National Institute for Health and Care Excellence

FINAL

Thyroid disease: assessment and management

[H] Tests for people with confirmed thyrotoxicosis

NICE guideline NG145

Diagnostic evidence review underpinning recommendations 1.6.1 to 1.6.4 in the guideline

2019

FINAL

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

Disclaimer

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

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

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 All NICE guidance is subject to regular review and may be updated or withdrawn.

ISBN: 978-1-4731-3595-6

Contents

1

1 Antibodies in hyperthyroidism

1.1 Review question: What is the accuracy of anti-thyroid peroxidase antibodies (anti-TPO) testing, thyroid stimulating hormone (TSH) receptor antibodies (TRAb) testing, ultrasound scanning and isotope scanning for diagnosing Graves' disease?

What is the clinical and cost effectiveness of using anti-TPO testing, TRAb testing, ultrasound scanning or isotope scanning in the diagnosis of Graves' disease?

1.2 Introduction

Graves' disease (autoimmune hyperthyroidism) is the commonest cause of thyrotoxicosis. A correct diagnosis of Graves' disease is important as treatment of thyrotoxicosis depends upon the cause. For example, whilst patients with Graves' disease usually need treatment with antithyroid drugs, radioiodine or thyroidectomy, thyrotoxicosis due to thyroiditis is self-limiting and patients only require treatment for symptom relief. Furthermore, patients with Graves' disease are at risk of developing other extra-thyroidal disorders, such as thyroid eye disease. Therefore, the correct diagnosis of Graves' disease will help the patient to be aware of the risks, allowing them to seek clinical advice promptly in case of new eye symptoms and to take steps to prevent thyroid eye disease (for example, stopping smoking).

Although careful clinical history and physical examination can provide clues to the cause of thyrotoxicosis, most patients require laboratory or imaging investigations to confirm the aetiological diagnosis. Several investigations are commonly used in the clinical practice to investigate a patient with suspected Graves' disease, including measurement of thyroid autoantibodies (TPO-Ab and TSHR-Ab), thyroid ultrasound and thyroid isotope uptake scan. In the past, TPO-Ab (and TG-Ab) concentrations have been widely used to investigate autoimmune thyroid diseases, including Graves' disease. However, in the recent years, second and third generation assays for TSHR-Ab have become more widely available for routine use in the clinical practice, with many centres (but not all) preferring TSHR-Ab to TPO-Ab for investigating a patient with suspected Graves' disease. Furthermore, some centres also use thyroid isotope uptake scan and thyroid ultrasound for the investigation of thyrotoxicosis. There is currently no national standard, and there is a variation in the choice and sequence of the investigations for thyrotoxicosis in routine clinical practice.

1.3 PICO table

For full details see the review protocol in appendix A.

Table 1: PICO characteristics of review question

Population	People diagnosed with hyperthyroidism who are being investigated for Graves' disease
Target condition	Graves' disease
Index tests	Anti-TPO testing TRAb testing Ultrasound scan Isotope scan

Reference standards	**Diagnostic accuracy data:** Reference standard to be determined by include studies, likely to include some composite of TRAb, multiple investigations, eventual clinical progression. To be specified in review on a study by study basis and impact on risk of bias considered **Test and treat data:** Any of above testing strategies compared with any other
Statistical measures [or] Outcomes	**Diagnostic accuracy data:** Sensitivity Specificity Specificity will be prioritised **Test and treat data:** • Critical Mortality (dichotomous) Quality of life (continuous) • Important Healthcare contacts (rates/dichotomous) Experience of care (continuous)
Study design	**Test and treat data:** RCTs preferred, if no RCTs available to consider non-randomised cohort studies in which key confounders (age, sex, co-existing conditions) are addressed, either through restriction or appropriate matching/statistical adjustment **Diagnostic accuracy data:** Two gate study designs will be excluded Prospective studies prioritised, retrospective studies included if insufficient prospective studies identified Minimum duration of follow-up 3 months Crossover studies excluded

1.4 Clinical evidence

1.4.1 Included studies

Seven studies were included in the review; [5, 34, 50, 55, 65, 66, 70] these are summarised in Table 2 below. Evidence from these studies is summarised in the clinical evidence summary below (Table 3).

Two studies were in children. Five studies were in adults. Five studies assessed accuracy of some form of TRAb, two studies assessed accuracy of ultrasound and one study assessed accuracy of Technetium 99 scans.

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

1.4.2 Excluded studies

See the excluded studies list in appendix H.

1.4.3 Summary of clinical studies included in the evidence review

Table 2: Summary of studies included in the evidence review

Study	Population	Target condition	Index test	Reference standard	Comments
Baskaran 2015 [5]	Children n=47, mean age (SD, range): 12.3 (4.6); GD (n=37) 11.7 years (4.4, 2.4-17.7 years); non-GD (n=10) 14.8 years (4.5, 5.5-18.6 years) USA	Graves' disease	Technetium 99 (^{99m}Tc) scan TSH receptor stimulating immunoglobulins (TSI)	Laboratory tests and clinical progress (clinical presentation, successful treatment with antithyroid medication, surgery or radioactive ablation)	^{99m}Tc uptake ≤ 0.4% was considered to be decreased/negative and suggestive of non-GD thyroiditis; any uptake that was either increased or inappropriately normal was considered positive and suggestive of GD.

Study	Population	Target condition	Index test	Reference standard	Comments
Lee 2016 [36]	Children n=113; mean age (range): 12 years (6-19 years); Patients with diffuse swelling of the anterior neck or an enlarged thyroid gland by ocular inspection or palpation finally included (n=86: autoimmune thyroiditis n=26; Graves' disease n=14; simple goiter n=46) South Korea	Graves' disease	Ultrasound (gray-scale & Doppler US)	Radioimmunoassay of antithyroid antibody levels (including anti-TPO, antithyroglobulin, anti-thyroid-stimulating hormone receptor antibodies)	Independent sonographic criteria for the identification of autoimmune thyroid disease (Hypoechogenicity, Coarse echotexture, micronodularity, increased vascularity) 12 out of 14 children with Graves' disease had overt hyperthyroidism (euthyroidism n=1, subclinical hyperthyroidism =1)
Paunkovic 2006[50]	Adults n=255; median age 52 Patients presenting to clinic with symptoms of hyperthyroidism Serbia	Graves' disease	TSH receptor assay (combination of TBII and TBIII, majority TBII)	Clinical impression (including eye signs) combined with biochemical criteria	Repeated TSH assay in those who were negative, both a TBII and TBIII test
Pishdad 2017 [55]	Adults n=149; Graves' disease n=34, mean age (SD): 36.8 (10.17) years; Hashimoto's thyroiditis n=62, mean age (SD): 33.4 (12.16) years; healthy controls n=53, mean age (SD): 34.74 (16.87)	Graves' disease	Ultrasound (gray scale)	Clinical and laboratory data including thyroid hormone levels, and anti-thyroid antibodies.	Diagnostic accuracy of different sonographic patterns (homogenously hypoechoic, peripherally hypoechoic, centrally hypoechoic, homogenously isoechoic, homogenously hyperechoic)

Study	Population	Target condition	Index test	Reference standard	Comments
	Iran				
Sulman 1990[65]	Adults n=190; clinically examined for hyper and hypometabolism symptoms, assessment of a possible goiter and signs of Graves' ocular or skin disease France	Graves' disease	TSH receptor assay (TB II)	Clinical examination and biological analysis combined	
Syme 2011[66]	Adults n=102; patients attending first appointment at thyroid clinic between 2008 and 2009 UK	Graves' disease	TSH receptor assay (TB III)	Clinical examination with biochemistry and t-99 scan in 70 patients to aid diagnosis	
Theodoraki 2011[70]	Adults n=244; two cohorts (one prospective, one retrospective), patients attending clinic where only those with hyperthyroid symptoms and no history of Graves or obvious clinical signs of Graves (assumed to be diagnostic) are investigated further UK	Graves' disease	TSH receptor assay (TB III)	Final recorded clinical diagnosis	

See appendix D for full evidence tables.

1.4.4 Quality assessment of clinical studies included in the evidence review

Table 3: Clinical evidence summary: diagnostic tests in adults

Index Test (Threshold)	Number of studies	n	Quality	Sensitivity % (95% CI)	Specificity % (95% CI)
TRAb					
TRAb TB II/III (threshold not specified	2	435	VERY LOW[a,b,c] due to risk of bias, serious inconsistency and serious imprecision	99 (96 to 100) 88 (80 to 93)	100 (84 to 100) 85 (74 to 93)
TRAb TB III, 0.4U/L	1	244	MODERATE[a] due to risk of bias	86 (80 to 91)	94 (87 to 98)
TRAb TB III, 0.9U/L	1	102	LOW[a,c] due to risk of bias, serious imprecision	100 (94 to 100)	89 (76 to 96)
TRAb TB III, 1.6U/L	1	102	MODERATE[a] due to risk of bias	95 (85 to 99)	98 (88 to 100)
TRAb TB III, 1.75U/L	1	102	MODERATE[a] due to risk of bias	93 (83 to 98)	100 (92 to 100)
TRAb TB III, 1.86U/L	1	102	MODERATE[a] due to risk of bias	91 (80 to 97)	100 (92 to 100)
Ultrasound					
Peripherally hypoechoic	1	149	LOW[ac] due to risk of bias, serious imprecision	15 (5 to 31)	100 (93 to 100)
Centrally hypoechoic	1	149	LOW[a,c] due to risk of bias, serious imprecision	18 (7 to 35)	100 (93 to 100)
Homogenously hypoechoic	1	149	LOW[a,c] due to risk of bias, serious imprecision	47 (30 to 65)	91 (79 to 97)
Homogenously isoechoic	1	149	LOW[a,c] due to risk of bias, serious imprecision	6 (1 to 20)	51 (37 to 65)
Homogenously	1	149	VERY LOW[a,c]	15 (5 to 31)	58 (44 to 72)

Index Test (Threshold)	Number of studies	n	Quality	Sensitivity % (95% CI)	Specificity % (95% CI)
hyperechoic			due to risk of bias, very serious imprecision		

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

(b) Inconsistency was assessed by inspection of the sensitivity and specificity plots. The evidence was

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

(c) Imprecision was assessed according to the range of confidence intervals in the individual studies. The evidence was downgraded by 1 increment when the range of the confidence intervals around the point estimate was 20–40%, and downgraded by 2 increments when there was a range of >40%

Table 4: Clinical evidence summary: diagnostic tests in children

Index Test (Threshold)	Number of studies	n	Quality	Sensitivity % (95% CI)	Specificity % (95% CI)
TRAb					
TSI	1	47	LOW[a,b] due to risk of bias, serious imprecision	84 (68 to 94)	100 (69 to 100)
Isotope					
Technetium 99	1	47	LOW[a,b] due to risk of bias, serious imprecision	100 (91 to 100)	100 (69 to 100)
Ultrasound					
Hypoechogenicity (US + Doppler)	1	113	LOW[b] due to very serious imprecision	86 (57 to 98)	67 (55 to 77)
Coarse echotexture (US + Doppler)	1	113	LOW[b] due to very serious imprecision	64 (35 to 87)	74 (62 to 83)
Micronodularity (US +	1	113	MODERATE[b]	7 (0 to 34)	81 (70 to 89)

Index Test (Threshold)	Number of studies	n	Quality	Sensitivity % (95% CI)	Specificity % (95% CI)
Doppler)			due to serious imprecision		
Increased vascularity (US + Doppler)	1	113	LOW[b] due to very serious imprecision	71 (42 to 92)	92 (83 to 97)

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

(b) Imprecision was assessed according to the range of confidence intervals in the individual studies. The evidence was downgraded by 1 increment when the range of the confidence interval around the point estimate was 20–40%, and downgraded by 2 increments when there was a range of >40%

1.5 Economic evidence

1.5.1 Included studies

No relevant health economic studies were identified.

1.5.2 Excluded studies

No health economic studies that were relevant to this question were excluded due to assessment of limited applicability or methodological limitations.

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

1.5.3 Health economic modelling

This area was not prioritised for new cost-effectiveness analysis.

1.5.4 Resource costs

Relevant unit costs are provided below to aid consideration of cost effectiveness.

Table 5: UK costs of different interventions in the diagnosis of Graves' disease

Interventions	Unit costs
Ultrasound scan (a)	£53.22
Thyroid Gland Scan, 19 years and over (b)	£258
Thyroid Gland Scan, 18 years and under (c)	£222
TRAb antibody testing (d)	£16.64
TPO antibody testing (e)	£12.32

Source[s]: NHS reference costs 2017-18
(a) Ultrasound Scan with duration of less than 20 minutes and over 20 minutes, without contrast, NHS ref cost code: RD40Z, RD42Z
(b) Thyroid gland scan, including the intravenous injection of radiotracer technetium, NHS ref cost code; RN32A
(c) Thyroid gland scan, including the intravenous injection of radiotracer technetium, NHS ref cost code; RN32B
(d) Average costs obtained from two hospitals from the GC members
(e) Average costs obtained from two hospitals from the GC members

1.6 Evidence statements

1.6.1 Clinical evidence statements

Seven studies, two of which were conducted in children were included in the review. Five studies examined the diagnostic accuracy of TRAb (Thyrotropin-Binding Inhibitory Immunoglobulin-TBI using different thresholds in adults, Thyroid-Stimulating Immunoglobulin- TSI in children) for Graves' disease; two studies assessed the diagnostic accuracy of ultrasound and one study assessed the diagnostic accuracy of Technetium 99 scan.

1.6.1.1 TRAb in Adults

- **TB II/III** (threshold not specified): very low quality evidence from two studies with 435 participants showed that TB II/III has a sensitivity range of 88-99% and a specificity of 85-100%.

- **TB III (0.4 U/L):** moderate quality evidence from one study with 244 participants showed that using a 0.4 U/L cut-off, TB III has a sensitivity of 86% and a specificity of 94%.
- **TB III (0.9 U/L):** low quality evidence from one study with 102 participants showed that using a 0.9 U/L cut-off, TB III has a sensitivity of 100% and a specificity of 89%
- **TB III (1.6 U/L):** moderate quality evidence from one study with 102 participants showed that using a 1.6 U/L cut-off, TB III has a sensitivity of 95% and a specificity of 98%.
- **TB III (1.75 U/L):** moderate quality evidence from one study with 102 participants showed that using a 1.75 U/L cut-off, TB III has a sensitivity of 93% and a specificity of 100%.
- **TB III (1.86 U/L):** moderate quality evidence from one study with 102 participants showed that using a 1.86 U/L cut-off, TB III has a sensitivity of 91 % and a specificity of 100%.

1.6.1.2 Ultrasound in adults (diagnostic accuracy of individual features)

- **Peripherally hypoechoic:** low quality evidence from one study with 149 participants showed that peripheral hypoechogenicity has a sensitivity of 15% and a specificity of 100%.
- **Centrally hypoechoic:** low quality evidence from one study with 149 participants showed that central hypoechogenicity has a sensitivity of 18% and a specificity of 100%.
- **Homogenously hypoechoic:** low quality evidence from one study with 149 participants showed that homogenous hypoechogenicity has a sensitivity of 47% and a specificity of 91%.
- **Homogenously isoechoic:** low quality evidence from one study with 149 participants showed that a homogenously isoechoic sonographic pattern has a sensitivity of 6% and a specificity of 51%.
- **Homogenously hyperechoic:** very low quality evidence from one study with 149 participants showed that a homogenously hyperechoic sonographic pattern has a sensitivity of 15% and a specificity of 58%.

1.6.1.3 TRAb in Children

- **TSI:** low quality evidence from one study with 47 participants showed that TSI has a sensitivity of 84% and a specificity of 100%.

1.6.1.4 Isotope scan in Children

- **Technetium 99:** low quality evidence from one study with 47 participants showed that ^{99m}Tc has sensitivity of 100% and a specificity of 100%.

1.6.1.5 Ultrasound in Children

- **Hypoechogenicity (US + Doppler):** low quality evidence from one study with 113 participants showed that hypoechogenicity on combined gray-scale and power Doppler ultrasound has a sensitivity of 86% and a specificity of 67%.
- **Coarse echotexture (US + Doppler):** low quality evidence from one study with 113 participants showed that coarse echotexture has a sensitivity of 64% and a specificity of 74%.
- **Micronodularity (US + Doppler):** moderate quality evidence from one study with 113 participants showed that micronodularity has a sensitivity of 7% and a specificity of 81%

- **Increased vascularity (US + Doppler):** low quality evidence from one study with 113 participants showed that increased vascularity has a sensitivity of 71% and a specificity of 92%.

1.6.2 Health economic evidence statements

- No relevant economic evaluations were identified.

1.7 The committee's discussion of the evidence

1.7.1 Interpreting the evidence

1.7.1.1 The outcomes that matter most

The diagnostic measures of sensitivity and specificity of TRAb, Ultrasound and isotope scanning for diagnosing Graves' disease were considered for this review. Specificity was deemed the most important measure by the committee and hence it was prioritised for decision making. No evidence was identified for the diagnostic accuracy of anti-TPO testing.

1.7.1.2 The quality of the evidence

Clinical evidence for the diagnostic accuracy of different forms of TRAb for Graves' disease was available from five studies, one of which was conducted in children. In adults the evidence identified was for the accuracy of second and third generation TRAb TB for the diagnosis of Graves' disease based on different thresholds, the majority being for the third generation TRAb TB. In children available evidence was for the diagnostic accuracy of TSI. The quality of the evidence for adults ranged from very low to moderate; the majority being of moderate quality and was downgraded due to risk of bias and occasionally inconsistency and imprecision. In children, the quality of the evidence was low and was downgraded due to risk of bias and imprecision.

Clinical evidence for the diagnostic accuracy of different sonographic patterns of ultrasound for Graves' disease was available from two studies, one of which was conducted in children and examined conventional ultrasound combined with power Doppler. Evidence for the different ultrasound patterns in adults ranged from very low to low, the majority being of very low quality and was downgraded for risk of bias and imprecision. In children, evidence for the different sonographic patterns ranged from low to moderate, the majority being of low quality and was downgraded due to imprecision.

Evidence was also available for the diagnostic accuracy of Technetium 99 scanning for Graves' disease in children. The quality of the evidence was low and was downgraded due to risk of bias and imprecision.

1.7.1.3 Benefits and harms

1.7.1.3.1 Diagnostic tests in adults

Evidence suggested that in adults, both measures of sensitivity and specificity were similarly high for the use of third and second generation TRAb TB and its different cut-off values, with sensitivity ranging from 88 to 100% and specificity ranging from 85 to 100%. Specifically, sensitivity of TRAb TB III was 100% for a cut-off at 0.9 U/L and unsurprisingly specificity was highest for TRAb TB III when higher cut-offs of 1.75 U/L and 1.86 U/L were used.

Evidence for the diagnostic accuracy of Anti-TPO testing was not available. However, based on clinical experience the committee agreed that anti-TPO testing alone is not as useful to confirm the diagnosis of Graves' disease as TRAb testing.

Evidence suggested that the diagnostic accuracy of the different sonographic patterns of ultrasound was consistently low in terms of sensitivity, with sensitivity being as low as 6% for a homogenously isoechoic pattern. Sensitivity of ultrasound was highest (47%) for a homogenously hypoechoic pattern. The specificity of ultrasound patterns was higher ranging from 58 to 100%, with highest specificity noted for a peripherally hypoechoic US pattern (100%), centrally hypoechoic US pattern (100%) and a homogenously hypoechoic US pattern (91%). The committee noted that these findings were derived from only one study and were thus not that informative. Based on the current evidence and their clinical experience, the committee agreed that ultrasound is of limited diagnostic value for Graves' disease. Although they noted that ultrasound can be informative in cases where nodules are present, previous tests for Graves' disease have not provided a definitive answer; or if surgery is planned, routine ultrasound of all goitre is likely to lead to over investigation of incidental findings.

The committee noted that biochemical results such as thyroid hormone levels are not informative of the cause of hyperthyroidism and are not used to diagnose Graves' disease. Due to a vague description of biochemical results being reported as the reference standard used to confirm the diagnosis of Graves' disease in some of the studies included, the committee could not be certain about the extent to which the reference standards used were sufficient, potentially reducing the validity of the findings.

Evidence for the diagnostic accuracy of technetium scanning was not available in adults. However, the committee noted that isotope scanning is likely to be useful in the diagnosis of Graves' in patients with history of thyroiditis and patients with painless thyroiditis including post-partum thyroiditis. It was noted that technetium scanning can be helpful in differentiating Graves' disease with other causes of thyroiditis including Hashimoto's thyroiditis. The committee also noted that it is possible for patients with Graves' disease to test negative on TRAb and that technetium scanning could be useful in cases where there is a negative TRAb test but Graves' disease is still suspected. The committee agreed that in adults, technetium scanning would be preferable to ultrasound. The committee also noted that it is important for any person receiving a test to be fully informed of the details of the investigation, for technetium scanning this includes the use of radioactive material.

1.7.1.3.2 Diagnostic tests in children

Evidence suggested that in children, the diagnostic accuracy of TSI TRAb was high showing 84% sensitivity and 100% specificity. The committee noted that this was demonstrated by only one study that included a relatively small number of children (n=47) but agreed on the diagnostic accuracy of TRAb testing for the diagnosis of Graves' disease. The committee also noted that although TPO testing alone is not likely to be as useful as TRAb testing for the diagnosis of Graves' disease, it could be useful as an adjunct in some cases where the absence of TRAb and presence of TPO indicates that thyrotoxicosis is more likely to resolve spontaneously.

Evidence from one study showed that the accuracy of Technetium 99 (T-99) scanning in diagnosis of Graves' disease in children was very high (resulting in 100% sensitivity and specificity). However, based on clinical experience the committee noted that in children ultrasound would be preferred over T-99 scanning.

Evidence for the diagnostic accuracy of ultrasound combined with power Doppler ultrasound in children varied across the individual ultrasound features with sensitivity ranging from 7 to 86% and specificity ranging from 67 to 92%. Both diagnostic accuracy measures were high for increased vascularity showing 71% sensitivity and 92 % specificity. The committee

agreed on the usefulness of ultrasound in children, but noted that TRAb testing is likely to be a more accurate diagnostic test.

1.7.2 Cost effectiveness and resource use

No health economic evidence was identified for this question.

The unit costs for the TRAb and TPO tests were obtained from two NHS hospitals and were presented to the committee. The average TRAb cost was £16.64 and the TPO was £12.32. It was noted that costs vary as pathology laboratories may add a handling fee to these costs. Additionally, the NHS reference unit cost (2017/18) for US and thyroid gland scans were presented to the committee. The weighted average cost of an US scan was £53.22 (NHS reference cost code RD40Z, RD42Z) and a thyroid gland scan that includes the technetium was estimated to be £258 for adults (NHS reference cost code RN32A), and £222 for patients 18 years, and under (NHS reference cost code RN32B).

The committee made a recommendation to offer TRAb testing to confirm Graves' disease, as it had a higher diagnostic accuracy (both higher sensitivity and specificity). Although TRAb appears to be slightly higher cost than TPO the committee noted that the higher diagnostic accuracy would mean less misdiagnosed patients (false negatives and false positives) who might go on to receive unnecessary treatment and in turn cost the NHS money. Furthermore, the committee noted that TPO testing was not sufficient alone for diagnosing Graves' disease and required further tests and scans. Hence increasing the cost of TPO testing as repeat tests and scans may be required. Overall, they agreed that TRAb was therefore likely to be more cost effective than TPO.

Based on their clinical experience, the committee agreed that in children, measuring TRAb and considering the measurement of TPOAb, to establish a diagnosis, was useful as their condition can deteriorate much quicker. TPOAb testing in children is also used to rule out Hashimoto's thyroiditis and guide treatment.

The recommendations made by the committee to consider technetium scanning would only be appropriate in a small population, that is, cases where there is a negative TRAb test in patients with thyrotoxicosis. The committee noted that this is likely to reduce the number of people with Graves' disease being missed (false negatives) and ensure they receive appropriate treatment in a timely manner. This should reduce any spending on the management of long-term complications such as, increased cardiovascular morbidity and bone-related complications, of undiagnosed Graves' disease and any unnecessary referrals and investigations of people whose symptoms are unexplained and who are looking for a cause for their symptoms. Furthermore, it will ensure that those who have a negative result from an initial test (TRAb) are appropriately managed and alternative diagnoses are explored.

In some centres, this recommendation might require a move to TRAb testing from anti-TPO testing and therefore this might have a significant resource impact. However, if TRAB testing enables more accurate differentiation between the different causes of thyrotoxicosis, there are likely to be reductions in unnecessary antithyroid treatment (including surgery) of people with transient thyroiditis and more timely and appropriate treatment choices for people with toxic nodular hyperthyroidism.

1.7.3 Other factors the committee took into account

The committee noted that, although thyroid eye disease (TED) was not in the scope of this guideline, patients with hyperthyroidism with negative TRAb test results but in whom thyroid TED was present, should still be assumed to have Graves' disease. The committee noted that thyrotoxicosis in a baby may reflect transplacental passage of maternal antibody or reflect a germline mutation in the TSH receptor.

References

1. Aleksic A, Aleksic Z, Stojanovic M. TSH receptor antibodies for confirming the diagnosis and prediction of remission duration, in newly diagnosed Graves' disease patients. Hellenic Journal of Nuclear Medicine. 2009; 12(2):146-50

2. Banaka I, Kaltsas G, Antoniou S, Kanakis G, Zilos A, Baltas CS et al. Prognostic value of vascularity index for the diagnosis of autoimmune thyroid disease. Jbr-Btr: Organe de la Societe Royale Belge de Radiologie. 2011; 94(4):185-90

3. Banaka I, Thomas D, Kaltsas G. Value of the left inferior thyroid artery peak systolic velocity in diagnosing autoimmune thyroid disease. Journal of Ultrasound in Medicine. 2013; 32(11):1969-78

4. Barbesino G, Tomer Y. Clinical review: Clinical utility of TSH receptor antibodies. Journal of Clinical Endocrinology and Metabolism. 2013; 98(6):2247-55

5. Baskaran C, Misra M, Levitsky LL. Diagnosis of pediatric hyperthyroidism: technetium 99 uptake versus thyroid stimulating immunoglobulins. Thyroid. 2015; 25(1):37-42

6. Bell L, Hunter AL, Kyriacou A, Mukherjee A, Syed AA. Clinical diagnosis of Graves' or non-Graves' hyperthyroidism compared to TSH receptor antibody test. Endocrine Connections. 2018; 7(4):504-510

7. Bosi E, Bianchi R, Ruotolo G, Bazzigaluppi E, Belloni C, Calori G et al. Diagnostic sensitivity of thyroid autoantibodies assessed in a population-based, cross-sectional study in adults. Autoimmunity Highlights. 2010; 1(2):83-6

8. Burman KD, Pandian R. Clinical utility of assays for TSH receptor antibodies. Endocrinologist. 1998; 8(4):284-290

9. Cappelli C, Gandossi E, Castellano M, Pizzocaro C, Agosti B, Delbarba A et al. Prognostic value of thyrotropin receptor antibodies (TRAb) in Graves' disease: a 120 months prospective study. Endocrine Journal. 2007; 54(5):713-20

10. Cardia MS, Lima N, Knobel M, Medeiros-Neto G. Evaluation of a coated-tube assay for antithyrotropin receptor antibodies in patients with Graves' disease and other thyroid disorders. Thyroid. 2004; 14(4):295-300

11. Carella C, Mazziotti G, Sorvillo F, Piscopo M, Cioffi M, Pilla P et al. Serum thyrotropin receptor antibodies concentrations in patients with Graves' disease before, at the end of methimazole treatment, and after drug withdrawal: evidence that the activity of thyrotropin receptor antibody and/or thyroid response modify during the observation period. Thyroid. 2006; 16(3):295-302

12. Costagliola S, Morgenthaler NG, Hoermann R, Badenhoop K, Struck J, Freitag D et al. Second generation assay for thyrotropin receptor antibodies has superior diagnostic sensitivity for Graves' disease. Journal of Clinical Endocrinology and Metabolism. 1999; 84(1):90-7

13. Diana T, Brown RS, Bossowski A, Segni M, Niedziela M, Konig J et al. Clinical relevance of thyroid-stimulating autoantibodies in pediatric graves' disease-a multicenter study. Journal of Clinical Endocrinology and Metabolism. 2014; 99(5):1648-55

14. Diana T, Wuster C, Kanitz M, Kahaly GJ. Highly variable sensitivity of five binding and two bio-assays for TSH-receptor antibodies. Journal of Endocrinological Investigation. 2016; 39(10):1159-65

15. Donkol RH, Nada AM, Boughattas S. Role of color Doppler in differentiation of Graves' disease and thyroiditis in thyrotoxicosis. World Journal of Radiology. 2013; 5(4):178-83

16. Doroudian S, Pedersen IB, Knudsen CS, Handberg A, Andersen SL. Comparison of three competitive immunoassays for measurement of TSH receptor antibodies in patients with Graves' disease. Scandinavian Journal of Clinical and Laboratory Investigation. 2017; 77(7):535-540

17. Duron F, Talbot JN, Feron R, Aubert P, Milhaud G. Clinical value of thyrotropin binding inhibiting immunoglobulins (TBII) assay in hyperthyroidism. Biomedicine and Pharmacotherapy. 1987; 41(7):383-8

18. Eckstein A, Esser J, Mann K, Schott M. Clinical value of TSH receptor antibodies measurement in patients with Graves' orbitopathy. Pediatric Endocrinology Reviews. 2010; 7 Suppl 2:198-203

19. Engler H, Riesen WF, Keller B. Anti-thyroid peroxidase (anti-TPO) antibodies in thyroid diseases, non-thyroidal illness and controls. Clinical validity of a new commercial method for detection of anti-TPO (thyroid microsomal) autoantibodies. Clinica Chimica Acta. 1994; 225(2):123-36

20. Gassner D, Stock W, Golla R, Roth HJ. First automated assay for thyrotropin receptor autoantibodies. Clinical Chemistry and Laboratory Medicine. 2009; 47(9):1091-5

21. Giovanella L, Ceriani L, Garancini S. Clinical applications of the 2nd generation assay for anti-TSH receptor antibodies in Graves' disease. Evaluation in patients with negative 1st generation test. Clinical Chemistry and Laboratory Medicine. 2001; 39(1):25-8

22. Giovanella L, Ceriani L, Garancini S. Evaluation of the 2nd generation radio-receptional assay for anti-TSH receptor antibodies (TRAb) in autoimmune thyroid diseases. Comparison with 1st generation and anti-thyroperoxidase antibodies (AbTPO). Quarterly Journal of Nuclear Medicine. 2001; 45(1):115-9

23. Heberling HJ, Bierwolf B, Lohmann D. Clinical experience with a radioreceptor assay for TSH-binding inhibiting immunoglobulins (TBII). Experimental and Clinical Endocrinology. 1988; 91(3):355-61

24. Hirooka Y, Li C, Takagi J, Gotoh M, Habu S, Yasaka-Nomura T et al. Comparison of new different assay systems for thyrotropin receptor antibodies with reference to thyroid-stimulating antibodies and thyroid stimulation-blocking antibodies in Graves' disease. International Journal of Clinical Pharmacology Research. 2004; 24(4):111-6

25. Iko BO. Grey scale ultrasonography of the thyroid gland, Nigeria. Tropical and Geographical Medicine. 1986; 38(1):21-7

26. Kamath C, Adlan MA, Premawardhana LD. The role of thyrotrophin receptor antibody assays in graves' disease. Journal of Thyroid Research. 2012; 2012:525936

27. Kamijo K. TSH-receptor antibody measurement in patients with various thyrotoxicosis and Hashimoto's thyroiditis: a comparison of two two-step assays, coated plate ELISA using porcine TSH-receptor and coated tube radioassay using human recombinant TSH-receptor. Endocrine Journal. 2003; 50(1):113-6

28. Kamijo K. Study on cutoff value setting for differential diagnosis between Graves' disease and painless thyroiditis using the TRAb (Elecsys TRAb) measurement via the fully automated electrochemiluminescence immunoassay system. Endocrine Journal. 2010; 57(10):895-902

29. Kamijo K, Murayama H, Uzu T, Togashi K, Olivo PD, Kahaly GJ. Similar clinical performance of a novel chimeric thyroid-stimulating hormone receptor bioassay and an automated thyroid-stimulating hormone receptor binding assay in Graves' disease. Thyroid. 2011; 21(12):1295-9

30. Kamijo K, Nagata A, Sato Y. Clinical significance of a sensitive assay for thyroid-stimulating antibodies in Graves' disease using polyethylene glycol at high concentrations and porcine thyroid cells. Endocrine Journal. 1999; 46(3):397-403

31. Khoo DH, Fok AC, Tan CE, Koh LK, Lim SC, Eng PH et al. Thyroid stimulating hormone receptor antibody levels in Singaporean patients with autoimmune thyroid disease. Annals of the Academy of Medicine, Singapore. 1997; 26(4):435-8

32. Kotwal A, Stan M. Thyrotropin receptor antibodies-an overview. Ophthalmic Plastic and Reconstructive Surgery. 2018; 34(4S Suppl 1):S20-S27

33. Laurberg P, Pedersen IB. Measurements of TSH receptor antibodies in differential diagnosis. Immuno-Analyse et Biologie Specialisee. 2006; 21(4):234-238

34. Lee B, Park JY, Shin HY, Park SH, Choi EB, Yoo J et al. What do korean women know and want to know about thyroid cancer? A qualitative study. Asian Pacific Journal of Cancer Prevention,. 2016; 17(6):2901-7

35. Lee JI, Jang HW, Kim SK, Choi JY, Kim JY, Hur KY et al. Diagnostic value of a chimeric TSH receptor (Mc4)-based bioassay for graves' disease. Korean Journal of Internal Medicine. 2011; 26(2):179-186

36. Lee SJ, Lim GY, Kim JY, Chung MH. Diagnostic performance of thyroid ultrasonography screening in pediatric patients with a hypothyroid, hyperthyroid or euthyroid goiter. Pediatric Radiology. 2016; 46(1):104-11

37. Lytton SD, Ponto KA, Kanitz M, Matheis N, Kohn LD, Kahaly GJ. A novel thyroid stimulating immunoglobulin bioassay is a functional indicator of activity and severity of graves' orbitopathy. Journal of Clinical Endocrinology and Metabolism. 2010; 95(5):2123-2131

38. Lytton SD, Schluter A, Banga PJ. Functional diagnostics for thyrotropin hormone receptor autoantibodies: bioassays prevail over binding assays. Frontiers in Bioscience (Landmark Edition). 2018; 23:2028-2043

39. Mariotti S, Ruf J, Caturegli P, Rossi V, Boniolo A, Piccolo P et al. Methodological approach and diagnostic usefulness of a new assay for anti-thyroid peroxidase autoantibodies. Annales de Biologie Clinique. 1989; 47(9):541-5

40. Marwaha RK, Tandon N, Kanwar R, Ganie MA, Bhattacharya V, Reddy DH et al. Evaluation of the role of ultrasonography in diagnosis of autoimmune thyroiditis in goitrous children. Indian Pediatrics. 2008; 45(4):279-84

41. Massart C, Sapin R, Gibassier J, Agin A, D'Herbomez M. Intermethod variability in TSH-receptor antibody measurement: Implication for the diagnosis of graves disease and for the follow-up of graves ophthalmopathy. Clinical Chemistry. 2009; 55(1):183-186

42. Maugendre D, Massart C. Clinical value of a new TSH binding inihibitory activity assay using human TSH receptors in the follow-up of antithyroid drug treated Graves' disease. Comparison with thyroid stimulating antibody bioassay. Clinical Endocrinology. 2001; 54(1):89-96

43. Meng Z, Zhang G, Sun H, Tan J, Yu C, Tian W et al. Differentiation between Graves' disease and painless thyroiditis by diffusion-weighted imaging, thyroid iodine uptake,

thyroid scintigraphy and serum parameters. Experimental and Therapeutic Medicine. 2015; 9(6):2165-2172

44. Morgenthaler NG, Nagata A, Katayama S, Bergmann A, Iitaka M. Detection of low titre TBII in patients with Graves' disease using recombinant human TSH receptor. Clinical Endocrinology. 2002; 57(2):193-8

45. Morris JC, 3rd, Hay ID, Nelson RE, Jiang NS. Clinical utility of thyrotropin-receptor antibody assays: comparison of radioreceptor and bioassay methods. Mayo Clinic Proceedings. 1988; 63(7):707-17

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

47. Nishihara E, Amino N, Kudo T, Ito M, Fukata S, Nishikawa M et al. Comparison of thyroglobulin and thyroid peroxidase antibodies measured by five different kits in autoimmune thyroid diseases. Endocrine Journal. 2017; 64(10):955-961

48. Ochi Y, Inui T, Kouki T, Yamashiro K, Takasu N, Kajita Y et al. Clinical usefulness of TSAb assay with high polyethylene glycol concentrations. Hormone Research. 1999; 51(3):142-9

49. Ochi Y, Kajita Y, Inui T, Yamashiro K, Takasu N, Sato Y et al. Sensitive thyroid-stimulating antibody assay in whole serum containing five percent polyethylene glycol using porcine thyroid cells. Thyroid. 2000; 10(8):653-7

50. Paunkovic J, Paunkovic N. Does autoantibody-negative Graves' disease exist? A second evaluation of the clinical diagnosis. Hormone and Metabolic Research. 2006; 38(1):53-6

51. Paunkovic N, Paunkovic J. Diagnostic sensitivity of two radio receptor assays (TRAK Assay and TRAK Dyno Human) for detection of TSH receptor antibodies. Nuclear Medicine Review. 2003; 6(2):119-22

52. Paunkovic N, Paunkovic J. The diagnostic criteria of Graves' disease and especially the thyrotropin receptor antibody; our own experience. Hellenic Journal of Nuclear Medicine. 2007; 10(2):89-94

53. Pedersen IB, Knudsen N, Perrild H, Ovesen L, Laurberg P. TSH-receptor antibody measurement for differentiation of hyperthyroidism into Graves' disease and multinodular toxic goitre: a comparison of two competitive binding assays. Clinical Endocrinology. 2001; 55(3):381-90

54. Pedersen OM, Aardal NP, Larssen TB, Varhaug JE, Myking O, Vik-Mo H. The value of ultrasonography in predicting autoimmune thyroid disease. Thyroid. 2000; 10(3):251-9

55. Pishdad P, Pishdad GR, Tavanaa S, Pishdad R, Jalli R. Thyroid ultrasonography in differentiation between Graves' disease and Hashimoto's thyroiditis. Journal of Biomedical Physics & Engineering. 2017; 7(1):21-26

56. Rago T, Chiovato L, Grasso L, Pinchera A, Vitti P. Thyroid ultrasonography as a tool for detecting thyroid autoimmune diseases and predicting thyroid dsfunction in apparently healthy subjects. Journal of Endocrinological Investigation. 2001; 24(10):763-9

57. Rosario PW, Santos JB, Nunes NS, da Silva AL, Calsolari MR. Color flow Doppler sonography for the etiologic diagnosis of thyrotoxicosis. Hormone and Metabolic Research. 2014; 46(7):505-9

58. Sapin R, d'Herbomez M, Gasser F, Meyer L, Schlienger JL. Increased sensitivity of a new assay for anti-thyroglobulin antibody detection in patients with autoimmune thyroid disease. Clinical Biochemistry. 2003; 36(8):611-6

59. Schott M, Feldkamp J, Bathan C, Fritzen R, Scherbaum WA, Seissler J. Detecting TSH-receptor antibodies with the recombinant TBII assay: technical and clinical evaluation. Hormone and Metabolic Research. 2000; 32(10):429-35

60. Schott M, Hermsen D, Broecker-Preuss M, Casati M, Mas JC, Eckstein A et al. Clinical value of the first automated TSH receptor autoantibody assay for the diagnosis of Graves' disease (GD): an international multicentre trial. Clinical Endocrinology. 2009; 71(4):566-73

61. Sekulic V, Rajic M, Vlajkovic M, Ilic S, Bogicevic M, Antic S et al. Thyroid blood flow and uptake of technetium-99m pertechnetate in Graves' disease. Hellenic Journal of Nuclear Medicine. 2006; 9(3):173-6

62. Smith J, Brown RS. Persistence of thyrotropin (TSH) receptor antibodies in children and adolescents with Graves' disease treated using antithyroid medication. Thyroid. 2007; 17(11):1103-7

63. Southgate K, Creagh F, Teece M, Kingswood C, Rees Smith B. A receptor assay for the measurement of TSH receptor antibodies in unextracted serum. Clinical Endocrinology. 1984; 20(5):539-48

64. Stozek K, Bossowski A, Ziora K, Bossowska A, Mrugacz M, Noczynska A et al. Functional TSH receptor antibodies in children with autoimmune thyroid diseases. Autoimmunity. 2018; 51(2):62-68

65. Sulman C, Gosselin P, Cazin JL, Cappoen JP, May JP. Thyrotropin binding inhibitory immunoglobulin (TBII) and thyroid diseases using ROC analysis. Pathologie Biologie. 1990; 38(2):113-8

66. Syme NR, Toft AD, Stoddart M, Beckett GJ. Clinical performance of the Roche cobas e411 automated assay system for thyrotropin-receptor antibodies for the diagnosis of Graves' disease. Annals of Clinical Biochemistry. 2011; 48(Pt 5):471-3

67. Szabolcs I, Bernard W, Horster FA. Thyroid autoantibodies in hospitalized chronic geriatric patients: Prevalence, effects of age, nonthyroidal clinical state, and thyroid function. Journal of the American Geriatrics Society. 1995; 43(6):670-673

68. Takasu N, Kamijo K, Sato Y, Yoshimura H, Nagata A, Ochi Y. Sensitive thyroid-stimulating antibody assay with high concentrations of polyethylene glycol for the diagnosis of Graves' disease. Clinical and Experimental Pharmacology and Physiology. 2004; 31(5-6):314-9

69. Takasu N, Oshiro C, Akamine H, Komiya I, Nagata A, Sato Y et al. Thyroid-stimulating antibody and TSH-binding inhibitor immunoglobulin in 277 Graves' patients and in 686 normal subjects. Journal of Endocrinological Investigation. 1997; 20(8):452-61

70. Theodoraki A, Jones G, Parker J, Woolman E, Martin N, Perera S et al. Performance of a third-generation TSH-receptor antibody in a UK clinic. Clinical Endocrinology. 2011; 75(1):127-33

71. Tozzoli R, Bagnasco M, Giavarina D, Bizzaro N. TSH receptor autoantibody immunoassay in patients with Graves' disease: improvement of diagnostic accuracy over different generations of methods. Systematic review and meta-analysis. Autoimmunity Reviews. 2012; 12(2):107-13

72. Tozzoli R, Kodermaz G, Villalta D, Bagnasco M, Pesce G, Bizzaro N. Accuracy of receptor-based methods for detection of thyrotropin-receptor autoantibodies: a new automated third-generation immunoassay shows higher analytical and clinical sensitivity for the differential diagnosis of hyperthyroidism. Autoimmunity Highlights. 2010; 1(2):95-100

73. Uchida T, Suzuki R, Kasai T, Onose H, Komiya K, Goto H et al. Cutoff value of thyroid uptake of (99m)Tc-pertechnetate to discriminate between Graves' disease and painless thyroiditis: a single center retrospective study. Endocrine Journal. 2016; 63(2):143-9

74. Varadhan L, Varughese GI, Sankaranarayanan S. Hyperthyroidism and Graves' disease: Is an ultrasound examination needed? Indian Journal of Endocrinology and Metabolism. 2016; 20(6):866-869

75. Vos XG, Smit N, Endert E, Tijssen JG, Wiersinga WM. Frequency and characteristics of TBII-seronegative patients in a population with untreated Graves' hyperthyroidism: a prospective study. Clinical Endocrinology. 2008; 69(2):311-7

76. Wallaschofski H, Orda C, Georgi P, Miehle K, Paschke R. Distinction between autoimmune and non-autoimmune hyperthyroidism by determination of TSH-receptor antibodies in patients with the initial diagnosis of toxic multinodular goiter. Hormone and Metabolic Research. 2001; 33(8):504-507

77. Yoshimura Noh J, Miyazaki N, Ito K, Takeda K, Hiramatsu S, Morita S et al. Evaluation of a new rapid and fully automated electrochemiluminescence immunoassay for thyrotropin receptor autoantibodies. Thyroid. 2008; 18(11):1157-64

78. Zophel K, Gruning T, Roggenbuck D, Wunderlich G, Kotzerke J. On specificity of 2nd generation TSH receptor autoantibody measurements. Clinical Laboratory. 2008; 54(7-8):243-9

79. Zophel K, Roggenbuck D, von Landenberg P, Wunderlich G, Gruning T, Kotzerke J et al. TSH receptor antibody (TRAb) assays based on the human monoclonal autoantibody M22 are more sensitive than bovine TSH based assays. Hormone and Metabolic Research. 2010; 42(1):65-9

80. Zophel K, Roggenbuck D, Wunderlich G, Schott M. Continuously increasing sensitivity over three generations of TSH receptor autoantibody assays. Hormone and Metabolic Research. 2010; 42(12):900-2

81. Zouvanis M, Panz VR, Kalk WJ, Joffe BI. Thyrotropin receptor antibodies in black South African patients with Graves' disease and their response to medical therapy. Journal of Endocrinological Investigation. 1998; 21(11):771-4

82. Zuhur SS, Ozel A, Kuzu I, Erol RS, Ozcan ND, Basat O et al. The Diagnostic Utility of Color Doppler Ultrasonography, Tc-99m Pertechnetate Uptake, and TSH-Receptor Antibody for Differential Diagnosis of Graves' Disease and Silent Thyroiditis: A Comparative Study. Endocrine Practice. 2014; 20(4):310-9

Appendices

Appendix A: Review protocols

Table 6:

ID	Field	Content
I	Review question	What is the accuracy of anti-TPO testing, TRAb testing, ultrasound scanning and isotope scanning for diagnosing Graves' disease? What is the clinical and cost effectiveness of using anti-TPO testing, TRAb testing, ultrasound scanning or isotope scanning in the diagnosis of Graves' disease?
II	Type of review question	Diagnostic accuracy Test and treat review A review of health economic evidence related to the same review question was conducted in parallel with this review. For details see the health economic review protocol for this NICE guideline.
III	Objective of the review	To determine the accuracy and clinical and cost effectiveness of anti-TPO testing, TRAb testing, ultrasound scanning and isotope scanning for diagnosing Graves' disease. Appropriate treatment of hyperthyroidism requires determining whether Graves' disease is the underlying cause. Anti-TSH testing, ultrasound scanning and isotope scanning may all be used for this purpose. This review seeks to clarify the accuracy of each in order to inform recommendations about which should be used.
IV	Eligibility criteria – population / disease / condition / issue / domain	• People diagnosed with hyperthyroidism who are being investigated for Graves' disease
V	Eligibility criteria – intervention(s) / exposure(s) / prognostic factor(s)	• Anti-TPO testing • TRAb testing • Ultrasound scan • Isotope scan
VI	Eligibility criteria – comparator(s) / control or reference (gold) standard	**Diagnostic accuracy data:** • Reference standard to be determined by include studies, likely to include some composite of TRAb, multiple investigations, eventual clinical progression. To be specified in review on a study by study basis and impact on risk of bias considered **Test and treat data:** • Any of above testing strategies compared with any other
VII	Outcomes and prioritisation	**Diagnostic accuracy data:** • Sensitivity • Specificity Specificity will be prioritised

		Test and treat data:
		• Critical
		  ○ Mortality (dichotomous)
		  ○ Quality of life (continuous)
		• Important
		  ○ Healthcare contacts (rates/dichotomous)
		  ○ Experience of care (continuous)

VIII	Eligibility criteria – study design	**Test and treat data:** RCTs preferred, if no RCTs available to consider non-randomised cohort studies in which key confounders (age, sex, co-existing conditions) are addressed, either through restriction or appropriate matching/statistical adjustment **Diagnostic accuracy data:** Two gate study designs will be excluded Prospective studies prioritised, retrospective studies included if insufficient prospective studies identified Minimum duration of follow-up 3 months Crossover studies excluded
IX	Other inclusion exclusion criteria	Nil else
X	Proposed sensitivity / subgroup analysis, or meta-regression	**Stratifications** • Age – infants (<4), children (4-18), adults (>18-65), older adults (>65) • Generation of TRAb assays – 1st vs 2nd vs 3rd • US type – appearance only vs flow based assessment **Subgroup analyses** • Architecture of TRAb assays – presence of antibodies vs function of antibodies
XI	Selection process – duplicate screening / selection / analysis	• A sample of at least 10% of the abstract lists were double-sifted by a senior research fellow and discrepancies rectified, with committee input where consensus could not be reached, for more information please see the separate Methods report for this guideline.
XII	Data management (software)	• Endnote was used for bibliography, citations, sifting and reference management • WinBUGS was used for meta-analysis of diagnostic accuracy outcomes
XIII	Information sources – databases and dates	• Medline, Embase and the Cochrane library
XIV	Identify if an update	Not an update
XV	Author contacts	https://www.nice.org.uk/guidance/indevelopment/gid-ng10074
XVI	Highlight if amendment to previous protocol	Not an amendment
XVI	Search	For details please see appendix B

I	strategy – for one database	
XVIII	Data collection process – forms / duplicate	A standardised evidence table format will be used, and published as appendix D of the evidence report.
XIX	Data items – define all variables to be collected	For details please see evidence tables in Appendix D (clinical evidence tables) or H (health economic evidence tables).
XX	Methods for assessing bias at outcome / study level	QUADAS-2 checklists were used to critically appraise individual studies. For details please see section 6.2 of Developing NICE guidelines: the manual The risk of bias across all available evidence was evaluated for each outcome using an adaptation of the 'Grading of Recommendations Assessment, Development and Evaluation (GRADE) toolbox' developed by the international GRADE working group http://www.gradeworkinggroup.org/
XXI	Criteria for quantitative synthesis	For details please see section 6.4 of Developing NICE guidelines: the manual.
XXII	Methods for quantitative analysis – combining studies and exploring (in)consistency	For details please see the separate Methods report for this guideline.
XXIII	Meta-bias assessment – publication bias, selective reporting bias	For details please see section 6.2 of Developing NICE guidelines: the manual.
XXIV	Confidence in cumulative evidence	For details please see sections 6.4 and 9.1 of Developing NICE guidelines: the manual.
XXV	Rationale / context – what is known	For details please see the introduction to the evidence review.
XXVI	Describe contributions of authors and guarantor	A multidisciplinary committee developed the evidence review. The committee was convened by the National Guideline Centre (NGC) and chaired by Sarah Fishburn in line with section 3 of Developing NICE guidelines: the manual. Staff from NGC undertook systematic literature searches, appraised the evidence, conducted meta-analysis and cost-effectiveness analysis where appropriate, and drafted the evidence review in collaboration with the committee. For details please see Developing NICE guidelines: the manual.
XXVII	Sources of funding / support	NGC is funded by NICE and hosted by the Royal College of Physicians.
XXVIII	Name of sponsor	NGC is funded by NICE and hosted by the Royal College of Physicians.
XXIX	Roles of sponsor	NICE funds NGC to develop guidelines for those working in the NHS, public health and social care in England.
XXX	PROSPERO registration	Not registered

	number	

Table 7: Health economic review protocol

Review question	All questions – health economic evidence
Objectives	To identify health economic studies relevant to any of the review questions.
Search criteria	• Populations, interventions and comparators must be as specified in the clinical review protocol above. • Studies must be of a relevant health economic study design (cost–utility analysis, cost-effectiveness analysis, cost–benefit analysis, cost–consequences analysis, comparative cost analysis). • Studies must not be a letter, editorial or commentary, or a review of health economic evaluations. (Recent reviews will be ordered although not reviewed. The bibliographies will be checked for relevant studies, which will then be ordered.) • Unpublished reports will not be considered unless submitted as part of a call for evidence. • Studies must be in English.
Search strategy	A health economic study search will be undertaken using population-specific terms and a health economic study filter – see appendix B below.
Review strategy	Studies not meeting any of the search criteria above will be excluded. Studies published before 2003, abstract-only studies and studies from non-OECD countries or the USA will also be excluded. Each remaining study will be assessed for applicability and methodological limitations using the NICE economic evaluation checklist which can be found in appendix H of Developing NICE guidelines: the manual (2014).[46] **Inclusion and exclusion criteria** • If a study is rated as both 'Directly applicable' and with 'Minor limitations' then it will be included in the guideline. A health economic evidence table will be completed and it will be included in the health economic evidence profile. • If a study is rated as either 'Not applicable' or with 'Very serious limitations' then it will usually be excluded from the guideline. If it is excluded then a health economic evidence table will not be completed and it will not be included in the health economic evidence profile. • If a study is rated as 'Partially applicable', with 'Potentially serious limitations' or both then there is discretion over whether it should be included. **Where there is discretion** The health economist will make a decision based on the relative applicability and quality of the available evidence for that question, in discussion with the guideline committee if required. The ultimate aim is to include health economic studies that are helpful for decision-making in the context of the guideline and the current NHS setting. If several studies are considered of sufficiently high applicability and methodological quality that they could all be included, then the health economist, in discussion with the committee if required, may decide to include only the most applicable studies and to selectively exclude the remaining studies. All studies excluded on the basis of applicability or methodological limitations will be listed with explanation in the excluded health economic studies appendix below. The health economist will be guided by the following hierarchies. *Setting:* • UK NHS (most applicable). • OECD countries with predominantly public health insurance systems (for example, France, Germany, Sweden). • OECD countries with predominantly private health insurance systems (for example, Switzerland).

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

Appendix B: Literature search strategies

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

For more detailed information, please see the Methodology Review.

B.1 Clinical search literature search strategy

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

Table 8: Database date parameters and filters used

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

Medline (Ovid) search terms

1.	exp goiter/
2.	exp Hyperthyroidism/
3.	(hyperthyroid* or thyrotoxicosis).ti,ab.
4.	(toxic adj4 (node* or nodul* or multi?nodul* or goitre or goiter)).ti,ab.
5.	(graves' disease or plummer's disease).ti,ab.
6.	or/1-5
7.	letter/
8.	editorial/
9.	news/
10.	exp historical article/
11.	Anecdotes as Topic/
12.	comment/
13.	case report/
14.	(letter or comment*).ti.
15.	or/7-14
16.	randomized controlled trial/ or random*.ti,ab.
17.	15 not 16
18.	animals/ not humans/
19.	exp Animals, Laboratory/
20.	exp Animal Experimentation/
21.	exp Models, Animal/
22.	exp Rodentia/
23.	(rat or rats or mouse or mice).ti.
24.	or/17-23
25.	6 not 24
26.	limit 25 to English language
27.	autoantibodies/
28.	anti-TPO.ti,ab.
29.	((anti thyroid or antithyroid or TPO) adj2 (peroxidase or antibod* or autoantibod*)).ti,ab.
30.	((iodide adj2 peroxidase) or thyroperoxidase or microsomal antigen).ti,ab.
31.	TRAbs.ti,ab.
32.	((TSH or thyrotropin) adj2 receptor* adj2 (antigen* or antibod* or anti bod*)).ti,ab.
33.	(TSI or TBI or TBII or (thyroid adj2 (antibod* or anti bod*)) or binding inhibitory immunoglobulin).ti,ab.
34.	Ultrasonography/
35.	(ultrasonic or ultra sonic or ultra sonograh* or ultrasonograph* or ultrasound* or ultra sound* or sonograph* or sonogram* or echograph* or echotomograph* or doppler).ti,ab.
36.	(computed adj3 tomography).ti,ab.
37.	((isotope* or radioisotope* or radio isotope) adj4 scan*).ti,ab.
38.	radionuclide imaging/
39.	iodine radioisotopes/
40.	((iodine 131 or 131-I or I-131 or iodine 123 or 123-I or I-123 or radioiodine or radio-iodine or radionuclide) adj4 (scan* or test* or imag* or image*)).ti,ab.
41.	(radioactive iodine uptake or RAI or RAUI or RAIU).ti,ab.
42.	or/27-41

43.	randomized controlled trial.pt.
44.	controlled clinical trial.pt.
45.	randomi#ed.ti,ab.
46.	placebo.ab.
47.	randomly.ti,ab.
48.	Clinical Trials as topic.sh.
49.	trial.ti.
50.	or/43-49
51.	Meta-Analysis/
52.	exp Meta-Analysis as Topic/
53.	(meta analy* or metanaly* or metaanaly* or meta regression).ti,ab.
54.	((systematic* or evidence*) adj3 (review* or overview*)).ti,ab.
55.	(reference list* or bibliograph* or hand search* or manual search* or relevant journals).ab.
56.	(search strategy or search criteria or systematic search or study selection or data extraction).ab.
57.	(search* adj4 literature).ab.
58.	(medline or pubmed or cochrane or embase or psychlit or psyclit or psychinfo or psycinfo or cinahl or science citation index or bids or cancerlit).ab.
59.	cochrane.jw.
60.	((multiple treatment* or indirect or mixed) adj2 comparison*).ti,ab.
61.	or/51-60
62.	exp "sensitivity and specificity"/
63.	(sensitivity or specificity).ti,ab.
64.	((pre test or pretest or post test) adj probability).ti,ab.
65.	(predictive value* or PPV or NPV).ti,ab.
66.	likelihood ratio*.ti,ab.
67.	likelihood function/
68.	((area under adj4 curve) or AUC).ti,ab.
69.	(receive* operat* characteristic* or receive* operat* curve* or ROC curve*).ti,ab.
70.	(diagnos* adj3 (performance* or accurac* or utilit* or value* or efficien* or effectiveness)).ti,ab.
71.	gold standard.ab.
72.	or/62-71
73.	Epidemiologic studies/
74.	Observational study/
75.	exp Cohort studies/
76.	(cohort adj (study or studies or analys* or data)).ti,ab.
77.	((follow up or observational or uncontrolled or non randomi#ed or epidemiologic*) adj (study or studies or data)).ti,ab.
78.	((longitudinal or retrospective or prospective or cross sectional) and (study or studies or review or analys* or cohort* or data)).ti,ab.
79.	Controlled Before-After Studies/
80.	Historically Controlled Study/
81.	Interrupted Time Series Analysis/
82.	(before adj2 after adj2 (study or studies or data)).ti,ab.
83.	or/73-82

84.	exp case control study/
85.	case control*.ti,ab.
86.	or/84-85
87.	83 or 86
88.	Cross-sectional studies/
89.	(cross sectional and (study or studies or review or analys* or cohort* or data)).ti,ab.
90.	or/88-89
91.	83 or 90
92.	83 or 86 or 90
93.	26 and 42 and (50 or 61 or 72 or 92)

Embase (Ovid) search terms

1.	goiter/
2.	hyperthyroidism/ or graves disease/ or thyrotoxicosis/ or toxic goiter/
3.	(hyperthyroid* or thyrotoxicosis).ti,ab.
4.	(toxic adj4 (node* or nodul* or multi?nodul* or goitre or goiter)).ti,ab.
5.	(graves' disease or plummer's disease).ti,ab.
6.	or/1-5
7.	letter.pt. or letter/
8.	note.pt.
9.	editorial.pt.
10.	case report/ or case study/
11.	(letter or comment*).ti.
12.	or/7-11
13.	randomized controlled trial/ or random*.ti,ab.
14.	12 not 13
15.	animal/ not human/
16.	nonhuman/
17.	exp Animal Experiment/
18.	exp Experimental Animal/
19.	animal model/
20.	exp Rodent/
21.	(rat or rats or mouse or mice).ti.
22.	or/14-21
23.	6 not 22
24.	limit 23 to English language
25.	Autoantibodies/
26.	anti-TPO.ti,ab.
27.	((anti thyroid or antithyroid or TPO) adj2 (peroxidase or antibod* or autoantibod*)).ti,ab.
28.	((iodide adj2 peroxidase) or thyroperoxidase or microsomal antigen).ti,ab.
29.	TRAbs.ti,ab.
30.	((TSH or thyrotropin) adj2 receptor* adj2 (antigen* or antibod* or anti bod*)).ti,ab.
31.	(TSI or TBI or TBII or (thyroid adj2 (antibod* or anti bod*)) or binding inhibitory immunoglobulin).ti,ab.
32.	echography/
33.	(ultrasonic or ultra sonic or ultra sonograh* or ultrasonograph* or ultrasound* or ultra

	sound* or sonograph* or sonogram* or echograph* or echotomograph* or doppler).ti,ab.
34.	(computed adj3 tomography).ti,ab.
35.	((isotope* or radioisotope* or radio isotope) adj4 scan*).ti,ab.
36.	scintiscanning/
37.	radioactive iodine/
38.	((iodine 131 or 131-I or I-131 or iodine 123 or 123-I or I-123 or radioiodine or radio-iodine or radionuclide) adj4 (scan* or test* or imag* or image*)).ti,ab.
39.	(radioactive iodine uptake or RAI or RAUI or RAIU).ti,ab.
40.	or/27-39
41.	random*.ti,ab.
42.	factorial*.ti,ab.
43.	(crossover* or cross over*).ti,ab.
44.	((doubl* or singl*) adj blind*).ti,ab.
45.	(assign* or allocat* or volunteer* or placebo*).ti,ab.
46.	crossover procedure/
47.	single blind procedure/
48.	randomized controlled trial/
49.	double blind procedure/
50.	or/41-49
51.	systematic review/
52.	meta-analysis/
53.	(meta analy* or metanaly* or metaanaly* or meta regression).ti,ab.
54.	((systematic* or evidence*) adj3 (review* or overview*)).ti,ab.
55.	(reference list* or bibliograph* or hand search* or manual search* or relevant journals).ab.
56.	(search strategy or search criteria or systematic search or study selection or data extraction).ab.
57.	(search* adj4 literature).ab.
58.	(medline or pubmed or cochrane or embase or psychlit or psyclit or psychinfo or psycinfo or cinahl or science citation index or bids or cancerlit).ab.
59.	cochrane.jw.
60.	((multiple treatment* or indirect or mixed) adj2 comparison*).ti,ab.
61.	or/51-60
62.	exp "sensitivity and specificity"/
63.	(sensitivity or specificity).ti,ab.
64.	((pre test or pretest or post test) adj probability).ti,ab.
65.	(predictive value* or PPV or NPV).ti,ab.
66.	likelihood ratio*.ti,ab.
67.	((area under adj4 curve) or AUC).ti,ab.
68.	(receive* operat* characteristic* or receive* operat* curve* or ROC curve*).ti,ab.
69.	(diagnos* adj3 (performance* or accurac* or utilit* or value* or efficien* or effectiveness)).ti,ab.
70.	diagnostic accuracy/
71.	diagnostic test accuracy study/
72.	gold standard.ab.
73.	or/62-72

74.	Clinical study/
75.	Observational study/
76.	family study/
77.	longitudinal study/
78.	retrospective study/
79.	prospective study/
80.	cohort analysis/
81.	follow-up/
82.	cohort*.ti,ab.
83.	81 and 82
84.	(cohort adj (study or studies or analys* or data)).ti,ab.
85.	((follow up or observational or uncontrolled or non randomi#ed or epidemiologic*) adj (study or studies or data)).ti,ab.
86.	((longitudinal or retrospective or prospective or cross sectional) and (study or studies or review or analys* or cohort* or data)).ti,ab.
87.	(before adj2 after adj2 (study or studies or data)).ti,ab.
88.	or/74-80,83-87
89.	exp case control study/
90.	case control*.ti,ab.
91.	or/89-90
92.	88 or 91
93.	cross-sectional study/
94.	(cross sectional and (study or studies or review or analys* or cohort* or data)).ti,ab.
95.	or/93-94
96.	88 or 95
97.	88 or 91 or 95
98.	24 and 40 and (50 or 61 or 73 or 97)

Cochrane Library (Wiley) search terms

#1.	MeSH descriptor: [Goiter] explode all trees
#2.	MeSH descriptor: [Hyperthyroidism] explode all trees
#3.	(hyperthyroid* or thyrotoxicosis):ti,ab
#4.	(toxic near/4 (node* or nodul* or multinodul* or multi-nodul* or goitre or goiter)):ti,ab
#5.	MeSH descriptor: [Graves Disease] explode all trees
#6.	(grave* near/4 (thyrotoxicos* or hyperthyr*)):ti,ab
#7.	graves' disease:ti,ab
#8.	(or #1-#7)
#9.	MeSH descriptor: [Autoantibodies] explode all trees
#10.	anti-TPO:ti,ab
#11.	((anti thyroid or antithyroid or TPO) near/2 (peroxidase or antibod* or autoantibod*)):ti,ab
#12.	((iodide near/2 peroxidase) or thyroperoxidase or microsomal antigen):ti,ab
#13.	TRAbs:ti,ab
#14.	((TSH or thyrotropin) near/2 receptor* near/2 (antigen* or antibod* or anti bod*)):ti,ab
#15.	(TSI or TBI or TBII or (thyroid near/2 (antibod* or anti bod*)) or binding inhibitory immunoglobulin):ti,ab
#16.	MeSH descriptor: [Ultrasonography] explode all trees

#17.	(ultrasonic or ultra sonic or ultra sonograh* or ultrasonograph* or ultrasound* or ultra sound* or sonograph* or sonogram* or echograph* or echotomograph* or doppler):ti,ab
#18.	(computed near/3 tomography):ti,ab
#19.	((isotope* or radioisotope* or radio isotope) near/4 scan*):ti,ab
#20.	MeSH descriptor: [Radionuclide imaging] explode all trees
#21.	MeSH descriptor: [Iodine radioisotopes] explode all trees
#22.	((iodine 131 or 131-I or I-131 or iodine 123 or 123-I or I-123 or radioiodine or radio-iodine or radionuclide) near/4 (scan* or test* or imag* or image*)):ti,ab
#23.	(radioactive iodine uptake or RAI or RAUI or RAIU):ti,ab
#24.	(or #9-#23)
#25.	#8 and #24

B.2 Health Economics literature search strategy

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

Table 9: Database date parameters and filters used

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

Medline (Ovid) search terms

1.	exp thyroid diseases/
2.	hyperthyroid*.ti,ab.
3.	hypothyroid*.ti,ab.
4.	thyrotoxicosis.ti,ab.
5.	(thyroid adj3 (swell* or dysfunction* or enlarg* or nodule* or node* or disease* or condition* or disorder*)).ti,ab.
6.	or/1-5
7.	letter/
8.	editorial/
9.	news/

10.	exp historical article/
11.	Anecdotes as Topic/
12.	comment/
13.	case report/
14.	(letter or comment*).ti.
15.	or/7-14
16.	randomized controlled trial/ or random*.ti,ab.
17.	15 not 16
18.	animals/ not humans/
19.	exp Animals, Laboratory/
20.	exp Animal Experimentation/
21.	exp Models, Animal/
22.	exp Rodentia/
23.	(rat or rats or mouse or mice).ti.
24.	or/17-23
25.	6 not 24
26.	limit 25 to English language
27.	Economics/
28.	Value of life/
29.	exp "Costs and Cost Analysis"/
30.	exp Economics, Hospital/
31.	exp Economics, Medical/
32.	Economics, Nursing/
33.	Economics, Pharmaceutical/
34.	exp "Fees and Charges"/
35.	exp Budgets/
36.	budget*.ti,ab.
37.	cost*.ti.
38.	(economic* or pharmaco?economic*).ti.
39.	(price* or pricing*).ti,ab.
40.	(cost* adj2 (effective* or utilit* or benefit* or minimi* or unit* or estimat* or variable*)).ab.
41.	(financ* or fee or fees).ti,ab.
42.	(value adj2 (money or monetary)).ti,ab.
43.	or/27-42
44.	exp models, economic/
45.	*Models, Theoretical/
46.	*Models, Organizational/
47.	markov chains/
48.	monte carlo method/
49.	exp Decision Theory/
50.	(markov* or monte carlo).ti,ab.
51.	econom* model*.ti,ab.
52.	(decision* adj2 (tree* or analy* or model*)).ti,ab.
53.	or/44-52

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

Embase (Ovid) search terms

1.	exp thyroid diseases/
2.	hyperthyroid*.ti,ab.
3.	hypothyroid*.ti,ab.
4.	thyrotoxicosis*.ti,ab.
5.	(thyroid adj3 (swell* or dysfunction* or enlarg* or nodule* or node* or disease* or condition* or disorder*)).ti,ab.
6.	or/1-5
7.	letter.pt. or letter/
8.	note.pt.
9.	editorial.pt.
10.	case report/ or case study/
11.	(letter or comment*).ti.
12.	or/7-11
13.	randomized controlled trial/ or random*.ti,ab.
14.	12 not 13
15.	animal/ not human/
16.	nonhuman/
17.	exp Animal Experiment/
18.	exp Experimental Animal/
19.	animal model/
20.	exp Rodent/

21.	(rat or rats or mouse or mice).ti.
22.	or/14-21
23.	6 not 22
24.	limit 23 to English language
25.	health economics/
26.	exp economic evaluation/
27.	exp health care cost/
28.	exp fee/
29.	budget/
30.	funding/
31.	budget*.ti,ab.
32.	cost*.ti.
33.	(economic* or pharmaco?economic*).ti.
34.	(price* or pricing*).ti,ab.
35.	(cost* adj2 (effective* or utilit* or benefit* or minimi* or unit* or estimat* or variable*)).ab.
36.	(financ* or fee or fees).ti,ab.
37.	(value adj2 (money or monetary)).ti,ab.
38.	or/25-37
39.	statistical model/
40.	exp economic aspect/
41.	39 and 40
42.	*theoretical model/
43.	*nonbiological model/
44.	stochastic model/
45.	decision theory/
46.	decision tree/
47.	monte carlo method/
48.	(markov* or monte carlo).ti,ab.
49.	econom* model*.ti,ab.
50.	(decision* adj2 (tree* or analy* or model*)).ti,ab.
51.	or/41-50
52.	quality adjusted life year/
53.	"quality of life index"/
54.	short form 12/ or short form 20/ or short form 36/ or short form 8/
55.	sickness impact profile/
56.	(quality adj2 (wellbeing or well being)).ti,ab.
57.	sickness impact profile.ti,ab.
58.	disability adjusted life.ti,ab.
59.	(qal* or qtime* or qwb* or daly*).ti,ab.
60.	(euroqol* or eq5d* or eq 5*).ti,ab.

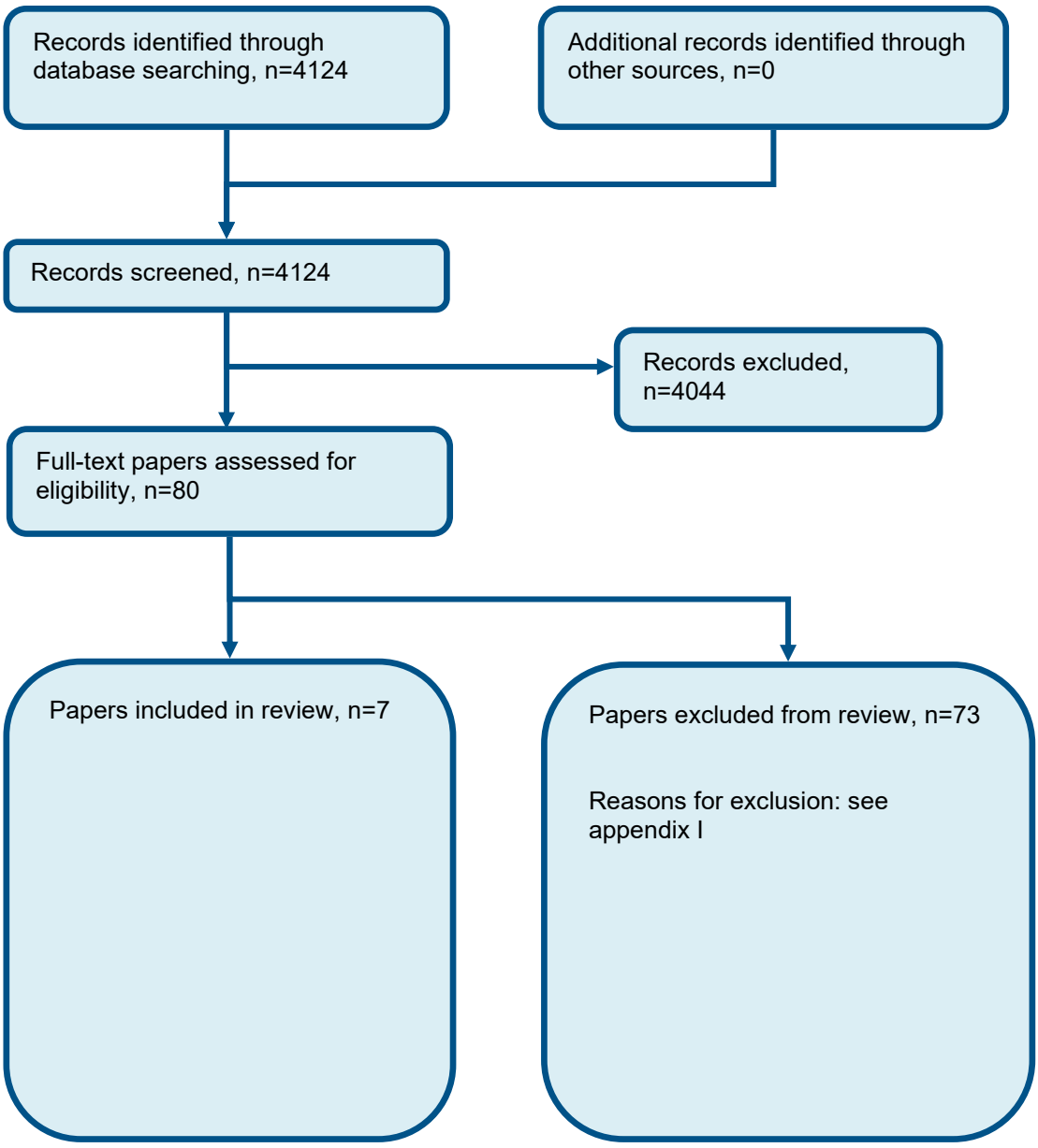
61.	(qol* or hql* or hqol* or h qol* or hrqol* or hr qol*).ti,ab.
62.	(health utility* or utility score* or disutilit* or utility value*).ti,ab.
63.	(hui or hui1 or hui2 or hui3).ti,ab.
64.	(health* year* equivalent* or hye or hyes).ti,ab.
65.	discrete choice*.ti,ab.
66.	rosser.ti,ab.
67.	(willingness to pay or time tradeoff or time trade off or tto or standard gamble*).ti,ab.
68.	(sf36* or sf 36* or short form 36* or shortform 36* or shortform36*).ti,ab.
69.	(sf20 or sf 20 or short form 20 or shortform 20 or shortform20).ti,ab.
70.	(sf12* or sf 12* or short form 12* or shortform 12* or shortform12*).ti,ab.
71.	(sf8* or sf 8* or short form 8* or shortform 8* or shortform8*).ti,ab.
72.	(sf6* or sf 6* or short form 6* or shortform 6* or shortform6*).ti,ab.
73.	or/52-72
74.	24 and (38 or 51 or 73)

NHS EED and HTA (CRD) search terms

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

Appendix C: Clinical evidence selection

Figure 1: Flow chart of clinical study selection for the review of antibodies for hyperthyroidism



Appendix D: Clinical evidence tables

Reference	**Baskaran 2015** [5]
Study type	Retrospective
Study methodology	Data source: patients presenting at paediatric endocrine unit between January 2002 and January 2014 Recruitment: not specified
Number of patients	n = 47 (37 GD; 10 non-GD thyroiditis)
Patient characteristics	Age, mean (SD): 12.3 (4.6); GD 11.7 (4.4); non-GD 14.8 (4.5) Gender (male to female ratio): 39:8 Ethnicity: not specified Setting: Massachusetts General Hospital for Children; Mayo Medical Laboratories; Nuclear Medicine Unit of Massachusetts General Hospital (MGH) Country: USA Inclusion criteria: patients presenting at the paediatric endocrine unit between January 2002 and January 2014 with symptoms of hyperthyroidism and suppressed TSH level associated with an elevated total triiodothyronine (T3) and/or elevated free thyroxine (T4), with both TSI levels and ^{99m}Tc scan at the time of diagnosis. Exclusion criteria: diagnoses such as thyroid nodules or thyroid malignancy
Target condition(s)	Graves' disease
Index test(s) and reference standard	Index test: ^{99m}Tc scan ^{99m}Tc uptake was performed in the MGH. ^{99m}Technetium pertechnetate was given as an intravenous injection, and the dose was calculated based on the patient's weight (0.15 mCi/kg). Standard and pinhole images were obtained 20 minutes after the intravenous injection, and an uptake was calculated. The lower limit of normal ^{99m}Tc uptake was based on the normal reference range described by the nuclear medicine department of MGH. Which is 0.5-3.75%. Therefore uptake ≤ 0.4% was considered to be decreased/negative and suggestive of destructive/ non-GD thyroiditis. Any uptake that was either increased or inappropriately normal was considered a positive

<table>
<tr><td>Reference</td><td colspan="5">Baskaran 2015 [5]</td></tr>
<tr><td></td><td colspan="5">result and suggestive of GD.

<u>Index test: TSI</u>
TSI assessment was performed at the Mayo Medical Laboratories by comparing cyclic AMP activity in TSH responsive cell lines after addition of patient’s serum with exposure to normal control serum. The test was performed using Diagnostic Hybrids kits with a coefficient of variation of <15%. The clinical sensitivity and specificity for the test was determined to be 92% and 99.4% respectively. In 11/37 patients with GD, the TSI was sent out to other clinical laboratories. Therefore, analysis was performed with the TSI value represented as multiples of the upper limit of normal for the respective labs. The test was considered positive if the TSI index was above the upper limit of normal for the lab.

<u>Reference standard: Clinical presentation & successful treatment for GD at follow-up</u>
Diagnosis of the cause of hyperthyroidism was established based on laboratory tests and clinical progress. Laboratory tests included levels of TSH, free T4, total T3, thyroid peroxidase and thyroglobulin antibodies. GD was diagnosed by clinical presentation (including signs and symptoms at diagnosis and physical exam findings such as thyroid enlargement) and successful treatment with antithyroid medication, surgery or radioactive ablation at follow-up.

Time between measurement of index test and reference standard: not specified</td></tr>
<tr><td>2×2 table</td><td>^{99m}Tc scan</td><td>Reference standard +</td><td>Reference standard −</td><td>Total</td><td rowspan="4">In 3/47 patients, the absolute value of uptake was not quantified, but the report indicated symmetrically increased uptake in both lobes and the results were considered positive.</td></tr>
<tr><td></td><td>Index test +</td><td>37</td><td>0</td><td>37</td></tr>
<tr><td></td><td>Index test −</td><td>0</td><td>10</td><td>10</td></tr>
<tr><td></td><td>Total</td><td>37</td><td>10</td><td>47</td></tr>
<tr><td>2×2 table</td><td>TSI</td><td>Reference standard +</td><td>Reference standard −</td><td>Total</td><td rowspan="4"></td></tr>
<tr><td></td><td>Index test +</td><td>31</td><td>0</td><td>31</td></tr>
<tr><td></td><td>Index test −</td><td>6</td><td>10</td><td>16</td></tr>
<tr><td></td><td>Total</td><td>37</td><td>10</td><td>47</td></tr>
<tr><td>Statistical measures</td><td colspan="5"><u>Index text : ^{99m}Tc scan</u>
Sensitivity : 100%
Specificity: 100%
PPV: 100%
NPV: 100%

<u>Index text: TSI</u>
Sensitivity : 83.8%</td></tr>
</table>

Reference	**Baskaran 2015** [5]
	Specificity: 100% PPV: 100% NPV: 62.5%
Source of funding	NIH grants
Limitations	Risk of bias: high due to risk of bias in the index test and reference standard Indirectness: none
Comments	Diagnostic accuracy of ^{99m}Tc scan and TSI for Graves' disease in children

Reference	**Lee 2016** [36]
Study type	Retrospective
Study methodology	Data source: patients <20 years of age who had undergone US between April 2008 and October 2013 Recruitment: unclear
Number of patients	n = 113 (132 US scans)
Patient characteristics	Age, mean (range): 12 years (6-19 years) Gender (male to female ratio): 23:90 Ethnicity: not specified Setting: St Mary's Hospital, The Catholic University of Korea Country: South Korea Inclusion criteria: patients <20 years of age with a diffuse goitre by inspection and palpation who had undergone thyroid US between April 2008 and October 2013 Exclusion criteria: patients with palpable thyroid nodules

Reference	**Lee 2016** [36]				
	Patients with diffuse swelling of the anterior neck or an enlarged thyroid gland by ocular inspection or palpation finally included (n=86: autoimmune thyroiditis n=26; Graves' disease n=14; simple goiter n=46); 12 out of 14 children with Graves' disease had overt hyperthyroidism (euthyroidism n=1, subclinical hyperthyroidism =1)				
Target condition(s)	Graves' disease				
Index test(s) and reference standard	Index test: US US was performed by a thyroid imaging specialist using a high-resolution US unit with a 7-to 12 MHz linear array transducer; at a single institution by two board-certified radiologists. All US images were reviewed retrospectively by two board-certified radiologists, one with 19 years' experience in paediatric imaging, the other with 10 years' experience in thyroid imaging and intervention. Doppler US was used to evaluate the vascularity of the glands and nodules. Increased vascularity was assessed subjectively during the examination and diagnosed by consensus. 'Thyroid inferno' was defined as increased vascularity. Reference standard: Clinical features The diagnosis of autoimmune thyroid disease was rendered based on the results of a radioimmunoassay of antithyroid antibody levels, including antithyroid peroxidase, antithyroglobulin and anti-thyroid-stimulating hormone receptor antibodies. Time between measurement of index test and reference standard: not specified				
2×2 table	**Hypoechogenicity**	Reference standard +	Reference standard −	Total	
	Index test +	12	24	36	
	Index test −	2	48	50	
	Total	14	72	86	
2×2 table	**Coarse echotexture**	Reference standard +	Reference standard −	Total	
	Index test +	9	19	28	
	Index test −	5	53	58	
	Total	14	72	86	
2×2 table	**Micronodularity**	Reference standard +	Reference standard −	Total	
	Index test +	1	14	15	
	Index test −	13	58	71	

Reference	**Lee 2016** [36]				
	Total	14	72	86	
2×2 table	**Increased vascularity**	Reference standard +	Reference standard −	Total	
	Index test +	10	6	16	
	Index test −	4	66	70	
	Total	14	72	86	
Statistical measures	Index text US (hypoechogenicity): Sensitivity : 85.7% Specificity: 66.7% PPV: 33.3% NPV: 96% Index text US (coarse echotexture): Sensitivity: 64.3% Specificity: 73.6% PPV: 32.1% NPV:91.4% Index test US (micronodularity): Sensitivity: 7.1% Specificity: 80.6% PPV: 6.7 NPV:81.7% Index text US (Increased vascularity): Sensitivity: 71.4% Specificity: 91.7% PPV: 62.5% NPV: 94.3%				
Source of funding	Not specified				
Limitations	Risk of bias: none Indirectness: none				
Comments	Diagnostic accuracy of US for Graves' disease in children				

Reference	**Paunkovic 2006[50]**
Study type	Retrospective test accuracy study
Study methodology	Data source: patients presenting with symptoms of hyperthyroidism between 1998 and 2000. Recruitment: consecutive
Number of patients	n = 255
Patient characteristics	Age, median (range): 52 (6-84) Gender (male to female ratio): 33:222 Ethnicity: not specified Setting: Medical centre, Department of Nuclear Medicine, Serbia Country: Serbia Inclusion criteria: patients presenting with symptoms of hyperthyroidism at the medical centre between 1998 and 2000 Exclusion criteria: low thyroid uptake on thyroid uptake test (^{131}I or ^{99m}Tc) 164 patients had newly manifested disease, 91 had relapse of known hyperthyroidism
Target condition(s)	Graves' disease
Index test(s) and reference standard	Index test: TRAb Conventional porcine TBII assay (TRAK assay) and second-generation TBII assay (TRAK human RRA) were performed according to the manufacturer's instructions. For TBII porcine assay used 15U/L as cut-off, for TBIII assay used 1.5IU/L as cut-off Reference standard: Clinical and biochemical criteria The same endocrinologist with experience in thyroidology for over 20 years established a diagnosis of Graves' disease in 255 consecutive patients using clinical and biochemical criteria. Presence of ophthalmopathy confirmed the immunological pathogenesis of hyperthyroidism, but the absence of ophthalmopathy did not exclude it.

Reference	**Paunkovic 2006[50]**			
	Time between measurement of index test and reference standard: not specified			
2×2 table		Reference standard +	Reference standard −	Total
	Index test +	231	0	231
	Index test −	3	21	24
	Total	234	21	255
Statistical measures	Index text: TRAb Sensitivity: 99% Specificity: 100% PPV: 100% NPV: 87.5%			
Source of funding	Not specified			
Limitations	Risk of bias: serious due to risk of bias in index test interpretation, flow and timing Indirectness: none			
Comments	Diagnostic accuracy of combined TBII and TBIII in adults			

Reference	**Pishdad 2017 [55]**
Study type	Test assessment study (prospective)
Study methodology	Data source: patients with definitive diagnosis of Graves' disease or Hashimoto's thyroiditis referred for sonographic examination Recruitment: not specified
Number of patients	n = 149
Patient characteristics	Age, mean (SD): Graves' disease 36.8 (10.17); Hashimoto's thyroiditis 33.4 (12.16); healthy controls 34.74 (16.87) Gender (male to female ratio): 32:117 Ethnicity: not specified

Reference	**Pishdad 2017** [55]				
	Setting: Shiraz University of Medical Sciences Country: Iran Inclusion criteria: not specified Exclusion criteria: uncertain diagnosis of Graves' disease or Hashimoto's thyroiditis, history of thyroid surgery, palpable nodules 86 patients were anti-TPO positive, 77 had higher than normal anti Tg levels.				
Target condition(s)	Graves' disease				
Index test(s) and reference standard	Index test Ultrasound: Ultrasound was performed by a single radiologist using MEDISON Accuvix V10 sonography unit with a 10 MHz linear transducer. Thyroid gland echogenicity was compared with patient's submandibular glands and the gain of sonographic system was set to produce an echo free appearance in the lumen of internal jugular vein and carotid artery. Reference standard: Clinical and lab data Laboratory data included measurements of thyroid hormone levels and anti-thyroid antibodies (anti-thyroid peroxidase, anti-thyroglobulin) Time between measurement of index test and reference standard: not specified				
2×2 table	**Homogenously hypoechoic**	Reference standard +	Reference standard −	Total	GD vs control group
	Index test +	16	5	21	
	Index test −	18	48	66	
	Total	34	53	87	
2×2 table	**Peripherally hypoechoic**	Reference standard +	Reference standard −	Total	GD vs control group
	Index test +	5	0	5	
	Index test −	29	53	82	
	Total	34	53	87	
2×2 table	**Centrally hypoechoic**	Reference standard +	Reference standard −	Total	GD vs control group

Reference	**Pishdad 2017** [55]				
	Index test +	6	0	6	
	Index test −	28	53	81	
	Total	34	53	87	
2×2 table	**Homogenously isoechoic**	Reference standard +	Reference standard −	Total	GD vs control group
	Index test +	2	26	28	
	Index test −	32	27	59	
	Total	34	53	87	
2×2 table	**Homogenously hyperechoic**	Reference standard +	Reference standard −	Total	GD vs control group
	Index test +	5	22	27	
	Index test −	29	31	60	
	Total	34	53	87	
Statistical measures	Index text US (homogenously hypoechoic) Sensitivity : 47.1% Specificity: 90.6% PPV: 76.2% NPV: 72.7% Index text US (peripherally hypoechoic) Sensitivity: 14.7% Specificity: 100% PPV: 100% NPV: 64.6% Index text US (centrally hypoechoic) Sensitivity: 17.6% Specificity: 100% PPV: 100% NPV: 65.4% Index text US (homogenously isoechoic) Sensitivity: 5.9% Specificity: 50.9%				

Reference	Pishdad 2017 [55]
	PPV: 7.1% NPV:45.8% Index test US (homogenously hyperechoic) Sensitivity: 14.7% Specificity:58.5% PPV:18.5% NPV:51.7%
Source of funding	Not specified
Limitations	Risk of bias: serious due to high risk of bias in patient selection Indirectness: none
Comments	Diagnostic accuracy of US for Graves' disease in Adults

Reference	Sulman 1990[65]
Study type	Prospective
Study methodology	Data source: patients clinically examined for hyper and hypo-metabolism symptoms, assessment of possible goiter and signs of any ocular and/or Graves' disease dermopathy. Recruitment: not specified
Number of patients	n = 190
Patient characteristics	Age, mean (SD): not specified Gender (male to female ratio): not specified Ethnicity: not specified Setting: not specified Country: France Inclusion criteria: pre-treatment patients clinically examined for hyper and hypo-metabolism symptoms, assessment of possible goiter and signs of any ocular and/or Graves' disease dermopathy

Reference	**Sulman 1990[65]**
	Exclusion criteria: not specified Based on anamnesis, clinical examination and biological analysis, of 128 auto-immune hyperthyroidisms 74 were Graves' disease (associating thyrotoxicosis, a diffuse goiter, ocular signs and/or a pretibial myxoedema type dermopathy), 54 were toxic diffuse goiters (which presented the same clinical picture as Graves' disease except for the ocular signs and dermopathy); of 35 patients with a priori non-immune hyperthyroidism, one had post-partum transitory hyperthyroidism, 8 secondary toxic goiters, 20 toxic nodules, 5 iodine-induced hyperthyroidisms and one chronic carcinoma. The other thyroid diseases included 6 hypothyroidisms of protothyroid source with elevated TSH, 13 thyroiditis (12 chronic Hashimoto's disease and one sub-acute Quervain's disease), 11 ordinary goiters and 3 isolated thyroid nodules.
Target condition(s)	Graves' disease
Index test(s) and reference standard	Index test: TRAb Detection of anti-TSH receptor antibodies (TBII) was performed using the radioreceptor assay Trak-assay of Behring Laboratories. The principle of this method is based on in vitro competition which uses the specific antibodies ability to inhibit labelled TSH binding to the TSH membrane receptor. TSH receptors used during this assay came from a detergent solubilisation of thyroid pig membranes. TBII, cut-off of 9%, derived from their own ROC curve, not clear what the % refers to Reference standard: Clinical examination and biological analysis All patients were clinically examined for hyper and hypo-metabolism symptoms, assessment of a possible goiter and signs of any ocular and/or Graves' disease dermopathy. Sera from all patients were assayed for thyroid hormones (T4 or FT3 and FT4) and thyrotropin (TSH or ultrasensitive TSH). In some patients, a study of the iodine uptake by the thyroid was performed. Time between measurement of index test and reference standard: not specified
2×2 table	

	Reference standard +	Reference standard −	Total
Index test +	112	9	121
Index test −	16	53	69
Total	128	62	190

Statistical measures	Index text: TRAb Sensitivity: 88% Specificity: 85% PPV: 92.6% NPV: 76.8%

Reference	**Sulman 1990[65]**
Source of funding	Not specified
Limitations	Risk of bias: serious due to patient selection, interpretation of index and reference standard Indirectness: none
Comments	Diagnostic accuracy of TB II in Adults

Reference	**Syme 2011[66]**
Study type	Prospective
Study methodology	Data source: new patients attending first appointment at thyroid clinic (Royal Infirmary of Edinburgh) between June 2008 and August 2009 Recruitment: consecutive
Number of patients	n = 102
Patient characteristics	Age, mean (SD): not specified Gender (male to female ratio): not specified Ethnicity: not specified Setting: Royal Infirmary of Edinburgh Country: UK Inclusion criteria: consecutive patients attending their first appointment at thyroid clinic between June 2008 and August 2009 Exclusion criteria: not specified Based on initial thyroid function test results, 58 of the 102 patients included had overt hyperthyroidism, seven had subclinical hyperthyroidism, one had hypothyroidism, five had subclinical hypothyroidism and 31 patients were euthyroid. 53 of the patients with overt hyperthyroidism were diagnosed with Graves' disease; the remaining five had diagnoses of autonomous nodule, postpartum thyroiditis, silent thyroiditis, type 2 amiodarone-induced thyroiditis or viral thyroiditis. Three of the patients with subclinical hyperthyroidism were diagnosed with Graves' disease; two of these received an isotope uptake scan and all three had TRAbs detected in their serum samples. The remaining four patients with subclinical hyperthyroidism all received isotope uptake scans; three were diagnosed with multi-nodular

Reference	**Syme 2011[66]**				
	goitre and one with toxic nodule.				
Target condition(s)	Graves' disease				
Index test(s) and reference standard	Index test: TRAb 3rd generation assay, TRAbs were measured using the cobas e411 analyser (Roche Diagnostics, Sussex, UK). The sensitivity, specificity, and positive and negative predictive values for the TRAbs assay in the diagnosis of Graves' disease were compared with published performance characteristics at cut-offs of 1.6, 1.75 and 1.86 IU/L, and also the manufacturer's stated functional sensitivity (0.9 IU/L). Reference standard: Clinical examination and t-99 (n=70) The diagnosis was made by the same consultant, independently of TRAb results, based on clinical examination with TSH, FT4, and total triiodothyronine concentrations measured on Architect analyser. 70 patients received a technetium-99 uptake scan to aid diagnosis. Time between measurement of index test and reference standard:				
2×2 table	TRAb (0.9 IU/L)	Reference standard +	Reference standard −	Total	
	Index test +	56	5	62	
	Index test −	0	41	41	
	Total	56	46	102	
2×2 table	TRAb (1.6 IU/L)	Reference standard +	Reference standard −	Total	
	Index test +	53	1	54	
	Index test −	3	45	48	
	Total	56	46	102	
2×2 table	TRAb (1.75 IU/L)	Reference standard +	Reference standard −	Total	
	Index test +	52	0	52	
	Index test −	4	46	50	
	Total	56	46	102	
2×2 table	TRAb (1.86 IU/L)	Reference standard +	Reference standard −	Total	
	Index test +	51	0	51	
	Index test −	5	46	51	
	Total	56	46	102	

Reference	Syme 2011[66]
Statistical measures	Index text TRAb (0.9 IU/L) Sensitivity: 100% Specificity: 89% PPV: 92% NPV: 100% Index text TRAb (1.6 IU/L) Sensitivity: 95% Specificity: 98% PPV: 98% NPV: 94% Index text TRAb (1.75 IU/L) Sensitivity: 93% Specificity: 100% PPV: 100% NPV: 92% Index text TRAb (1.86 IU/L) Sensitivity: 91% Specificity: 100% PPV: 100% NPV: 90%
Source of funding	NHS Research Scotland (NRS)
Limitations	Risk of bias: serious risk of bias dues to index test, flow and timing Indirectness:
Comments	Diagnostic accuracy of TRAb using different cut-offs

Reference	Theodoraki 2011[70]
Study type	Prospective & retrospective cohort
Study methodology	Data source: medical records of patients with TRAb requests between May 2008 and July 2009 (only hyperthyroid patients with indeterminate clinical diagnosis, with Graves' eye disease and pregnant women with past or present Graves' disease); hospital and primary care records of patients with newly recorded undetectable serum TSH from all sources identified at the Biochemistry laboratory

Reference	Theodoraki 2011[70]
	Recruitment: consecutive
Number of patients	n = 244
Patient characteristics	Age, mean (range): 45.8 (11-97) Gender (male to female ratio): 46:198 Ethnicity: not specified Setting: Department of Endocrinology, Clinical Immunology and Clinical Biochemistry, Royal Free Hampstead NHS Trust Country: UK Inclusion criteria: hospital medical records of patients with TRAb requests at the Department of Clinical Immunology between May 2008 and July 2009 (only hyperthyroid patients with indeterminate clinical diagnosis, with Graves' eye disease and pregnant women with past or present Graves' disease are tested for thyroid antibodies at the centre); samples of patients with newly identified undetectable serum TSH (<0.02 mIU/l) Exclusion criteria: patients with inadequate clinical information or duplicate requests; patients with known hyperthyroidism (for the prospective recruitment)
Target condition(s)	Graves' disease
Index test(s) and reference standard	Index test: TRAb (TBII) The TRAb assay used was a commercial third-generation TSH receptor autoantibody enzyme-linked immunosorbent assay (ELISA) kit supplied by RSR Limited. It quantified the presence of TRAb in patients' sera based on the inhibition of binding of the biotin labelled human monoclonal antibody M22 with immobilized TSH receptors in ELISA plates. Streptavidin peroxidase and tetramethylbenzidine were added to determine the amount of M22 bound to the plate. The absorbance of the mixture at 450 nm was read using an ELISA plate reader. Cut off 0.4U/L (manufacturer's suggested cut-off) Reference standard: Final recorded clinical diagnosis Four consultants and two trainees in Endocrinology participated in general endocrine, thyroid and antenatal outpatient clinics. Patients

<table>
<tr><td>Reference</td><td colspan="5">Theodoraki 2011[70]</td></tr>
<tr><td></td><td colspan="5">with suspected thyroid disease were tested for serum TSH and FT4. FT3 was measured when TSH was below reference range (0.3-4.2 mIU/l) and FT4 was normal. Hyperthyroid patients with clinical features of Graves' (diffusely enlarged thyroid, dysthyroid eye disease) or with previous history of Graves' disease were diagnosed with Graves' disease. Hyperthyroid patients with clinical diagnosis of indeterminate aetiology, nodular goitre, clinically or suspected thyroiditis, diagnostic thyroid scintigraphy with Tc-99m Pertechnetate was performed.
For the retrospective sample, hospital medical records were reviewed twice by independent reviewers and the final diagnosis was recorded.

Time between measurement of index test and reference standard: not specified</td></tr>
<tr><td>2×2 table</td><td></td><td>Reference standard +</td><td>Reference standard −</td><td>Total</td><td></td></tr>
<tr><td></td><td>Index test +</td><td>125</td><td>6</td><td>131</td><td></td></tr>
<tr><td></td><td>Index test −</td><td>20</td><td>93</td><td>113</td><td></td></tr>
<tr><td></td><td>Total</td><td>145</td><td>99</td><td>244</td><td></td></tr>
<tr><td>Statistical measures</td><td colspan="5">Index text: TRAb (TBIII)
Sensitivity: 86.2%
Specificity 93.9%
PPV: 95.4%
NPV: 82.3%</td></tr>
<tr><td>Source of funding</td><td colspan="5">Not stated</td></tr>
<tr><td>Limitations</td><td colspan="5">Risk of bias: serious due to risk of bias in interpretation of the index test.
Indirectness: none</td></tr>
<tr><td>Comments</td><td colspan="5">Diagnostic accuracy of third generation TRAb (TBIII) in adults.</td></tr>
</table>

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

E.1 Coupled sensitivity and specificity forest plots

Figure 2: TRAb, TB II/III, in adults

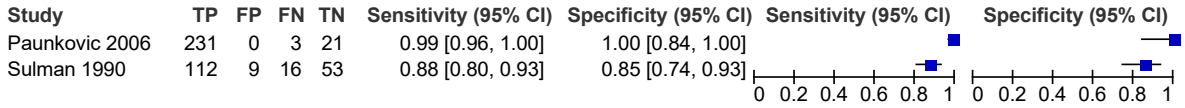


Figure 3: TRAb, TB III only, 0.4U/L, in adults

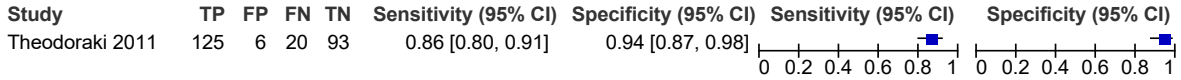


Figure 4: TRAb, TB III only, 0.9IU/L, in adults

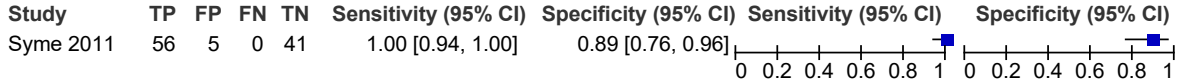


Figure 5: TRAb, TB III only, 1.6IU/L, in adults

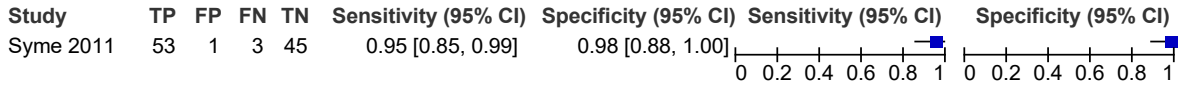


Figure 6: TRAb, TB III only, 1.75IU/L, in adults

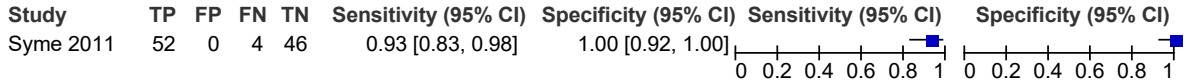


Figure 7: TRAb, TB III only, 1.75IU/L, in adults

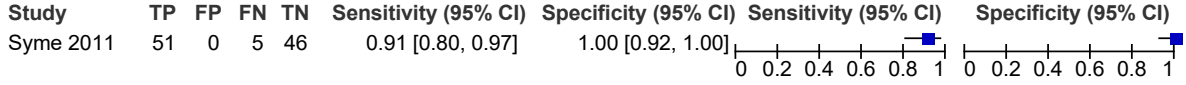


Figure 8: US, peripherally hypoechoic, in adults

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Pishdad 2017	5	0	29	53	0.15 [0.05, 0.31]	1.00 [0.93, 1.00]

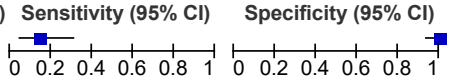


Figure 9: US, centrally hypoechoic, in adults

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Pishdad 2017	6	0	28	53	0.18 [0.07, 0.35]	1.00 [0.93, 1.00]

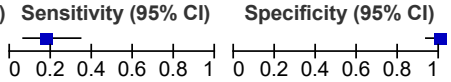


Figure 10: US, homogenously hypoechoic, in adults

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Pishdad 2017	16	5	18	48	0.47 [0.30, 0.65]	0.91 [0.79, 0.97]

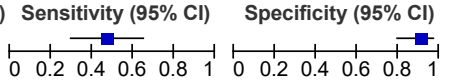


Figure 11: US, homogenously isoechoic, in adults

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Pishdad 2017	2	26	32	27	0.06 [0.01, 0.20]	0.51 [0.37, 0.65]

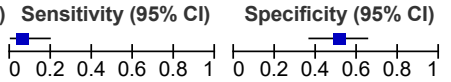


Figure 12: US, homogenously hyperechoic, in adults

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Pishdad 2017	5	22	29	31	0.15 [0.05, 0.31]	0.58 [0.44, 0.72]

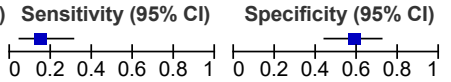


Figure 13: TSI, in children

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Baskaran 2015	31	0	6	10	0.84 [0.68, 0.94]	1.00 [0.69, 1.00]

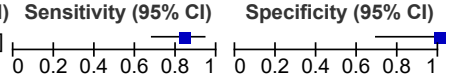


Figure 14: Technetium 99, in children

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
Baskaran 2015	37	0	0	10	1.00 [0.91, 1.00]	1.00 [0.69, 1.00]

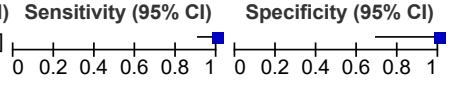


Figure 15: US, hypoechogenicity, in children

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
Lee 2016	12	24	2	48	0.86 [0.57, 0.98]	0.67 [0.55, 0.77]		

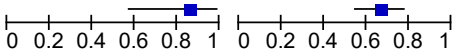


Figure 16: US, coarse echotexture, in children

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
Lee 2016	9	19	5	53	0.64 [0.35, 0.87]	0.74 [0.62, 0.83]		

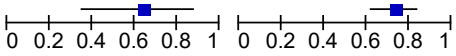


Figure 17: US, micronodularity, in children

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
Lee 2016	1	14	13	58	0.07 [0.00, 0.34]	0.81 [0.70, 0.89]		

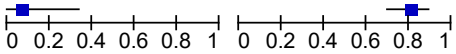
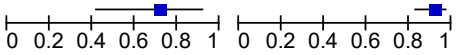


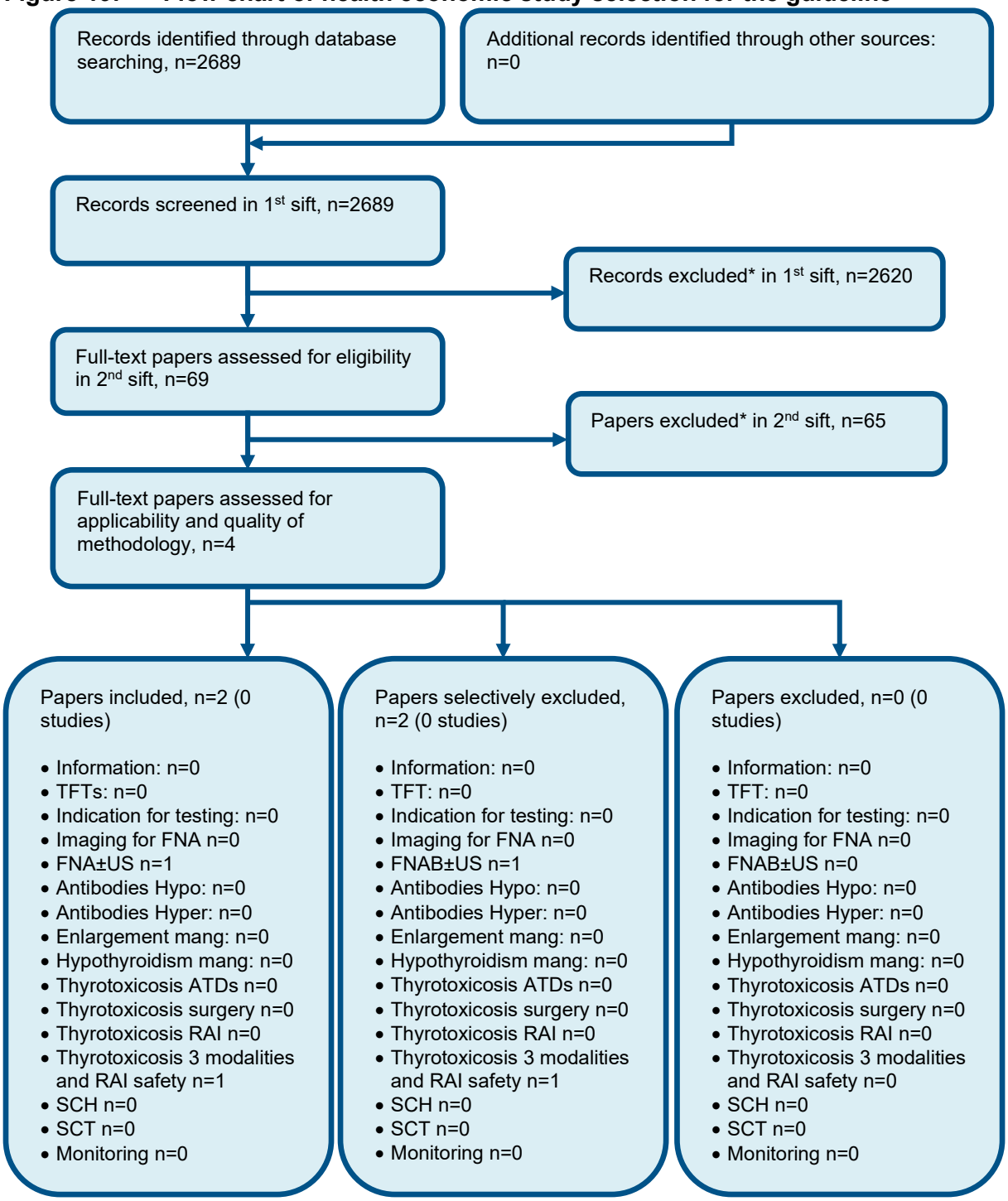
Figure 18: US, increased vascularity, in children

Study	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)
Lee 2016	10	6	4	66	0.71 [0.42, 0.92]	0.92 [0.83, 0.97]		



Appendix F: Health economic evidence selection

Figure 19: Flow chart of health economic study selection for the guideline



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

Appendix G: Health economic evidence tables

None

Appendix H: Health economic analysis

None

Appendix I: Excluded studies

I.1 Excluded clinical studies

Table 10: Studies excluded from the clinical review

Export title	Exclusion reason
Aleksic 2009[1]	Two gate study design
Banaka 2011[2]	Two gate study design
Banaka 2013[3]	Two gate study design
Barbesino 2013[4]	SR, references checked
Bell 2018[6]	Wrong study design
Bosi 2010[7]	Inappropriate population
Burman 1998[8]	SR, references checked
Cappelli 2007[9]	No usable outcomes
Cardia 2004[10]	Two gate study design
Carella 2006[11]	Two gate study design
Costagliola 1999[12]	Two gate study design
Diana 2014[13]	No usable outcomes
Diana 2016[14]	Two gate study design
Donkol 2013[15]	Inappropriate reference standard
Doroudian 2017[16]	Two gate study design
Duron 1987[17]	Two gate study design
Eckstein 2010[18]	SR, references checked
Engler 1994[19]	Two gate study design
Gassner 2009[20]	SR, references checked
Giovanella 2001[22]	Two gate study design
Giovanella 2001[21]	Two gate study design
Heberling 1988[23]	Two gate study design
Hirooka 2004[24]	Two gate study design
Iko 1986[25]	No usable outcomes
Kamath 2012[26]	SR, references checked
Kamijo 1999[30]	No usable outcomes
Kamijo 2003[27]	Two gate study design
Kamijo 2010[28]	Two gate study design
Kamijo 2011[29]	Two gate study design
Khoo 1997[31]	Two gate study design
Kotwal 2018[32]	SR, references checked
Laurberg 2006[33]	Two gate study design
Lee 2011[35]	Two gate study design
Lytton 2010[37]	Two gate study design
Lytton 2018[38]	SR, references checked
Mariotti 1989[39]	Two gate study design
Marwaha 2008[40]	Inappropriate population
Massart 2009[41]	Two gate study design
Maugendre 2001[42]	No usable outcomes
Meng 2015[43]	Two gate study design

Export title	Exclusion reason
Morgenthaler 2002[44]	Two gate study design
Morris 1988[45]	Two gate study design
Nishihara 2017[47]	Two gate study design
Ochi 1999[48]	Inappropriate population
Ochi 2000[49]	Two gate study design
Paunkovic 2003[51]	Two gate study design
Paunkovic 2007[52]	SR, references checked
Pedersen 2000[54]	Two gate study design
Pedersen 2001[53]	Two gate study design
Rago 2001[56]	Inappropriate population
Rosario 2014[57]	Inappropriate population
Sapin 2003[58]	Two gate study design
Schott 2000[59]	Two gate study design
Schott 2009[60]	Two gate study design
Sekulic 2006[61]	No usable outcomes
Smith 2007[62]	Two gate study design
Southgate 1984[63]	Two gate study design
Stozek 2018[64]	Two gate study design
Szabolcs 1995[67]	No usable outcomes
Takasu 1997[69]	Two gate study design
Takasu 2004[68]	Two gate study design
Tozzoli 2010[72]	Two gate study design
Tozzoli 2012[71]	SR, references checked
Uchida 2016[73]	Inappropriate population
Varadha 2016[74]	No usable outcomes
Vos 2008[75]	Inappropriate population
Wallaschofski 2001[76]	Two gate study design
Yoshimura Noh 2008[77]	Two gate study design
Zophel 2008[78]	Two gate study design
Zophel 2010[79]	Two gate study design
Zophel 2010[80]	No usable outcomes
Zouvanis 1998[81]	Two gate study design
Zuhur 2014[82]	Two gate study design

I.2 Excluded health economic studies

None