National Institute for Health and Care Excellence

Draft for Consultation

Thyroid disease: assessment and management

[F] Monitoring Thyroid Disease

NICE guideline

Intervention evidence review underpinning recommendations 1.4.1 to 1.4.6, 1.7.1 to 1.7.12 and 1.9.7 to 1.9.11 in the guideline. See also evidence review P

June 2019

Draft for Consultation

This evidence review was developed by the National Guideline Centre

1

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 All NICE guidance is subject to regular review and may be updated or withdrawn.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

Contents

1

1 Monitoring Thyroid Disease

1.1 Review question: How should non-malignant thyroid enlargement, hypothyroidism, thyrotoxicosis and subclinical thyroid dysfunction be monitored?

1.2 Introduction

1.2.1 Hypothroidism

Hypothyroidism is prevalent in 2% of the UK population and in more than 5% of people aged over 60, with women being 5–10 times more commonly affected than men. Long-term consequences of hypothyroidism include cardiovascular disease and an increase in cardiovascular risk factors including hypercholesterolaemia. These effects can be mitigated when hypothyroidism is effectively treated.

There is variation in how thyroid disease is managed and monitored in primary and secondary care settings. Standardisation in thyroid hormone replacement strategies and monitoring protocols for people with hypothyroidism are currently lacking.

Some patient support groups reported dissatisfaction with thyroid hormone replacement monitoring. In the UK patients may have different thyroid function tests to monitor their medication, and differing treatment targets, depending on where they live and which health care professional is responsible for treatment advice. By developing national guidance it is hoped that patients (and their support groups) will have evidence to advocate for their own monitoring and treatment, and doctors will have clear guidance to assist with consultations.

1.2.2 Thyrotoxicosis

Thyrotoxicosis occurs when there is an excess of thyroid hormones. It can be subdivided in to primary (caused by an abnormality in thyroid gland) and secondary (abnormality is in the pituitary gland or hypothalamus). It has a prevalence of 0.5-2% with a much higher female prevalence.

This NICE review will focus on monitoring of thyrotoxicosis following treatment with medication, radioactive iodine, surgery or a combination of the above.

Current practice is following treatment, patients' thyroid function is monitored with thyroid function tests (usually T4 and TSH) and frequency of testing varies depending on modality of treatment, whether the patient has been rendered hypothyroid with the treatment (more commonly with surgery and radioactive iodine) and any recurrence of symptoms. In addition thionamides (e.g. carbimazole and propylthioracil) are associated with rare and idiosyncratic side effects such as agranulocytosis and liver failure which need to be taken into consideration when monitoring during treatment. Patients who have been rendered hypothyroid following treatment for thyrotoxicosis should have replacement levothyroxine and monitored accordingly.

1.2.3 Subclinical Thyroid Dysfunction

Subclinical thyroid dysfunction encompasses both subclinical hypothyroidism (SCH) and subclinical thyrotoxicosis (STX). Both are common conditions and are characterised by serum TSH concentrations that are either above (SCH) or below (STX) the reference range,

with normal circulating thyroid hormone levels (FT4, FT3). Both conditions increase in prevalence with advancing age, and by the age of 70 years around 5% of the population have SCH and and 1% STX. Whilst the majority of these will revert to a euthyroid state following a period of observation, a small percentage in each group progress to overt hypo or hyperthroidism each year.

Strong evidence to favour active treatment of either condition is currently lacking, although some observational epidemiological studies have demonstrated an association with adverse health outcomes. At present there are no set standards for the monitoring of subclinical thyroid disease and the optimal monitoring strategy for each condition remains uncertain.

1.2.4 Nodule Monitoring

Thyroid nodules are lumps in the thyroid gland which are present in 15% of the UK population. Most are harmless although around 5% can be cancerous. They can be solid or cystic (fluid-filled lumps) and are more common in women than men. The frequency of nodules increases with age with the lifetime risk of developing a thyroid nodule being estimated at 5–10%. Some nodules produce too much thyroid hormone leading to thyrotoxicosis (an overactive thyroid).

In the UK there is variation in the monitoring of thyroid nodules. Some hospital clinics discharge patients when cancer has been excluded, whereas other clinics monitor nodules and potentially repeat investigations over a period of time.

This guidance covers the monitoring of nodules which have been proven to be benign (non-cancerous) in patients with normal thyroid function (euthyroid).

1.3 PICO table

For full details see the review protocols in Appendix A:.

Table 1: PICO characteristics of review question

Population	People with thyroid disease, stratified by: - thyrotoxicosis - hypothyroidism - subclinical thyroid dysfunction - thyroid enlargement
Interventions	No monitoring Monitoring +/- TSH, T4, T3, FBC, LFTs *Hypothyroidism* only – monitoring aimed at specific TSH target (<3 U/mL) vs TSH within reference range only
Comparisons	Any of the above vs any other at any frequency (6 monthly, annual, less than once a year)
Outcomes	**Critical** • Mortality (dichotomous, ≥1 year) • Quality of life (continuous) **Important** • Thyroid ophthalmopathy (dichotomous) • Euthyroidism (dichotomous) • Hypothyroidism (dichotomous) • Relapse of hyperthyroidism (dichotomous) • Cardiovascular morbidity (ischaemic heart disease, dichotomous)

	• Arrhythmia (dichotomous) • Osteoporosis (dichotomous) • Cognitive impairment (dichotomous) • Pain (continuous) • Symptom scores (continuous) • Patient/family/carer experience (continuous) • Healthcare contacts (rates/dichotomous) • Agranulocytosis (dichotomous) • Liver failure (dichotomous) • Minor drug related adverse effects (dichotomous) • Infertility (dichotomous) • Malignancy (dichotomous) • Growth abnormalities (dichotomous)
Study design	• RCTs preferred, if no RCTs available to consider non-randomised cohort studies in which key confounders (age, sex, co-existing conditions, treatment received) are addressed, either through restriction (see stratifications below) or appropriate matching/statistical adjustment

1.4 Clinical evidence

1.4.1 Included studies

One study was included in the review;[8] this is summarised in Table 2 below. Evidence from this study is summarised in the clinical evidence summary below (Table 3).

One study was an RCT in adults comparing different TSH targets in the treatment of hypothyroidism.

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

1.4.2 Excluded studies

See the excluded studies list in Appendix J:.

1.4.3 Summary of clinical studies included in the evidence review

Table 2: Summary of studies included in the evidence review

Study	Intervention and comparison	Population	Outcomes	Comments
Samuels 2018[8]	Aimed at specific TSH target, n = 46. Target of low-normal TSH (0.34 to 2.5 mU/L), mean dose at end of study 1.5ug/kg/day Aimed at normal TSH, n = 47. Target of high-normal TSH (2.51 to 5.6 mU/L), mean dose at end of study 1.32ug/kg/day	Adults (mean age 49.2, SD1) Hypothyroid, taking stable dose of levothyroxine for at least 3 months USA	Quality of life Symptom scores Reported at 6 months	17% of low normal arm did not achieve target 64% of high normal did not achieve target

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: aiming at low normal TSH vs aiming at high normal TSH

Outcomes	No of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with High-normal TSH	Risk difference with Low-normal TSH (95% CI)
Quality of life - SF-36, mental component Scale from: 0 to 100.	93 (1 study) 6 months	⨁⨁⊖⊖ LOW1,2 due to risk of bias, imprecision		The mean quality of life - sf-36, mental component in the control groups was 39.8	The mean quality of life - sf-36, mental component in the intervention groups was 1.2 higher (1.58 lower to 3.98 higher)
Quality of life - SF-36, physical component	93 (1 study)	⨁⨁⊖⊖ LOW1,2		The mean quality of life - sf-36, physical component in the control	The mean quality of life - sf-36, physical component in the intervention groups

Outcomes	No of Participants (studies) Follow up	Quality of the evidence (GRADE)	Relative effect (95% CI)	Anticipated absolute effects	
				Risk with High-normal TSH	Risk difference with Low-normal TSH (95% CI)
Scale from: 0 to 100.	6 months	due to risk of bias, imprecision		groups was 50.6	was 2.1 lower (4.03 to 0.17 lower)
Quality of life - Thyroid specific, TDQ-AW Scale from: -9 to +3.	93 (1 study) 6 months	⨁⊖⊖⊖ VERY LOW1,2 due to risk of bias, imprecision		The mean quality of life - thyroid specific, tdq-aw in the control groups was -1.4	The mean quality of life - thyroid specific, tdq-aw in the intervention groups was 0.2 lower (0.77 lower to 0.37 higher)
Symptom scores - Billewicz Scale from: -47 to +67.	93 (1 study) 6 months	⨁⊖⊖⊖ VERY LOW1,2 due to risk of bias, imprecision		The mean symptom scores - Billewicz in the control groups was 4	The mean symptom scores - Billewicz in the intervention groups was 1.2 lower (2.45 lower to 0.05 higher)

1 Downgraded by 1 increment if the majority of the evidence was at high risk of bias, and downgraded by 2 increments if the majority of the evidence was at very high risk of bias
2 Downgraded by 1 increment if the confidence interval crossed one MID or by 2 increments if the confidence interval crossed both MIDs

See Appendix F: for full GRADE tables.

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 G:.

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 4: UK costs of blood tests

Tests	*Unit cost*
TSH (a)	£2.15
FT3	£3.12
FT4	£2.10
TRAb antibody testing	£16.64
TPO antibody testing	£12.32
Phlebotomy (b)	£3.04

Source: Costs obtained from different hospitals
(a) Costs quoted include reagent, consumables and staff pay
(b) NHS reference cost 2016-17, currency code DAPS08

1.6 Evidence statements

1.6.1 Clinical evidence statements

Aiming for low normal TSH vs aiming for high normal TSH

No clinically important difference was identified for quality of life – SF-36 mental component (1 study, low quality), SF-36 physical component (1 study, low quality), quality of life – TDQ-AW (1 study, very low quality), symptom scores (1 study, very low quality).

1.6.2 Health economic evidence statements

- No relevant economic evaluations were identified.

F1. If hyperthyroidism persists after radioactive iodine treatment in adults, children and young people, consider antithyroid drugs until the 6-month appointment.

F2. If hyperthyroidism persists 6 months after radioactive iodine treatment in adults, children and young people, consider further treatment.

1.7 The committee's discussion of the evidence

1.7.1 Interpreting the evidence

1.7.1.1 The outcomes that matter most

Critical outcomes for this review were mortality and quality of life. Important outcomes for this review were thyroid ophthalmopathy, euthyroidism, hypothyroidism, cardiovascular morbidity, arrhythmia, osteoporosis, cognitive impairment, pain, symptom scores, and patient/family/carer experience and healthcare contacts.

Thyrotoxicosis specific important outcomes were relapse of hyperthyroidism, agranulocytosis, liver failure, minor drug related adverse effects, infertility, malignancy and growth abnormalities.

Hypothyroidism specific important outcomes included TSH suppression.

Enlargement specific important outcomes included stage/timing of malignancy diagnosis.

1.7.1.2 The quality of the evidence

Thyrotoxicosis

There was no clinical evidence identified in this review.

Hypothyroidism

No evidence was available on the optimum timing or intensity of monitoring. One study was available comparing different TSH targets. This evidence was low to very low quality due to risk of bias and imprecision.

Thyroid enlargement

There was no clinical evidence identified in this review.

Subclinical thyroid dysfunction

There was no clinical evidence identified in this review.

1.7.1.3 Benefits and harms

General

The benefits of more intense monitoring are that they may allow for earlier detection of deviations in thyroid function that could benefit from alterations in treatment strategy. The harms of more intense monitoring are that they incur treatment burden for the person with thyroid disease and the healthcare service and they may encourage interventions that will not actually have a significant impact. Overall the committee made consensus recommendations that they agreed struck an acceptable balance between safe monitoring and preventing excessive investigation.

Thyrotoxicosis

The committee noted that the need for monitoring during antithyroid drug therapy may vary depending on the regimen followed. Block and replace treatment, as opposed to titration, is considered by some to provide a more stable dosing and therefore may require less frequent monitoring. However this was not based on evidence identified in the review and indeed

some committee members noted anecdotally that this had not been their experience in analyses of their own services.

Hypothyroidism

The only available evidence in this review related to the comparison of aiming for a low-normal TSH target or a high-normal TSH target. There was no clinically important difference between these options for quality of life (generic physical, generic mental or disease specific) and for symptoms scores. The committee noted that for some people aiming for a low-normal range would require aggressive overtreatment which will increase the risks of adverse effects of treatment.

The committee noted that reference ranges are ranges and that for each individual person there will be variability as to what exact TSH targets are most appropriate for them.

In the monitoring of people with primary hypothyroidism being treated with levothyroxine who are not experiencing symptoms, the committee was in agreement that TSH alone was a sufficient investigation. The committee agreed that in this situation they commonly see a variety of TFTs conducted and that these did not provide any additional benefit. The committee noted that in some situations when people have symptoms there may be some benefit of looking at additional TFTs (for example FT4 or FT3) however there was not full consensus over this and given the potential resource impact, the committee did not make a specific recommendation over practice in this scenario.

The committee noted that which tests are done is often dictated by the setting of the investigation. In primary care it was not uncommon for people to do a FT4 test if TSH was normal in people presenting with symptoms.

The committee agreed that in some cases monitoring FT4 during treatment may cause harm as healthcare professionals may inappropriately titrate levothyroxine to the FT4 level as opposed to the TSH. However for some people who struggle with compliance to levothyroxine therapy, FT4 can be a marker of adherence.

Subclinical thyroid dysfunction

There was no evidence identified in this review to guide recommendation making. The committee agreed that in general it may be useful to monitor people with subclinical thyroid dysfunction that does not warrant treatment in case it transitions to clinical disease or a severity warranting treatment. However they were also keen to insure that people were not monitored indefinitely when this was not necessary, therefore they included recommendations about when to consider stopping monitoring. For people with treated subclinical thyroid dysfunction, monitoring should be as per monitoring for clinical variants of the relevant condition.

1.7.2 Cost effectiveness and resource use

No health economic evidence was identified for these monitoring questions. The committee could not assess the cost effectiveness of how people with thyroid dysfunction (thyroid enlargement, subclinical thyroid, hypothyroidism and thyrotoxicosis) should be monitored. The unit costs of the different thyroid function tests were presented to the committee.

The costs of TSH was estimated as approximate £2.15, FT3 costs £3.12 and FT4 costs £2.10, these costs include the reagent, medical and staffing cost involved in analysing the results. The cost of staff time and equipment required for taking blood for testing (phlebotomy cost) was taken separately from the NHS reference cost 2016-2017 at £3.04.

The cost of the different thyroid function tests was obtained from five different hospitals. The committee recognised that there is great heterogeneity between the costs of blood tests

obtained from different hospital laboratories, which was mainly due to the way staffing cost is calculated across different hospitals. The committee agreed that a median cost should be used as the results are skewed. Overall, the committee agreed that FT3 was the most expensive thyroid function test and should only be used in some cases as outlined in the recommendations. In addition, it was highlighted that doing combined tests when necessary could reduce the costs of the phlebotomy and have benefits for patients QALYs, as they may avoid additional blood tests and a quicker diagnosis may be obtained.

In general, more frequent monitoring will be associated with higher monitoring costs, as more healthcare professional time is required to see patients and review results. However, as described above there may also be benefits of more intense monitoring in terms of resource use and/or patients' health. Overall, the committee made consensus recommendations that they agreed struck an acceptable balance between safe monitoring and preventing excessive investigation.

Thyrotoxicosis

The committee agreed that monitoring thyrotoxicosis should vary depending on the treatment received. In general and in line with current practice TFTs should be repeated 6 weeks after diagnosis and then only if people have symptoms. The group emphasised that a TSH test is sufficient over the long term and additional tests should not be requested unless in special circumstances.

RAI – the committee made explicit how monitoring should be carried out after RAI therapy. This is because RAI was being recommended as first line treatment for thyrotoxicosis and had common side effects such as hypothyroidism. The committee recognised that carrying out TSH, FT4 and FT3 testing every 6 weeks for the first 6 months could improve detection of these side effects and reduce complications, which could be expensive to treat. Additionally, detecting hypothyroidism or persistent thyrotoxicosis is likely to have a positive effect on their quality of life. The committee recommended treating side effects such as hypothyroidism with levothyroxine, as this is likely to reduce the need for monitoring and avoid unnecessary risk of developing other conditions due to untreated hypothyroidism. Overall RAI monitoring strategy was in line with current practice and was not likely to have a substantial resource impact. However, as treatment of thyrotoxicosis is changing and RAI would be offered as first line more people are expected to be monitored. Nevertheless, the cost is likely to be justified, as RAI is more cost-effective than the other treatments for thyrotoxicosis as shown in the included study, Donovan et al. 2016[35], for the management of thyrotoxicosis review question.

Surgery – the group emphasised the importance of offering levothyroxine to people who have had total thyroidectomy, as this will reduce the need for monitoring. This improves patient's quality of life as early treatment avoids unnecessary risk of developing other conditions due to untreated hypothyroidism. Furthermore, money can be saved by reducing healthcare visits and monitoring. People who have had subtotal thyroidectomy need monitoring as there is a risk of thyrotoxicosis recurring and monitoring could help with early detection, hence earlier treatment and better quality of life in addition to reducing costs of complication associated with untreated thyrotoxicosis and mortality.

ATDs – the committee considered measuring the TSH, FT4, and FT3 in line with current practice, until thyroid function is normalised then to continue with TSH testing alone until end of treatment. It emphasised that patients' full blood count and liver function tests should not be monitored unless there is suspicion of liver dysfunction or agranulocytosis, currently it is common, in some areas full blood count, and liver function tests are routine practice for all people on ATDs. Therefore, this recommendation is likely to reduce costs of unnecessary blood tests requested by healthcare professionals.

The committee noted that monitoring should continue in people who have stopped treatment to detect early signs and symptoms of relapse or complications, which is likely to avoid downstream costs.

Hypothyroidism

The clinical review identified evidence comparing maintaining TSH within a high-normal range and a lower-normal range. The committee noted that maintaining TSH at a lower range was likely to be a higher cost strategy as it would require more monitoring. Given this and the lack of evidence of a clinical benefit of maintaining TSH at a lower range they felt it was important to make a good practice recommendation in regards to TSH being maintained within a reference range rather than specifically at the lower end of the range.

The committee also wanted to alert clinicians to the prolonged period sometimes needed to normalise TSH levels when adjusting the dose of levothyroxine used, particularly in people with high TSH levels before starting treatment or people who have had long periods of untreated hypothyroidism. This is likely to be cost saving as it reduces resource use such as healthcare practitioners time and blood tests.

The committee also considered monitoring people with TSH testing alone, after starting levothyroxine treatment until TSH level has stabilised. The committee noted that practitioners should not be testing FT4, as it does not give additional information, unless the patient is experiencing symptoms. There was a strong consensus that these recommendations may potentially improve early detection of uncontrolled thyroid disease, and hence improve treatment, survival time, and quality of life. They are also likely to be cost saving as they reduce unnecessary FT4 testing.

The committee did not make a recommendation to measure FT3 or FT4 as well as TSH in people who do not have symptoms of hypothyroidism after starting levothyroxine given there was insufficient evidence that this conferred additional benefit and as there is variation in practice currently there would be potential for a substantial resource impact to the NHS in England if it were to be recommended as well as TSH.

The committee understood that children would need more intense monitoring, as the impact of poorly treated hypothyroidism is likely to be more severe. Frequent monitoring would therefore reduce unnecessary downstream costs and improve quality of life.

These recommendations are likely to be cost saving as they reduce unnecessary monitoring and tests and are currently in line with current practice hence not resulting in a substantial resource impact.

Thyroid enlargement

The committee wanted to make clear that only a small number of people with untreated benign thyroid enlargement would need any type of monitoring, and therefore this recommendation is likely to reduce costs through unnecessary testing. The committee noted that tests should be carried out via TSH, which is considered a low cost test, and US scan, which is also noted to be the cheapest scan and that these forms of testing would be sufficient. By monitoring the condition, quality of life can be improved for patients.

Subclinical thyroid dysfunction (subclinical hypothyroidism and subclinical thyrotoxicosis)

No clinical evidence was available comparing different testing strategies in people with subclinical thyroid disease but the committee's consensus was that annual TSH and FT4

testing should be considered in people with features suggesting subclinical thyroid dysfunction. They also agreed that monitoring could be stopped if TSH is within the reference range and there are no features suggesting underlying thyroid disease without adversely affecting patients' health outcomes. The committee understood that children would need more intense monitoring, as the impact of untreated thyroid disease is likely to be more severe. Frequent monitoring would therefore reduce unnecessary downstream costs and improve quality of life. The committee highlighted that current practice for monitoring of subclinical disease was not standardised and so varied considerably with unnecessary testing often being undertaken. They agreed it is likely that these recommendation would reduce the number of tests overall and therefore reducing costs in the NHS.

1.7.3 Other factors the committee took into account

The committee discussed the importance of repeating thyroid function tests shortly after diagnosis. The committee agreed that in their experience there is considerable variability in the time between diagnosis of thyrotoxicosis and initiation of treatment, in some places people may be prescribed antithyroid drugs in primary care as a stop gap measure until specialist referral is available. However in other places primary healthcare professionals are unwilling to initiate antithyroid drugs without specialist input. The committee agreed that it was unacceptable to leave people with thyrotoxicosis without treatment antithyroid treatment and noted that a number of models are available to address this problem (for example providing specialist advice remotely before appointments).

References

1. Ajmal S, Rapoport S, Ramirez Batlle H, Mazzaglia PJ. The natural history of the benign thyroid nodule: what is the appropriate follow-up strategy? Journal of the American College of Surgeons. 2015; 220(6):987-92

2. Balhara B, Misra M, Levitsky LL. Clinical monitoring guidelines for congenital hypothyroidism: laboratory outcome data in the first year of life. Journal of Pediatrics. 2011; 158(4):532-7

3. Helfand M, Crapo LM. Monitoring therapy in patients taking levothyroxine. Annals of Internal Medicine. 1990; 113(6):450-4

4. Lee S, Skelton TS, Zheng F, Schwartz KA, Perrier ND, Lee JE et al. The biopsy-proven benign thyroid nodule: is long-term follow-up necessary? Journal of the American College of Surgeons. 2013; 217(1):81-8; discussion 88-9

5. Medici M, Liu X, Kwong N, Angell TE, Marqusee E, Kim MI et al. Long- versus short-interval follow-up of cytologically benign thyroid nodules: a prospective cohort study. BMC Medicine. 2016; 14:11

6. 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

7. Phitayakorn R, McHenry CR. Follow-up after surgery for benign nodular thyroid disease: evidence-based approach. World Journal of Surgery. 2008; 32(7):1374-84

8. Samuels MH, Kolobova I, Niederhausen M, Janowsky JS, Schuff KG. Effects of Altering Levothyroxine (L-T4) Doses on Quality of Life, Mood, and Cognition in L-T4 Treated Subjects. The Journal of Clinical Endocrinology & Metabolism. 2018; 103(5):1997-2008

Appendices

Appendix A: Review protocols

Table 5: Review protocol: Monitoring Thyrotoxicosis

ID	Field	Content
I	Review questions	How should thyrotoxicosis be monitored?
II	Type of review question	Intervention 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	Determine the most clinically and cost effective way to monitor thyrotoxicosis
IV	Eligibility criteria – population / disease / condition / issue / domain	People with thyrotoxicosis (clinical or subclinical thyrotoxicosis requiring treatment)
V	Eligibility criteria – intervention(s) / exposure(s) / prognostic factor(s)	No monitoring Monitoring • with or without thyroid function tests (TSH +/- T4 +/- T3, exact composition free to be determined by studies) • with or without FBC, LFTs • at any frequency (stratified into 6 monthly, annual, less frequently)
VI	Eligibility criteria – comparator(s) / control or reference (gold) standard	No monitoring vs monitoring Any monitoring strategy vs any other monitoring strategy
VII	Outcomes and prioritisation	**Critical** • Mortality (dichotomous, ≥1 year) • Quality of life (continuous) **Important** • Thyroid ophthalmopathy (dichotomous) • Euthyroidism (dichotomous) • Hypothyroidism (dichotomous) • Relapse of hyperthyroidism (dichotomous) • Cardiovascular morbidity (ischaemic heart disease, dichotomous) • Arrhythmia (dichotomous) • Osteoporosis (dichotomous) • Cognitive impairment (dichotomous) • Pain (continuous) • Symptom scores (continuous) • Patient/family/carer experience (continuous) • Healthcare contacts (rates/dichotomous) • Agranulocytosis (dichotomous) • Liver failure (dichotomous)

		• Minor drug related adverse effects (dichotomous) • Infertility (dichotomous) • Malignancy (dichotomous) • Growth abnormalities (dichotomous) Minimum duration as for the minimum duration for inclusion of studies unless specified.
VIII	Eligibility criteria – study design	• RCTs preferred, if no RCTs available to consider non-randomised cohort studies in which key confounders (age, sex, co-existing conditions, treatment received) are addressed, either through restriction (see stratifications below) or appropriate matching/statistical adjustment • Minimum duration of 1 year • Crossover studies excluded
IX	Other inclusion exclusion criteria	-
X	Proposed sensitivity / subgroup analysis, or meta-regression	**Stratifications** • Age - infants <4, children (4-18), adults (>18-65), older adults (>65) • Underlying condition – clinical thyrotoxicosis vs subclinical thyrotoxicosis • Treatment received (modality or submodality level) • Treatment stage – for example during ATD treatment vs following ATD treatment • Status of population – normalised thyroid hormones vs uncontrolled thyrotoxicosis **Subgroup analyses** • Age subdivisions (18-50, 50-65, 65-85, >85)
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
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
XVII	Search strategy – for one database	For details please see Appendix B:.
XVI	Data collection	Not applicable

II	process – forms / duplicate	
XIX	Data items – define all variables to be collected	Not applicable
XX	Methods for assessing bias at outcome / study level	Not applicable
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 [to add link to history page of the guideline after publication] 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 number	Not registered

1
2

1 **Table 6: Review protocol: Monitoring hypothyroidism**

ID	Field	Content
I	Review questions	How should hypothyroidism be monitored?
II	Type of review question	Intervention 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	Determine the most clinically and cost effective way to monitor hypothyroidism
IV	Eligibility criteria – population / disease / condition / issue / domain	People with hypothyroidism (or subclinical hypothyroidism requiring treatment)
V	Eligibility criteria – intervention(s) / exposure(s) / prognostic factor(s)	Monitoring • aimed at specific TSH target (<3 U/mL) vs TSH within reference range only • with or without thyroid function tests (TSH +/- T4 +/- T3, exact composition free to be determined by studies) • at any frequency (stratified into 6 monthly, annual, less frequently)
VI	Eligibility criteria – comparator(s) / control or reference (gold) standard	Any monitoring strategy vs any other monitoring strategy
VII	Outcomes and prioritisation	**Critical** • Mortality (dichotomous, ≥1 year) • Quality of life (continuous) **Important** • Cardiovascular morbidity - ischemic heart disease, heart failure (dichotomous) • Arrhythmias (dichotomous) • Osteoporosis (dichotomous) • Impaired cognitive function (dichotomous) • Depression (dichotomous) • Patient/family/carer experience (continuous) • Healthcare contacts (rates/dichotomous) • Symptom scores (continuous) • Growth (continuous) • TSH suppression (dichotomous) Minimum duration as for the minimum duration for inclusion of studies unless specified.
VIII	Eligibility criteria – study design	• 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 (see stratifications below) or appropriate matching/statistical adjustment • Minimum duration of 3 months for comparison of TSH targets • Minimum duration of 1 year for all other comparisons

		• Crossover studies excluded
IX	Other inclusion exclusion criteria	Excluding studies in population with hypothyroidism post-thyroid cancer treatment
X	Proposed sensitivity / subgroup analysis, or meta-regression	**Stratifications** • Age – infants (<4), children (4-18), adults (>18-65), older adults (>65) • Underlying condition – subclinical hypothyroidism vs clinical hypothyroidism • Treatment stage – early vs stable TFTs, as defined by studies • Cause of primary hypothyroidism – endogenous, iatrogenic **Subgroup analyses** • Age subdivisions (18-50, 50-65, 65-85, >85)
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)	• Pairwise meta-analyses were performed using Cochrane Review Manager (RevMan5). • GRADEpro was used to assess the quality of evidence for each outcome. • Endnote was used for bibliography, citations, sifting and reference management
XIII	Information sources – databases and dates	[List sources to be searched, limits to be applied to the search, and plans to use any supplemental search techniques with rationale for their use. Descibe other sources of evidence (calls for evience). List key papers if known.] • Consider cut-off dates if an update.]
XIV	Identify if an update	Not an update
XV	Author contacts	[Add link to the In development page for the guideline on the NICE website]
XVI	Highlight if amendment to previous protocol	Not an amendment
XVII	Search strategy – for one database	For details please see Appendix B:.
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 Appendix H: (health economic evidence tables).
XX	Methods for assessing bias at outcome / study level	Standard study 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 [to add link to history page of the guideline after publication] 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 number	Not registered

1
2

1 **Table 7: Review protocol: Monitoring non-malignant thyroid enlargement**

ID	Field	Content
I	Review questions	How should non-malignant thyroid enlargement be monitored?
II	Type of review question	Intervention 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	Determine the most clinically and cost effective way to monitor non-malignant thyroid enlargement that has not required intervention
IV	Eligibility criteria – population / disease / condition / issue / domain	People with non-malignant thyroid enlargement that has not required intervention
V	Eligibility criteria – intervention(s) / exposure(s) / prognostic factor(s)	No monitoring Monitoring • with or without thyroid function tests (TSH +/- T4 +/- T3, exact composition free to be determined by studies), imaging • at any frequency (stratified into 6 monthly, annual, less frequently)
VI	Eligibility criteria – comparator(s) / control or reference (gold) standard	Monitoring vs no monitoring Any monitoring strategy vs any other monitoring strategy
VII	Outcomes and prioritisation	**Critical** • Mortality (dichotomous, ≥1 year) • Quality of life (continuous) **Important** • Stage/timing of malignancy diagnosis (continuous/dichotomous) • Hyperthyroidism (dichotomous) • Patient/family/carer experience (continuous) • Healthcare contacts (rates/dichotomous) Minimum duration as for the minimum duration for inclusion of studies unless specified.
VIII	Eligibility criteria – study design	• RCTs preferred, if no RCTs available to consider non-randomised cohort studies in which key confounders (age, sex, co-existing conditions, size of nodule, type of nodule) are addressed, either through restriction (see stratifications below) or appropriate matching/statistical adjustment • Minimum duration of 1 year • Crossover studies excluded
IX	Other inclusion exclusion criteria	-
X	Proposed sensitivity / subgroup analysis, or meta-	**Stratifications** • Nodule vs goitre • Age – infants (<4), children (4-18), adults (>18-65), older adults (>65) • Type of nodule (benign vs indeterminate nodule, cystic vs non-cystic)

	regression	**Subgroup analyses** • Age subdivisions (18-50, 50-65, 65-85, >85)
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
XIII	Information sources – databases and dates	• Medline (OVID), Embase (OVID) an the Cochrane Library (Wiley)
XIV	Identify if an update	Not an update
XV	Author contacts	[Add link to the In development page for the guideline on the NICE website]
XVI	Highlight if amendment to previous protocol	Not an amendment
XVII	Search strategy – for one database	For details please see Appendix B:.
XVIII	Data collection process – forms / duplicate	Not applicable
XIX	Data items – define all variables to be collected	Not applicable
XX	Methods for assessing bias at outcome / study level	Not applicable
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	For details please see sections 6.4 and 9.1 of Developing NICE guidelines: the manual.

	evidence	
XXV	Rationale / context – what is known	For details please see the introduction to the evidence review.
XXVI	Describe contributions of authors and guarantor	[STANDARD TEXT] A multidisciplinary committee [add link to history page of the guideline] 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 number	Not registered

1
2

1 **Table 8: Review protocol: Monitoring subclinical thyroid dysfunction**

ID	Field	Content
I	Review questions	How should subclinical thyroid dysfunction be monitored?
II	Type of review question	Intervention 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	Determine the most clinically and cost effective way to monitor subclinical thyroid dysfunction
IV	Eligibility criteria – population / disease / condition / issue / domain	People with subclinical thyroid dysfunction that does not require treatment
V	Eligibility criteria – intervention(s) / exposure(s) / prognostic factor(s)	No monitoring Monitoring • with or without thyroid function tests (TSH +/- T4 +/- T3, exact composition free to be determined by studies) • at any frequency (stratified into 6 monthly, annual, less frequently)
VI	Eligibility criteria – comparator(s) / control or reference (gold) standard	Monitoring vs no monitoring Any monitoring strategy vs any other monitoring strategy
VII	Outcomes and prioritisation	**Critical** • Mortality (dichotomous, ≥1 year) • Quality of life (continuous) **Important** • Thyroid ophthalmopathy (dichotomous) • Euthyroidism (dichotomous) • Hypothyroidism (dichotomous) • Cardiovascular morbidity (ischaemic heart disease, dichotomous) • Arrhythmia (dichotomous) • Osteoporosis (dichotomous) • Cognitive impairment (dichotomous) • Pain (continuous) • Depression (dichotomous) • Symptom scores (continuous) • Patient/family/carer experience (continuous) • Healthcare contacts (rates/dichotomous) • Number of people initiating treatment for thyroid disease (dichotomous) Minimum duration as for the minimum duration for inclusion of studies unless specified.
VIII	Eligibility criteria – study design	• 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 (see stratifications below) or appropriate matching/statistical adjustment

		• Minimum duration of 1 year • Crossover studies excluded
IX	Other inclusion exclusion criteria	-
X	Proposed sensitivity / subgroup analysis, or meta-regression	**Stratifications** • Age – infants (<4), children (4-18), adults (>18-65), older adults (>65) • Thyroid disease – SCH vs SCT • TSH abnormality –SCT: <0.1 U/mL vs 0.1 U/mL – lower limit of reference range) **Subgroup analyses** • Age subdivisions (18-50, 50-65, 65-85, >85)
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
XIII	Information sources – databases and dates	• Medline (OVID), Embase (OVID) and the Cochrane Library (Wiley)
XIV	Identify if an update	Not an update
XV	Author contacts	
XVI	Highlight if amendment to previous protocol	Not an amendment
XVII	Search strategy – for one database	For details please see Appendix B:.
XVIII	Data collection process – forms / duplicate	Not applicable
XIX	Data items – define all variables to be collected	Not applicable
XX	Methods for assessing bias at outcome / study level	Not applicable
XXI	Criteria for quantitative synthesis	For details please see section 6.4 of Developing NICE guidelines: the manual.
XXII	Methods for quantitative analysis –	For details please see the separate Methods report for this guideline.

	combining studies and exploring (in)consistency	
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 [to add link to history page of the guideline after publication] 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 number	Not registered

1
2

1 **Table 9: 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).[6] **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.

1
2

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. [Add cross reference after publication]

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 10: 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
Embase (OVID)	1974 – 07 January 2019	Exclusions Randomised controlled trials Systematic review studies Observational 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 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.	exp goiter/
7.	exp Hyperthyroidism/
8.	(hyperthyroid* or thyrotoxicosis).ti,ab.
9.	(toxic adj4 (node* or nodul* or multi?nodul* or goitre or goiter)).ti,ab.
10.	(graves' disease or plummer's disease).ti,ab.
11.	(((thyroid adj4 (swell* or enlarg* or nodule* or node*)) or (goitre* or goiter*)) adj5 (non-malignan* or nonmalignan* or benign)).ti,ab.
12.	Thyroid Nodule/

13.	or/1-12
14.	letter/
15.	editorial/
16.	news/
17.	exp historical article/
18.	Anecdotes as Topic/
19.	comment/
20.	case report/
21.	(letter or comment*).ti.
22.	or/14-21
23.	randomized controlled trial/ or random*.ti,ab.
24.	22 not 23
25.	animals/ not humans/
26.	exp Animals, Laboratory/
27.	exp Animal Experimentation/
28.	exp Models, Animal/
29.	exp Rodentia/
30.	(rat or rats or mouse or mice).ti.
31.	or/25-30
32.	13 not 31
33.	Monitoring, Physiologic/
34.	Patient Outcome Assessment/
35.	monitoring.ti,ab.
36.	((followup or follow-up or check* or evaluat* or appointment* or observation or observations) adj3 (timing* or interval* or year* or annual* or month* or periodic* or frequen* or routine*)).ti,ab.
37.	((monitor* or time point* or interval*) adj3 (year* or annual* or month* or periodic* or frequen* or routine*)).ti,ab.
38.	or/33-37
39.	32 and 38
40.	limit 39 to English language
41.	randomized controlled trial.pt.
42.	controlled clinical trial.pt.
43.	randomi#ed.ti,ab.
44.	placebo.ab.
45.	randomly.ti,ab.
46.	Clinical Trials as topic.sh.
47.	trial.ti.
48.	or/41-47
49.	Meta-Analysis/
50.	exp Meta-Analysis as Topic/
51.	(meta analy* or metanaly* or metaanaly* or meta regression).ti,ab.
52.	((systematic* or evidence*) adj3 (review* or overview*)).ti,ab.
53.	(reference list* or bibliograph* or hand search* or manual search* or relevant journals).ab.
54.	(search strategy or search criteria or systematic search or study selection or data

	extraction).ab.
55.	(search* adj4 literature).ab.
56.	(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.
57.	cochrane.jw.
58.	((multiple treatment* or indirect or mixed) adj2 comparison*).ti,ab.
59.	or/49-58
60.	Epidemiologic studies/
61.	Observational study/
62.	exp Cohort studies/
63.	(cohort adj (study or studies or analys* or data)).ti,ab.
64.	((follow up or observational or uncontrolled or non randomi#ed or epidemiologic*) adj (study or studies or data)).ti,ab.
65.	((longitudinal or retrospective or prospective or cross sectional) and (study or studies or review or analys* or cohort* or data)).ti,ab.
66.	Controlled Before-After Studies/
67.	Historically Controlled Study/
68.	Interrupted Time Series Analysis/
69.	(before adj2 after adj2 (study or studies or data)).ti,ab.
70.	or/60-69
71.	exp case control study/
72.	case control*.ti,ab.
73.	or/71-72
74.	70 or 73
75.	Cross-sectional studies/
76.	(cross sectional and (study or studies or review or analys* or cohort* or data)).ti,ab.
77.	or/75-76
78.	70 or 77
79.	70 or 73 or 77
80.	40 and (48 or 59 or 79)

1 **Embase (Ovid) search terms**

1.	exp thyroid disease/
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.	(((thyroid adj4 (swell* or enlarg* or nodule* or node*)) or (goitre* or goiter*)) adj5 (non-malignan* or nonmalignan* or benign)).ti,ab.
7.	Thyroid Nodule/
8.	hyperthyroidism/ or graves disease/ or thyrotoxicosis/ or toxic goiter/
9.	(hyperthyroid* or thyrotoxicosis).ti,ab.
10.	(toxic adj4 (node* or nodul* or multi?nodul* or goitre or goiter)).ti,ab.
11.	(graves' disease or plummer's disease).ti,ab.
12.	or/1-11
13.	letter.pt. or letter/

14.	note.pt.
15.	editorial.pt.
16.	case report/ or case study/
17.	(letter or comment*).ti.
18.	or/13-17
19.	randomized controlled trial/ or random*.ti,ab.
20.	18 not 19
21.	animal/ not human/
22.	nonhuman/
23.	exp Animal Experiment/
24.	exp Experimental Animal/
25.	animal model/
26.	exp Rodent/
27.	(rat or rats or mouse or mice).ti.
28.	or/20-27
29.	12 not 28
30.	physiologic monitoring/
31.	patient monitoring/
32.	patient monitoring/
33.	monitoring.ti,ab.
34.	((followup or follow-up or check* or evaluat* or appointment* or observation or observations) adj3 (timing* or interval* or year* or annual* or month* or periodic* or frequen* or routine*)).ti,ab.
35.	((monitor* or time point* or interval*) adj3 (year* or annual* or month* or periodic* or frequen* or routine*)).ti,ab.
36.	or/30-35
37.	29 and 36
38.	limit 37 to English language
39.	random*.ti,ab.
40.	factorial*.ti,ab.
41.	(crossover* or cross over*).ti,ab.
42.	((doubl* or singl*) adj blind*).ti,ab.
43.	(assign* or allocat* or volunteer* or placebo*).ti,ab.
44.	crossover procedure/
45.	single blind procedure/
46.	randomized controlled trial/
47.	double blind procedure/
48.	or/39-47
49.	systematic review/
50.	meta-analysis/
51.	(meta analy* or metanaly* or metaanaly* or meta regression).ti,ab.
52.	((systematic* or evidence*) adj3 (review* or overview*)).ti,ab.
53.	(reference list* or bibliograph* or hand search* or manual search* or relevant journals).ab.
54.	(search strategy or search criteria or systematic search or study selection or data extraction).ab.

55.	(search* adj4 literature).ab.
56.	(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.
57.	cochrane.jw.
58.	((multiple treatment* or indirect or mixed) adj2 comparison*).ti,ab.
59.	or/49-58
60.	Clinical study/
61.	Observational study/
62.	family study/
63.	longitudinal study/
64.	retrospective study/
65.	prospective study/
66.	cohort analysis/
67.	follow-up/
68.	cohort*.ti,ab.
69.	67 and 68
70.	(cohort adj (study or studies or analys* or data)).ti,ab.
71.	((follow up or observational or uncontrolled or non randomi#ed or epidemiologic*) adj (study or studies or data)).ti,ab.
72.	((longitudinal or retrospective or prospective or cross sectional) and (study or studies or review or analys* or cohort* or data)).ti,ab.
73.	(before adj2 after adj2 (study or studies or data)).ti,ab.
74.	or/60-66,69-73
75.	exp case control study/
76.	case control*.ti,ab.
77.	or/75-76
78.	74 or 77
79.	cross-sectional study/
80.	(cross sectional and (study or studies or review or analys* or cohort* or data)).ti,ab.
81.	or/79-80
82.	74 or 81
83.	74 or 77 or 81
84.	38 and (48 or 59 or 83)

1 **Cochrane Library (Wiley) search terms**

#1.	MeSH descriptor: [Thyroid Diseases] explode all trees
#2.	hyperthyroid*:ti,ab
#3.	hypothyroid*:ti,ab
#4.	thyrotoxicosis:ti,ab
#5.	(thyroid near/3 (swell* or dysfunction* or enlarg* or nodule* or node* or disease* or condition* or disorder*)):ti,ab
#6.	MeSH descriptor: [Goiter] explode all trees
#7.	MeSH descriptor: [Hyperthyroidism] explode all trees
#8.	(hyperthyroid* or thyrotoxicosis):ti,ab
#9.	(toxic near/4 (node* or nodul* or multi?nodul* or goitre or goiter)):ti,ab
#10.	(graves' disease or plummer's disease):ti,ab
#11.	(((thyroid near/4 (swell* or enlarg* or nodule* or node*)) or (goitre* or goiter*)) near/5

	(non-malignan* or nonmalignan* or benign)):ti,ab
#12.	MeSH descriptor: [Thyroid Nodule] explode all trees
#13.	(or #1-#12)
#14.	MeSH descriptor: [Monitoring, Physiologic] explode all trees
#15.	MeSH descriptor: [Patient Outcome Assessment] explode all trees
#16.	monitoring:ti,ab
#17.	((followup or follow-up or check* or evaluat* or appointment* or observation or observations) near/3 (timing* or interval* or year* or annual* or month* or periodic* or frequen* or routine*)):ti,ab
#18.	((monitor* or time point* or interval*) near/3 (year* or annual* or month* or periodic* or frequen* or routine*)):ti,ab
#19.	(or #14-#18)
#20.	#13 and #19

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 11: 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)

1 **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)

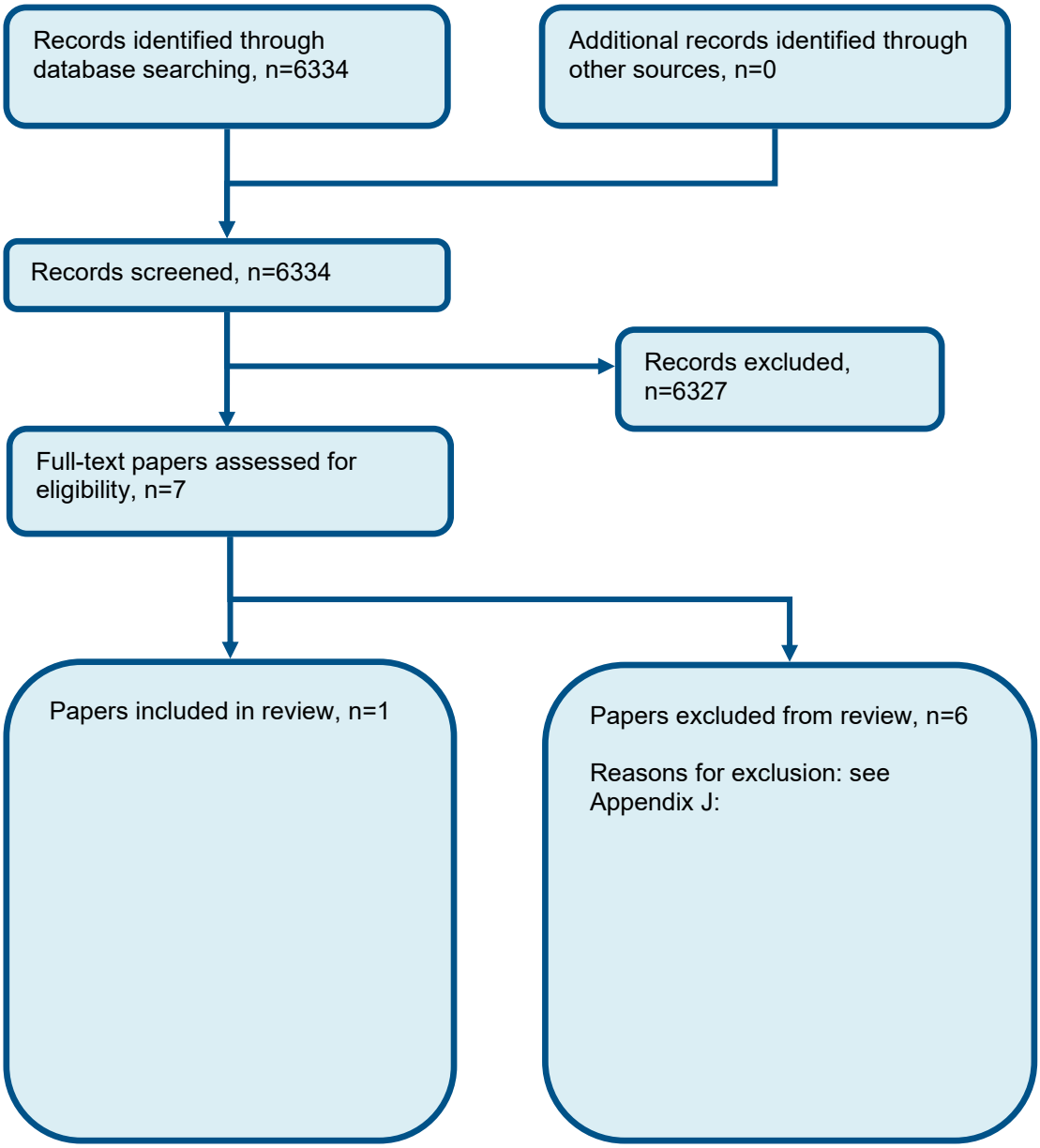
1 **NHS EED and HTA (CRD) search terms**

#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

2

Appendix C: Clinical evidence selection

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



Appendix D: Clinical evidence tables

Study	Samuels 2018[8]
Study type	RCT (Patient randomised; Parallel)
Number of studies (number of participants)	1 (n=138)
Countries and setting	Conducted in USA; Setting: Nil stated
Line of therapy	1st line
Duration of study	Intervention + follow up: 6 months
Method of assessment of guideline condition	Adequate method of assessment/diagnosis
Stratum	Overall
Subgroup analysis within study	Not applicable
Inclusion criteria	Hypothyroid, receiving LT4, adults, previously elevated TSH levels, stable doses for at least 3 months, no acute or chronic illnesses or medications that affect thyroid hormone levels,
Exclusion criteria	Nil stated
Recruitment/selection of patients	Nil stated
Age, gender and ethnicity	Age - Mean (SD): 49.2 (1). Gender (M:F): 9:91. Ethnicity: 92% white
Further population details	
Extra comments	Average duration of LT4 treatment 12 years, average dose 1.44ug/kg.
Indirectness of population	No indirectness
Interventions	(n=46) Intervention 1: T4 only - T4 - high dose start. Taking into account baseline levels, physician started initial dose (usual or adjusted up or down by 25 to 50ug). At 6, 12 and 18 weeks re-assessed and LT4 adjusted if not in target range (adjusted by 12.5 to 50ug). Additional interim visits allowed if not in range at 18 weeks, once in range, no more interim visits. Targeted 0.34 to 2.5 mU/L.. Duration 6 months. Concurrent medication/care: Usual care. Indirectness: No indirectness Further details: 1. T4 dosing: 2. T4 formulations: (n=47) Intervention 2: T4 only - T4 - high dose start. Taking into account baseline levels, physician started initial dose (usual or adjusted up or down by 25 to 50ug). At 6, 12 and 18 weeks re-assessed and LT4 adjusted if not in target range

	(adjusted by 12.5 to 50ug). Additional interim visits allowed if not in range at 18 weeks, once in range, no more interim visits. Targeted 2.51 to 5.6 mU/L.. Duration 6 months. Concurrent medication/care: Usual care. Indirectness: No indirectness Further details: 1. T4 dosing: 2. T4 formulations:
Funding	Academic or government funding

RESULTS (NUMBERS ANALYSED) AND RISK OF BIAS FOR COMPARISON: SPECIFIC TSH TARGET versus NORMAL TSH TARGET

Protocol outcome 1: Quality of life
- Actual outcome: SF-36 mental component summary (0-100, higher is better) at end of follow-up at 6 months; Group 1: mean 41 (SD 6.1); n=46, Group 2: mean 39.8 (SD 7.5); n=47
Risk of bias: All domain - High, Selection - High, Blinding - Low, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: , Reason: ~10% of overall population withdrew, not arm specific; Group 2 Number missing: , Reason: ~10% of overall population withdrew, not arm specific
- Actual outcome: SF-36 physical component summary (0-100, higher is better) at end of follow-up at 6 months; Group 1: mean 48.5 (SD 4.7); n=46, Group 2: mean 50.6 (SD 4.8); n=47
Risk of bias: All domain - High, Selection - High, Blinding - Low, Incomplete outcome data - High, Outcome reporting - Low, Measurement - Low, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: , Reason: ~10% of overall population withdrew, not arm specific; Group 2 Number missing: , Reason: ~10% of overall population withdrew, not arm specific
- Actual outcome: Thyroid specific quality of life (TDQ-AWI, -9 to +3, higher is better) at end of follow-up at 6 months; Group 1: mean -1.6 (SD 1.4); n=46, Group 2: mean -1.4 (SD 1.4); n=47
Risk of bias: All domain - Very high, Selection - High, Blinding - Low, Incomplete outcome data - High, Outcome reporting - Low, Measurement - High, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: , Reason: ~10% of overall population withdrew, not arm specific; Group 2 Number missing: , Reason: ~10% of overall population withdrew, not arm specific
- Actual outcome: SF-36 physical component summary (0-100, higher is better) at end of follow-up at 6 months;
Risk of bias: All domain - ; Indirectness of outcome: No indirectness

Protocol outcome 2: Symptom scores
- Actual outcome: Thyroid symptoms (Billewicz score, -47 to +67, higher is better) at end of follow-up at 6 months; Group 1: mean 2.8 (SD 2.7); n=46, Group 2: mean 4 (SD 3.4); n=47
Risk of bias: All domain - Very high, Selection - High, Blinding - Low, Incomplete outcome data - High, Outcome reporting - Low, Measurement - High, Crossover - Low; Indirectness of outcome: No indirectness ; Group 1 Number missing: , Reason: ~10% of overall population withdrew, not arm specific; Group 2 Number missing: , Reason: ~10% of overall population withdrew, not arm specific

Protocol outcomes not reported by the study	Mortality ; Ischaemic heart disease ; Heart failure ; Arrhythmia ; Osteoporosis ; Impaired cognitive function ; Depression ; Experience of care ; Healthcare contacts ; Growth ; TSH suppression

Appendix E: Forest plots

E.1 Low normal TSH vs high normal TSH

Figure 2: Quality of life, SF-36, mental component

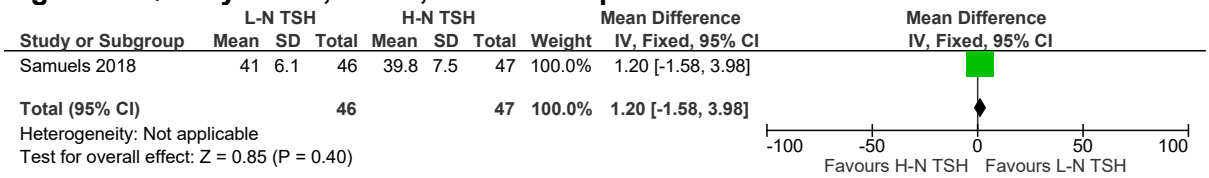


Figure 3: Quality of life, SF-36, physical component

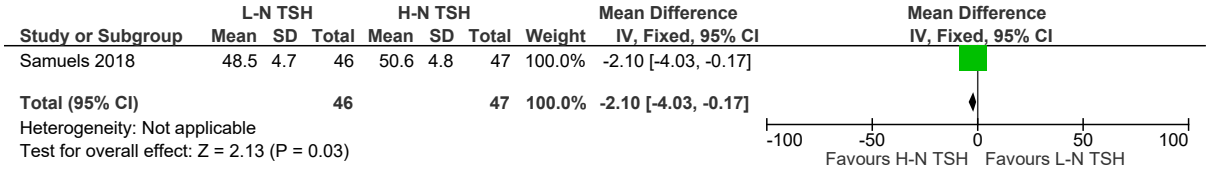


Figure 4: Quality of life, thyroid specific, TDQ-AQ

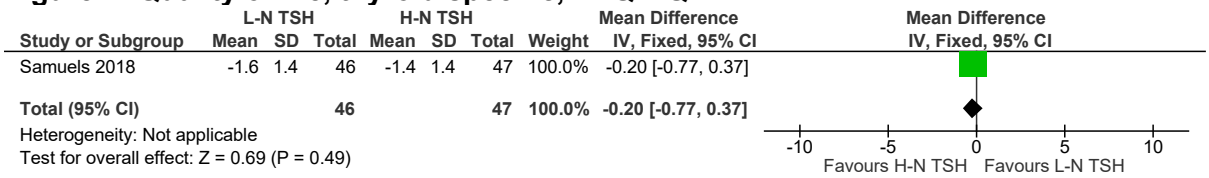
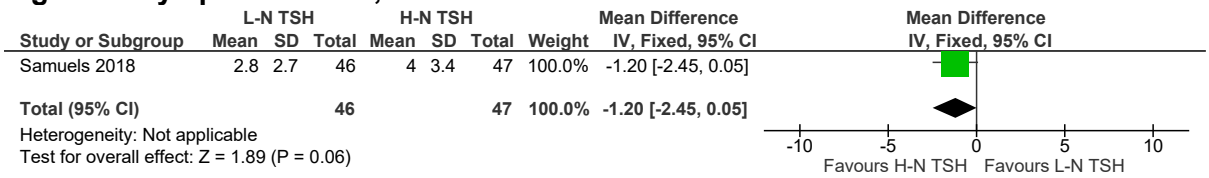


Figure 5: Symptom scores, Billewicz



Appendix F: GRADE tables

Table 12: Clinical evidence profile: low normal vs high normal

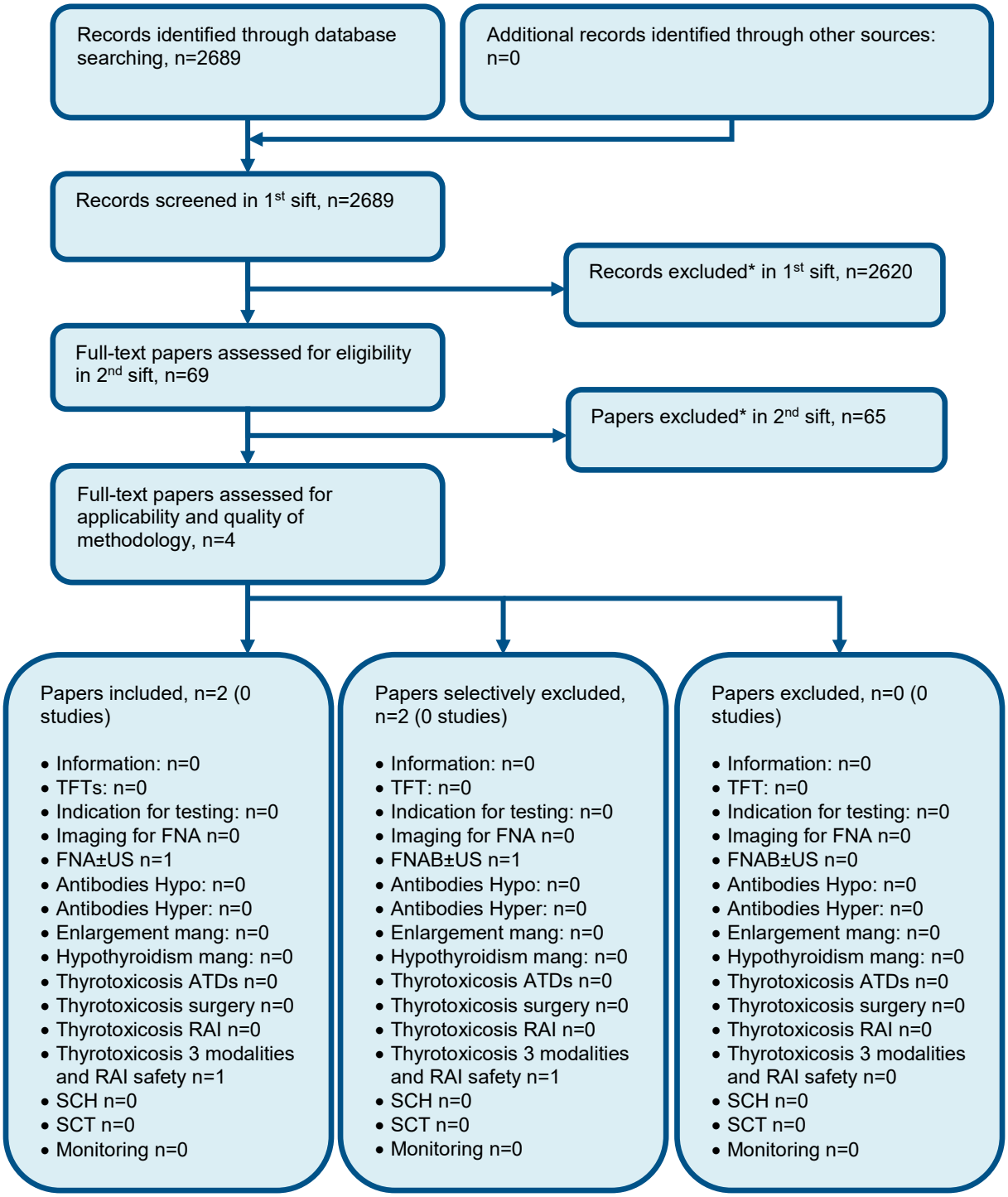
Quality assessment							No of patients		Effect		Quality	Importance
No of studies	Design	Risk of bias	Inconsistency	Indirectness	Imprecision	Other considerations	Low-normal TSH	High-normal TSH	Relative (95% CI)	Absolute		
Quality of life - SF-36, mental component (follow-up 6 months; range of scores: 0-100; Better indicated by higher values)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	46	47	-	MD 1.2 higher (1.58 lower to 3.98 higher)	⊕⊕OO LOW	CRITICAL
Quality of life - SF-36, physical component (follow-up 6 months; range of scores: 0-100; Better indicated by higher values)												
1	randomised trials	serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	46	47	-	MD 2.1 lower (4.03 to 0.17 lower)	⊕⊕OO LOW	CRITICAL
Quality of life - Thyroid specific, TDQ-AW (follow-up 6 months; range of scores: -9-+3; Better indicated by higher values)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	serious[2]	none	46	47	-	MD 0.2 lower (0.77 lower to 0.37 higher)	⊕OOO VERY LOW	CRITICAL
Symptom scores - Billewicz (follow-up 6 months; range of scores: -47-+67; Better indicated by higher values)												
1	randomised trials	very serious[1]	no serious inconsistency	no serious indirectness	very serious[2]	none	46	47	-	MD 1.2 lower (2.45 lower to 0.05 higher)	⊕OOO VERY	CRITICAL

											LOW	

[1] Downgraded by 1 increment if the majority of the evidence was at high risk of bias, and downgraded by 2 increments if the majority of the evidence was at very high risk of bias
[2] Downgraded by 1 increment if the confidence interval crossed one MID or by 2 increments if the confidence interval crossed both MIDs

Appendix G: Health economic evidence selection

Figure 6: 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.

1

Appendix H: Health economic evidence tables

None

Appendix I: Health economic analysis

None

Appendix J: Excluded studies

J.1 Excluded clinical studies

Table 13: Studies excluded from the clinical review

Study	Exclusion reason
Ajmal 2015[1]	Incorrect interventions
Balhara 2011[2]	Incorrect interventions
Helfand 1990[3]	SR, checked for references
Lee 2013[4]	Incorrect interventions
Medici 2016[5]	NRS without adequate adjustment
Phitayakorn 2008[7]	SR, checked for references

J.2 Excluded health economic studies

None