National Institute for Health and Care Excellence

FINAL

Thyroid disease: assessment and management

[C] Thyroid function tests

NICE guideline

Intervention evidence review underpinning recommendations 1.2.8 to 1.2.10 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 All NICE guidance is subject to regular review and may be updated or withdrawn.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 Thyroid function tests

1.1 Review question: Which thyroid function tests should be requested?

1.2 Introduction

A Thyroid function test (TFT) commonly refers to the quantitation of thyroid stimulating hormone (TSH) and circulating thyroid hormones in serum to assess the ability of the thyroid gland to produce and regulate thyroid hormone production. TFTs are used for diagnosis and to monitor treatment of common thyroid gland disorders. These biochemical tests have both high analytical sensitivity and specificity and well established clinical utility. Consequently TFTs are amongst the most widely requested blood tests and are the first line investigation when thyroid disease is suspected. Despite their widespread use and high clinical efficacy, there is still considerable debate as to the optimum testing strategy for both the diagnosis and monitoring of thyroid dysfunction. Measurement of serum TSH concentration is considered as the most effective single marker for the exclusion of primary thyroid dysfunction. TSH is secreted by the pituitary gland in response to circulating thyroid hormone concentration in a classic endocrine feedback loop and can therefore be used as a marker of thyroid status. However there are notable exceptions when serum TSH concentration alone may not accurately reflect thyroid hormone production. This is most frequently encountered when the homeostatic mechanism has been impaired by long standing primary thyroid disease or the pituitary-thyroid axis has not reached equilibrium following changes to thyroid therapy. Secondary thyroid dysfunction due to pituitary disease is a less common confounder of TFTs. Both pituitary deficiency and autonomous TSH production by pituitary adenomas, can lead to erroneous classification of thyroid function if serum TSH is used as a sole biomarker. For these reasons the direct analysis of thyroid hormones is also recommended as an adjunct to TSH testing. Most UK labs would advocate the measurement of free thyroid hormones rather than total thyroid hormone (thyroid hormone that is not bound to thyroid carrier proteins) as the best marker for biologically active hormone. This avoids situations where variation in thyroid hormone binding proteins rather than thyroid function *per se* is responsible for abnormal test results.

Various TFT protocols are used in current clinical practice with no particular procedure dominating. Common protocols include an initial TSH measurement with thyroid hormone analysis (either free thyroxine - FT4 or free triiodothyronine - FT3) cascaded when TSH is outside the established reference interval, simultaneous analysis of both TSH and FT4 or open access to either test depending on requestor preference. Many laboratories offer FT3 testing but this is not universally, available. Of the laboratories that offer FT3 analysis, some provide open access where others will use FT3 instead of FT4 to confirm the diagnosis of hyperthyroidism when TSH is suppressed. Other indications for FT3 include exclusion of T3 toxicosis (when TSH is suppressed but FT4 is within the reference interval), the detection of non-thyroidal illness, to adjust levothyroxine dosage if TSH remains elevated in the presence of FT4 concentrations above the reference interval or more contentiously to direct thyroid replacement therapy in patients with TSH within the reference interval but with persistence of hypothyroid symptoms.

The latest authoritative UK guidance on TFT came from the British Thyroid Association in 2006. This guide recommends using both TSH and FT4 at the first investigation of thyroid disease, at the initial optimisation of therapy in both hyper or hypothyroidism, in patients treated with thionamides, in pregnancy or when pituitary disease is suspected. In other situations, such as longer term follow-up, the testing of TSH alone may be appropriate. The authors recognised the need for further studies in particular the use of TFTs as a screening test and for diagnosing and treating subclinical disease.

Given the widespread use of TFTs the addition of FT4 and FT3 to the TFT represents a considerable financial burden to the UK health economy. The purpose of this review is to establish if objective evidence is available to support the premise that the use of more complex thyroid function testing strategies leads to demonstrable health benefit.

1.3 PICO table

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

Table 1: PICO characteristics of review question

Population	People being investigated for hypothyroidism, hyperthyroidism or thyroid enlargement
Interventions	TSH +/- FT4 +/- FT3 +/- tissue markers of thyroid hormone action (for example reverse T3, sex hormone binding globulin (SHBG), cholesterol)
Comparisons	Any of the combinations of diagnostic tests above compared with any other
Outcomes	**Critical outcomes** • Mortality (dichotomous, ≥1 year) • Quality of life (continuous) **Important outcomes** • Symptom scores (continuous) • Patient/family/carer experience (continuous) • Healthcare contacts (rates/dichotomous) • Number of people receiving treatment (dichotomous) • Growth (children and young people only, continuous) • Neurodevelopment (children and young people only, continuous)
Study design	RCTs, non-randomised cohort studies to be considered if adjusted for key confounders (age, co-existing conditions) and insufficient RCTs evidence found for committee decision making

1.4 Clinical evidence

1.4.1 Included studies

As per the protocol, given the lack of randomised controlled trials, non-randomised studies were also considered for inclusion for this evidence review. No relevant clinical studies comparing any thyroid function testing strategy with any other were identified.

See also the study selection flow chart in Appendix C:.

1.4.2 Excluded studies

See the excluded studies list in Appendix G:.

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 2: UK costs of Tests

Tests	*A*	B	C	D	E	Median
TSH	£5.65	£2.94	£2.15	£0.86	£1.79	£2.15
TSH+FT4	£11.05	£6.31	£4.41	£1.8	£3.72	£4.41
FT3	£4.96	£3.51	£3.12	£1.86	£2.04	£3.12
FT4	- (a)	£3.37	£2.26	£0.94	£1.93	£2.10

Source: Costs obtained from committee members working in different hospitals
(a) This hospital does not do FT4 tests alone
(b) Costs quoted include reagent, any consumables and staff pay

Tests	*Unit costs*
Phlebotomy (a)	£3.04

Source: NHS reference cost 2016-17
(a) Currency code DAPS08

1.6 Evidence statements

1.6.1 Clinical evidence statements

- No relevant published evidence was identified.

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 committee considered mortality and quality of life to be critical outcomes. The committee considered symptom scores, patient/family/carer experience, healthcare contacts, the number of people receiving treatment, growth and neurodevelopment to be important outcomes.

No evidence was identified for this review.

1.7.1.2 The quality of the evidence

No evidence was identified for this review.

1.7.1.3 Benefits and harms

In the first investigation of possible thyroid disease the committee agreed based on their experience that a testing strategy of only using TSH with a FT4 measurement if TSH is above the reference range would be appropriate for the majority of adults. If TSH is below the reference range both FT4 and FT3 will be necessary as the results of both FT4 and FT3 will dictate decisions around which type of treatment the thyrotoxicosis may require. If TSH is normal, abnormal thyroid FT4 is unlikely based on the committee's experience. If TSH is abnormal, it is important to determine whether a person has clinical or subclinical thyroid dysfunction (in other words if the abnormality is limited only to the TSH or not) as this impacts management decisions.

The committee agreed there may be some situations in which both tests could be ordered simultaneously, particularly if there is suspicion of secondary thyroid disease. This would apply generally for children as the aetiology of thyroid disease in children tends to include more secondary causes and fewer primary or autoimmune causes. The committee also noted that in children, it is often more distressing and challenging to obtain multiple blood samples and therefore strategies should aim to limit samples.

1.7.2 Cost effectiveness and resource use

There was no economic evidence identified in this review comparing thyroid function testing strategies, for people being investigated for hypothyroidism, hyperthyroidism, or thyroid enlargement. Unit costs of thyroid function tests were presented to the committee. The committee made consensus recommendations on the groups of people that different tests may be appropriate for.

The costs of different blood tests were collected from committee members at 5 different hospitals, the costs reported included cost of reagents, consumables, and staff pay. There were large variations between the costs reported and the committee agreed that this was due to differences in the way staff pay is calculated in the hospitals cost data was collected from. Due to the skewed data, a median was calculated (TSH £2.15, FT3 £3.12 and FT4 £2.10) and presented to the committee as estimates of the different costs of blood tests.

The committee highlighted that the reported costs did not include phlebotomy costs and therefore in instances when tests are ordered separately an additional cost of £3.04 will be incurred (NHS reference cost 2016-17, currency code DAPS08), each time. The committee were aware that some laboratories performing TFTs would conduct an additional test (FT4 and/ or TF3) on the same blood sample (cascade), if they detect an abnormal TSH. This approach is not being followed everywhere but it can reduce unnecessary testing and avoid incurring the additional phlebotomy costs. The committee noted that TSH should be considered, as a first line blood test for anyone with a potential thyroid dysfunction or enlargement, and depending on the results, via cascade, FT4 then FT3 might be required. It was noted that this cascade approach, where a small population would require further tests depending on the results from the initial test, is likely to be cost saving to the NHS. It was also agreed that measuring both TSH and FT4 should be considered in people with suspected secondary thyroid disease, this would help speed the diagnosis and people can be treated for their condition sooner.

In addition, the committee acknowledged that repeat testing is often done too soon after diagnosis and that a recommendation considering repeat testing at least 6 weeks after diagnosis can help avoid misinterpretation of test results. This can be cost saving to the NHS and beneficial to the patient as misinterpreted test results can lead to unnecessary treatment or dose adjustments.

The recommendations broadly reflect current practice, although not all laboratories currently follow the cascading approach to testing. Where FT4 is currently a routine test for thyroid dysfunction, cascading will reduce NHS costs by avoiding extra tests for people with a TSH within the reference range. In areas where FT3 is not currently being measured, cascading will mean a cost increase. But this will be offset by the benefits of correctly diagnosing and managing thyrotoxicosis.

References

1. Brochmann H, Bjoro T, Gaarder PI, Hanson F, Frey HM. Prevalence of thyroid dysfunction in elderly subjects. A randomized study in a Norwegian rural community (Naeroy). Acta Endocrinologica. 1988; 117(1):7-12

2. Feldkamp CS, Carey JL. An algorithmic approach to thyroid function testing in a managed care setting. 3-year experience. American Journal of Clinical Pathology. 1996; 105(1):11-6

3. Henze M, Brown SJ, Hadlow NC, Walsh JP. Rationalizing Thyroid Function Testing: Which TSH Cutoffs Are Optimal for Testing Free T4? Journal of Clinical Endocrinology and Metabolism. 2017; 102(11):4235-4241

4. Koulouri O, Gurnell M. How to interpret thyroid function tests. Clinical Medicine. 2013; 13(3):282-6

5. 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

6. Notas G, Kampa M, Malliaraki N, Petrodaskalaki M, Papavasileiou S, Castanas E. Implementation of thyroid function tests algorithms by clinical laboratories: A four-year experience of good clinical and diagnostic practice in a tertiary hospital in Greece. European Journal of Internal Medicine. 2018; 54:81-86

7. Snabboon T, Sridama V, Sunthornyothin S, Suwanwalaikorn S, Vongthavaravat V. A more appropriate algorithm of thyroid function test in diagnosis of hyperthyroidism for Thai patients. Journal of the Medical Association of Thailand. 2004; 87 (Suppl 2):S19-21

Appendices

Appendix A: Review protocols

Table 3:

ID	Field	Content
I	Review question	Which thyroid function tests should be requested?
II	Type of review question	Diagnostic intervention review (test and treat approach) 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 clinical and cost effectiveness of using various thyroid function tests in the investigation of thyroid disease
IV	Eligibility criteria – population / disease / condition / issue / domain	• People being investigated for hypothyroidism, hyperthyroidism or thyroid enlargement
V	Eligibility criteria – intervention(s) / exposure(s) / prognostic factor(s)	• TSH +/- T4 +/- T3 +/- tissue markers of thyroid hormone action (for example reverse T3, sex hormone binding globulin (SHBG), cholesterol) • Serum tests only • Free T3/T4 only Each arm must provide similar interventions to people diagnosed with thyroid disease regardless of test used for diagnosis
VI	Eligibility criteria – comparator(s) / control or reference (gold) standard	• Any of the combinations of diagnostic tests above compared with any other
VII	Outcomes and prioritisation	**Critical outcomes** • Mortality (dichotomous, ≥1 year) • Quality of life (continuous) **Important outcomes** • Symptom scores (continuous) • Patient/family/carer experience (continuous) • Healthcare contacts (rates/dichotomous) • Number of people receiving treatment (dichotomous) • Growth (children and young people only, continuous) • Neurodevelopment (children and young people only, continuous)
VIII	Eligibility criteria – study design	• RCTs • Non-randomised cohort studies to be considered if adjusted for key confounders (age, co-existing conditions) and insufficient RCTs evidence found for committee decision making • Minimum follow-up 3 months

IX	Other inclusion exclusion criteria	• None specified
X	Proposed sensitivity / subgroup analysis, or meta-regression	**Stratifications** • People being investigated for hypothyroidism • People being investigated for hyperthyroidism • People being investigated for thyroid enlargement • Age – young children (0-4), children and young people (4-18), adults (>18-65), older adults (>65)
XI	Selection process – duplicate screening / selection / analysis	No duplicate screening was deemed necessary for this question, 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:.
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	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	For details please see the separate Methods report for this guideline.

	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 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

Table 4: 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 D: below.
Review	Studies not meeting any of the search criteria above will be excluded. Studies

strategy

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).[5]

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 5: 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 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.	thyroid function tests/
27.	Thyroxine/ or Triiodothyronine/ or thyrotropin/
28.	(T3 or T4 or FT3 or FT4 or T3RU or T3 uptake or T3 resin uptake or TSH or hTSH).ti,ab.
29.	(thyroxine or triidothyronine).ti,ab.
30.	(Thyroid stimulating hormone or thyroid-stimulating hormone or thyrotropin or thyrotropic hormone or hormone).ti,ab.
31.	or/27-30
32.	25 and 31
33.	limit 32 to English language
34.	randomized controlled trial.pt.
35.	controlled clinical trial.pt.
36.	randomi#ed.ti,ab.
37.	placebo.ab.
38.	randomly.ti,ab.
39.	Clinical Trials as topic.sh.
40.	trial.ti.
41.	or/34-40
42.	Meta-Analysis/
43.	exp Meta-Analysis as Topic/
44.	(meta analy* or metanaly* or metaanaly* or meta regression).ti,ab.
45.	((systematic* or evidence*) adj3 (review* or overview*)).ti,ab.
46.	(reference list* or bibliograph* or hand search* or manual search* or relevant journals).ab.
47.	(search strategy or search criteria or systematic search or study selection or data extraction).ab.
48.	(search* adj4 literature).ab.
49.	(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.
50.	cochrane.jw.
51.	((multiple treatment* or indirect or mixed) adj2 comparison*).ti,ab.

52.	or/42-51
53.	exp "sensitivity and specificity"/
54.	(sensitivity or specificity).ti,ab.
55.	((pre test or pretest or post test) adj probability).ti,ab.
56.	(predictive value* or PPV or NPV).ti,ab.
57.	likelihood ratio*.ti,ab.
58.	likelihood function/
59.	((area under adj4 curve) or AUC).ti,ab.
60.	(receive* operat* characteristic* or receive* operat* curve* or ROC curve*).ti,ab.
61.	(diagnos* adj3 (performance* or accurac* or utilit* or value* or efficien* or effectiveness)).ti,ab.
62.	gold standard.ab.
63.	or/53-62
64.	Epidemiologic studies/
65.	Observational study/
66.	exp Cohort studies/
67.	(cohort adj (study or studies or analys* or data)).ti,ab.
68.	((follow up or observational or uncontrolled or non randomi#ed or epidemiologic*) adj (study or studies or data)).ti,ab.
69.	((longitudinal or retrospective or prospective or cross sectional) and (study or studies or review or analys* or cohort* or data)).ti,ab.
70.	Controlled Before-After Studies/
71.	Historically Controlled Study/
72.	Interrupted Time Series Analysis/
73.	(before adj2 after adj2 (study or studies or data)).ti,ab.
74.	or/64-73
75.	exp case control study/
76.	case control*.ti,ab.
77.	or/75-76
78.	74 or 77
79.	Cross-sectional studies/
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.	33 and (41 or 52 or 63 or 83)

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.	thyroid function test/
25.	Thyroxine/ or Triiodothyronine/ or thyrotropin/
26.	(T3 or T4 or FT3 or FT4 or T3RU or T3 uptake or T3 resin uptake or TSH or hTSH).ti,ab.
27.	(thyroxine or triidothyronine).ti,ab.
28.	(Thyroid stimulating hormone or thyroid-stimulating hormone or thyrotropin or thyrotropic hormone or hormone).ti,ab.
29.	or/24-28
30.	23 and 29
31.	limit 30 to English language
32.	random*.ti,ab.
33.	factorial*.ti,ab.
34.	(crossover* or cross over*).ti,ab.
35.	((doubl* or singl*) adj blind*).ti,ab.
36.	(assign* or allocat* or volunteer* or placebo*).ti,ab.
37.	crossover procedure/
38.	single blind procedure/
39.	randomized controlled trial/
40.	double blind procedure/
41.	or/32-40
42.	systematic review/
43.	meta-analysis/
44.	(meta analy* or metanaly* or metaanaly* or meta regression).ti,ab.
45.	((systematic* or evidence*) adj3 (review* or overview*)).ti,ab.
46.	(reference list* or bibliograph* or hand search* or manual search* or relevant journals).ab.
47.	(search strategy or search criteria or systematic search or study selection or data extraction).ab.
48.	(search* adj4 literature).ab.
49.	(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.
50.	cochrane.jw.

51.	((multiple treatment* or indirect or mixed) adj2 comparison*).ti,ab.
52.	or/42-51
53.	exp "sensitivity and specificity"/
54.	(sensitivity or specificity).ti,ab.
55.	((pre test or pretest or post test) adj probability).ti,ab.
56.	(predictive value* or PPV or NPV).ti,ab.
57.	likelihood ratio*.ti,ab.
58.	((area under adj4 curve) or AUC).ti,ab.
59.	(receive* operat* characteristic* or receive* operat* curve* or ROC curve*).ti,ab.
60.	(diagnos* adj3 (performance* or accurac* or utilit* or value* or efficien* or effectiveness)).ti,ab.
61.	diagnostic accuracy/
62.	diagnostic test accuracy study/
63.	gold standard.ab.
64.	or/53-63
65.	Clinical study/
66.	Observational study/
67.	family study/
68.	longitudinal study/
69.	retrospective study/
70.	prospective study/
71.	cohort analysis/
72.	follow-up/
73.	cohort*.ti,ab.
74.	72 and 73
75.	(cohort adj (study or studies or analys* or data)).ti,ab.
76.	((follow up or observational or uncontrolled or non randomi#ed or epidemiologic*) adj (study or studies or data)).ti,ab.
77.	((longitudinal or retrospective or prospective or cross sectional) and (study or studies or review or analys* or cohort* or data)).ti,ab.
78.	(before adj2 after adj2 (study or studies or data)).ti,ab.
79.	or/65-71,74-78
80.	exp case control study/
81.	case control*.ti,ab.
82.	or/80-81
83.	79 or 82
84.	cross-sectional study/
85.	(cross sectional and (study or studies or review or analys* or cohort* or data)).ti,ab.
86.	or/84-85
87.	79 or 86
88.	79 or 82 or 86
89.	31 and (41 or 52 or 64 or 88)

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.	(or #1-#5)
#7.	MeSH descriptor: [Thyroid function tests] explode all trees
#8.	MeSH descriptor: [Thyroxine] explode all trees
#9.	MeSH descriptor: [Triiodothyronine] explode all trees
#10.	MeSH descriptor: [Thyrotropin] explode all trees
#11.	(T3 or T4 or FT3 or FT4 or T3RU or T3 uptake or T3 resin uptake or TSH or hTSH):ti,ab
#12.	(thyroxine or triidothyronine):ti,ab
#13.	(Thyroid stimulating hormone or thyroid-stimulating hormone or thyrotropin or thyrotropic hormone or hormone):ti,ab
#14.	(or #7-#13)
#15.	#6 and #14

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 6: 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/

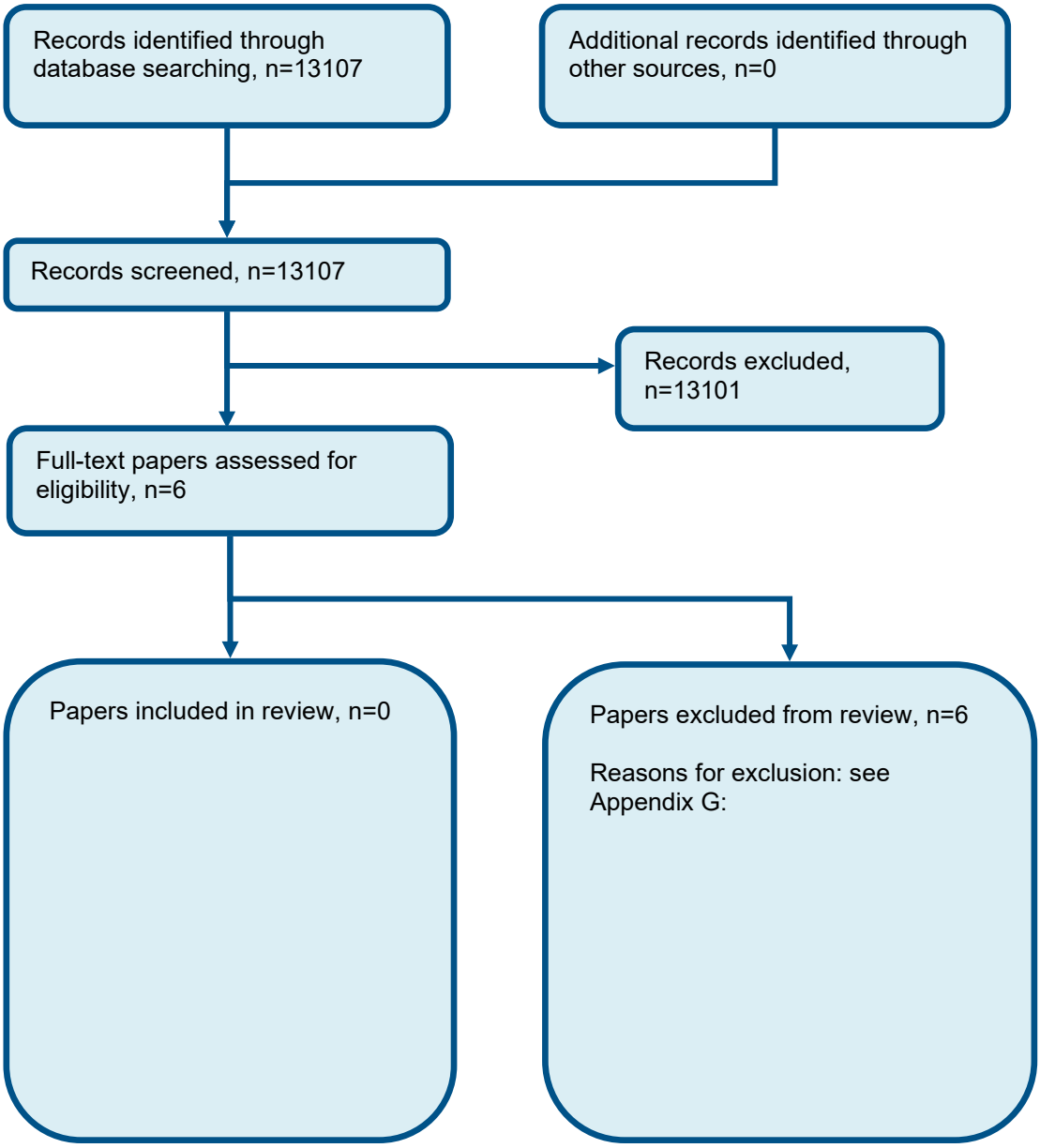
55.	sickness impact profile/
56.	(quality adj2 (wellbeing or well being)).ti,ab.
57.	sickness impact profile.ti,ab.
58.	disability adjusted life.ti,ab.
59.	(qal* or qtime* or qwb* or daly*).ti,ab.
60.	(euroqol* or eq5d* or eq 5*).ti,ab.
61.	(qol* or hql* or hqol* or h qol* or hrqol* or hr qol*).ti,ab.
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

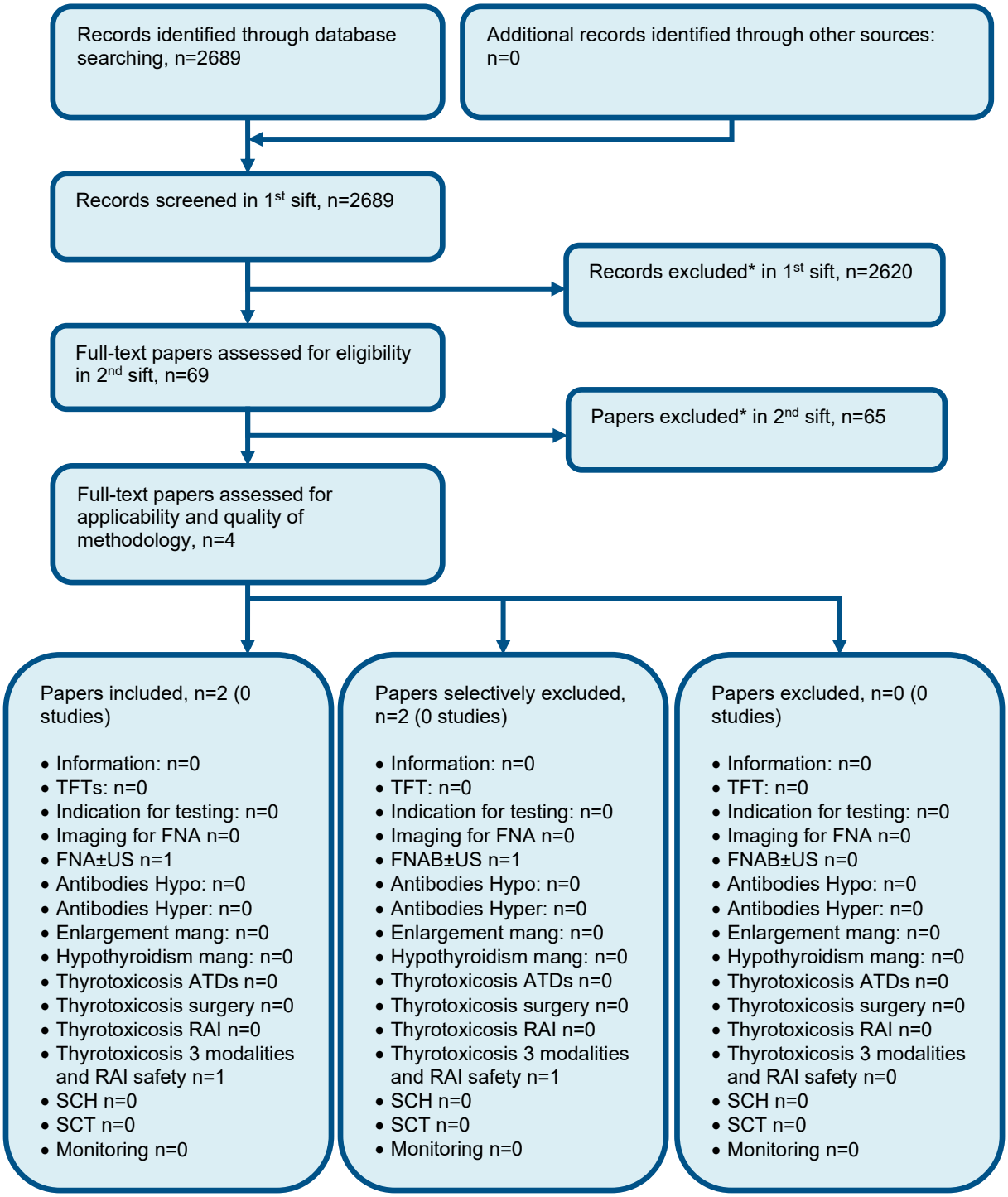
Appendix C: Clinical evidence selection

Figure 1: Flow chart of clinical study selection for the review of thyroid function tests



Appendix D: Health economic evidence selection

Figure 2: 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 E: Health economic evidence tables

None

Appendix F:Health economic analysis

None

Appendix G: Excluded studies

G.1 Excluded clinical studies

Table 7: Studies excluded from the clinical review

Reference	Reason for exclusion
Feldkamp 1996[2]	Incorrect study design: non-randomised study not meeting protocol.
Koulouri 2013[4]	Incorrect study design: non-randomised descriptive study.
Notas 2018[6]	Wrong population
Snabboon 2004[7]	Incorrect study design: non-randomised study; no usable outcomes meeting protocol.
Brochmann 1988[1]	Incorrect study design: epidemiological study with no usable outcomes.
Henze 2017[3]	Incorrect study design: non-randomised study not meeting protocol.

G.2 Excluded health economic studies

None