committee concluded that UGFNAC without benign repeat is also better then UGFNAC with benign repeat, because it results in far fewer false negatives. This will reduce costs but also improve patient's quality of life. This analysis was assessed as having minor limitations. Since it does not estimate QALYs and does not have a lifetime horizon it is considered to be partially applicable.

A.4.5 References

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

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

3. Farihah AG, Nurismah MI, Husyairi H, Shahrun Niza AS, Radhika S. Reliability of the ultrasound classification system of thyroid nodules in predicting malignancy. Medical Journal of Malaysia. 2018; 73(1):9-15

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

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

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

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

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

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