Health Technology Assessment (HTA)

HTA Protocol

Title	Thyroid function tests for the diagnosis of suspected primary or secondary thyroid dysfunction
Author/Affiliation	**Remziye Zaim, Ingelin Kvamme, Matthijs Versteegh** Institute for Medical Technology Assessment, Erasmus University, Rotterdam, The Netherlands **Anouk Oordt, Femke van Kessel, Eveline Bunge** Pallas Health Research and Consultancy, Rotterdam, The Netherlands
Technology	(free) Triiodothyronine ((f)T3), (free) Thyroxine ((f)T4), and Thyroid Stimulating Hormone (TSH) testing
Date	29 September 2021

Executive Summary

Thyroid function tests are used for the diagnosis and monitoring of patients with thyroid disorders. The most common thyroid function tests include thyroid stimulating hormone (TSH) tests, total and/or free thyroxine (T4/fT4) tests and total, free and/or reverse triiodothyroxine (T3, fT3, rT3) tests. Depending on the clinical situation, the choice of test(s) can be modified. For the diagnosis of suspected primary or secondary thyroid dysfunction two typical test approaches include a one-step test approach (i.e., measurement of TSH and (f)T4/(f)T3 simultaneously) and a two-step test approach (i.e., testing TSH first, only followed by the measurement of (f)T4/(f)T3 when TSH is out of the reference range). Although most guidelines recommend a two-step test approach for the diagnosis of suspected primary or secondary thyroid dysfunction, it is observed that TSH and (f)T4/(f)T3 are often measured simultaneously (one-step approach). The focus of this HTA is to evaluate the two thyroid function test approaches for the diagnosis of patients with suspected primary or secondary thyroid dysfunction.

For the clinical review, a systematic literature search of the Pubmed (MEDLINE) and Embase.com databases will be conducted adhering to international methodological standards. Comparative non-randomised studies and randomised controlled trials including adults with suspected primary or secondary thyroid dysfunction who are tested with a one-step approach will be included. We will search worldwide, in English, German, French and Italian with the publication period limited from 1990 to 2021. Studies on thyroid disorder screening and monitoring, and those including pregnant women will be excluded. The options for clinically relevant data merging/stratification will be investigated after data-extraction. Outcomes for which there are no possibility to calculate pooled estimates will be presented narratively using summary tables and accompanying text.

For the economic review, costs and budget impact searches will follow the principles of the clinical systematic literature search. After applying the quality control measures, data from the included articles found in the reviewed literature will be summarised. Data synthesis will be done using descriptive comparisons of the study question, methods and results. Following the data collection, a budget impact model will be populated with the obtained parameters. Subgroup analyses will be performed, where possible, to differentiate budget impact by pre-defined patient groups.

Legal, social, ethical, and organisational issues addressed in the studies included in the clinical, costs and budget impact systematic literature searches will be extracted and narratively summarised. In addition, grey literature searches will be conducted on these HTA domains.

Zusammenfassung

Schilddrüsenhormontests werden zur Diagnose und zur Überwachung von Patientinnen und Patienten mit Schilddrüsenerkrankungen eingesetzt. Zu den gebräuchlichsten Schilddrüsenhormontests gehören Tests auf das Thyreoidea-stimulierende Hormon (TSH), das gesamte und/oder freie Thyroxin (T4/fT4) sowie das gesamte, freie und/oder reverse Trijodthyroxin (T3, fT3, rT3). Je nach klinischer Situation kann sich die Wahl der Tests ändern. Für die Diagnose einer vermuteten primären oder sekundären Schilddrüsenfunktionsstörung gibt es zwei typische Testansätze: einen einstufigen Testansatz (d. h. gleichzeitige Messung von TSH und (f)T4/(f)T3) und einen zweistufigen Testansatz (d. h. zunächst TSH-Test und erst dann Messung von (f)T4/(f)T3, wenn TSH ausserhalb des Referenzbereichs liegt). Obwohl die meisten Richtlinien ein zweistufiges Vorgehen für die Diagnose einer vermuteten primären oder sekundären Schilddrüsenfunktionsstörung empfehlen, ist zu beobachten, dass TSH und (f)T4/(f)T3 häufig gleichzeitig gemessen werden (einstufiges Vorgehen). Der Schwerpunkt dieses HTA liegt auf der Bewertung der beiden Schilddrüsenhormontest-Ansätze für die Diagnose bei Patientinnen und Patienten mit Verdacht auf eine primäre oder sekundäre Schilddrüsenfunktionsstörung.

Für die klinische Prüfung wird eine systematische Literaturrecherche in den Datenbanken Pubmed (MEDLINE) und Embase.com unter Einhaltung internationaler methodischer Standards durchgeführt. Einbezogen werden vergleichende nicht-randomisierte Studien und randomisierte kontrollierte Studien, die mit einem einstufigen Ansatz getestete Erwachsene mit Verdacht auf eine primäre oder sekundäre Schilddrüsenfunktionsstörung einschliessen. Die Recherche erfolgt weltweit in den Sprachen Englisch, Deutsch, Französisch und Italienisch, wobei der Veröffentlichungszeitraum auf die Jahre 1990 bis 2021 begrenzt ist. Studien zum Screening und zur Überwachung von Schilddrüsenerkrankungen sowie Studien, die schwangere Frauen einschliessen, werden ausgeschlossen. Nach der Datenextraktion werden die Möglichkeiten zur Zusammenführung/Stratifizierung klinisch relevanter Daten untersucht. Ergebnisse, für die es keine Möglichkeit zur Berechnung gepoolter Schätzungen gibt, werden in zusammenfassenden Tabellen und im Begleittext narrativ dargestellt.

Für die Wirtschaftlichkeitsprüfung erfolgt die Recherche zu den Kosten und den Budgetauswirkungen nach den Grundsätzen der klinischen systematischen Literaturrecherche.
Nach der Qualitätskontrolle werden die Daten der einbezogenen Artikel aus der Literatursuche zusammengefasst. Die Datensynthese erfolgt anhand deskriptiver Vergleiche der Studienfragen, Methoden und Ergebnisse. Im Anschluss an die Datenerhebung wird ein Budgetauswirkungsmodell mit den ermittelten Parametern entwickelt. Wo möglich, werden Untergruppenanalysen durchgeführt, um die Budgetauswirkungen nach vordefinierten Patientengruppen zu differenzieren.

Rechtliche, soziale, ethische und organisatorische Fragen, welche in den einbezogenen Studien zu den klinischen Aspekten, Kosten und Budgetauswirkungen behandelt werden, werden extrahiert und narrativ zusammengefasst. Zudem werden zu diesen HTA-Bereichen Recherchen in grauer Literatur durchgeführt.

Résumé

Les tests de la fonction thyroïdienne sont utilisés pour diagnostiquer et suivre les patients présentant des troubles de la thyroïde. Les plus courants sont, notamment, les tests de la thyréostimuline (TSH), de la thyroxine totale et/ou libre (T4/fT4) et de la triiodothyronine totale, libre et/ou inverse (T3, fT3, rT3). Le choix du ou des tests peut être modifié en fonction de la situation clinique. Il existe deux approches courantes de diagnostic lorsqu'un dysfonctionnement thyroïdien primaire ou secondaire est soupçonné : une approche en une étape (mesure simultanée de la TSH et de la (f)T4/(f)T3) et une autre en deux étapes (tester d'abord la TSH, puis, seulement si le taux de TSH se situe en dehors de l'intervalle de référence, mesurer la (f)T4/(f)T3). Bien que la plupart des lignes directrices recommandent une approche en deux étapes, on observe que la TSH et la (f)T4/(f)T3 sont souvent mesurées en même temps (approche en une étape). La présente ETS se concentre sur l'évaluation des deux approches.

Pour l'évaluation clinique, une recherche systématique de la littérature sera effectuée dans les bases de données Pubmed (MEDLINE) et Embase.com, en suivant les normes méthodologiques internationales. Elle inclura les études comparatives non randomisées et les essais cliniques randomisés portant sur des adultes chez lesquels un dysfonctionnement thyroïdien primaire ou secondaire est soupçonné et qui passent un test selon une approche en une étape. Nous rechercherons les articles publiés entre 1990 et 2021 dans les revues du monde entier en anglais, en allemand, en français et en italien. Les études sur le dépistage et le suivi des troubles thyroïdiens seront exclues, de même que celles incluant des femmes enceintes. S'agissant de la fusion/stratification des données, les options pertinentes d'un point de vue clinique seront examinées après l'extraction des données. Les résultats pour lesquels des estimations groupées ne peuvent être calculées seront présentés de manière narrative, au moyen de tableaux récapitulatifs légendés.

Pour l'évaluation économique, des recherches concernant l'impact sur les coûts et les budgets seront menées selon les principes de la recherche systématique de la littérature clinique. Une fois les mesures de contrôle de la qualité appliquées, les données issues des articles inclus trouvés dans des revues évaluées par les pairs seront synthétisées. Cette synthèse sera effectuée par le biais de comparaisons descriptives de la problématique, de méthodes et des résultats de chaque étude. Après collecte des données, un modèle d'impact budgétaire sera alimenté avec les paramètres obtenus.

Des analyses par sous-groupes seront effectuées, dans la mesure du possible, afin de différencier l'impact budgétaire entre des catégories de patients prédéfinies.

Les questions juridiques, sociales, éthiques et organisationnelles abordées par les études incluses dans les recherches systématiques de la littérature portant sur les aspects cliniques, les coûts et l'impact budgétaire seront extraites et synthétisées de manière narrative. En outre, il sera procédé à des recherches de la littérature grise dans ces domaines d'ETS.

Sintesi

Gli esami della funzionalità tiroidea sono impiegati per la diagnosi e il monitoraggio dei disturbi tiroidei. Gli esami più comuni comprendono gli esami dell'ormone tireostimolante (TSH), della tiroxina totale e/o libera (T4/fT4) e della triiodotironina totale, libera e/o inversa (T3, fT3, rT3). La scelta dell'esame o degli esami varia a seconda del quadro clinico. Per la diagnosi di una sospetta disfunzione tiroidea primaria o secondaria vi sono due approcci di esami tipici: un approccio a una fase (misurazione simultanea di TSH e [f]T4/[f]T3) e uno a due fasi (esame del TSH per primo, seguito dalla misurazione della [f]T4/[f]T3 solo se il TSH non rientra nei valori di riferimento). Sebbene per la diagnosi di una sospetta disfunzione tiroidea primaria o secondaria la maggior parte delle linee guida raccomandi l'approccio a due fasi, si osserva che il TSH e la (f)T4/(f)T3 spesso sono misurati simultaneamente (approccio a una fase). Lo scopo del presente *Health Technology Assessment* (HTA) è valutare i due approcci per gli esami della funzionalità tiroidea finalizzati alla diagnosi dei pazienti con una sospetta disfunzione tiroidea primaria o secondaria.

Per la revisione clinica sarà condotto uno studio sistematico della letteratura nelle banche dati di Pubmed (MEDLINE) ed Embase.com, in linea con gli standard metodologici internazionali. Saranno inclusi studi comparativi non randomizzati e studi controllati randomizzati che comprendono adulti con una sospetta disfunzione tiroidea primaria o secondaria esaminati con l'approccio a una fase.

Lo studio, condotto su scala mondiale, prenderà in esame riviste in lingua inglese, tedesca, francese e italiana con un periodo di pubblicazione limitato dal 1990 al 2021. Non saranno presi in considerazione gli studi sullo screening e sul monitoraggio del disturbo tiroideo e quelli che includono le donne incinte. Le possibilità di raggruppare/stratificare i dati clinicamente rilevanti saranno esaminate una volta effettuata l'estrazione dei dati. I risultati per i quali risulterà impossibile calcolare stime aggregate saranno illustrati in modo discorsivo mediante tabelle riassuntive e un testo di accompagnamento.

Per la valutazione economica, gli studi relativi a costi e impatto sul bilancio avverranno sulla scorta dei principi dello studio sistematico della letteratura clinica. Dopo aver compiuto le misurazioni di controllo della qualità, i dati degli articoli presi in considerazione trovati nella letteratura peer review

saranno riassunti. La sintesi dei dati sarà effettuata confrontando in modo descrittivo il quesito, i metodi e i risultati dello studio. Dopo aver raccolto i dati, i parametri ottenuti confluiranno in un modello di impatto sul bilancio. Se possibile, saranno condotte analisi di sottogruppo per distinguere l'impatto sul bilancio per gruppi di pazienti predefiniti.

Le questioni legali, sociali, etiche e organizzative affrontate negli studi inclusi nelle ricerche sistematiche della letteratura clinica, dei costi e dell'impatto sul bilancio saranno estrapolate e riassunte in modo discorsivo. Infine, su questi ambiti dell'HTA saranno condotti studi della letteratura grigia.

Table of contents

Abbreviations and acronyms

CHF	Swiss Franc
FOPH	Federal Office of Public Health
fT3	free Triiodothyronine
fT4	free Thyroxine
GRADE	Grading of Recommendations, Assessment, Development, and Evaluations
HTA	Health Technology Assessment
LC-MS/MS	Liquid Chromatography with tandem Mass Spectrometry
NICE	National Institute for Health and Care Excellence
NHS EED	National Health Service Economic Evaluation Database
OECD	Organisation for Economic Cooperation and Development
PICO	Population Intervention Comparator Outcome
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
QUADAS-2	Quality Assessment of Diagnostic Accuracy Studies – 2
RCT	Randomised Controlled Trial
SR	Systematic Review
TSH	Thyroid-stimulating hormone

Objective of the HTA Protocol

Based on a preliminary screening of the literature the objective of the health technology assessment (HTA) protocol is to formulate the research question, to define the population, intervention, comparator, outcomes (PICO), and describe the methodology to conduct a systematic literature search, extraction, analyses, and syntheses of the data in the HTA report on the topic. Key questions are defined, addressing the main HTA domains, including the clinical outcomes and budget impact, as well as legal, social, ethical, and organisational issues.

1 Policy question

Thyroid function tests are used for the diagnosis and monitoring of patients with thyroid disorders. Hypothyroidism results from an impairment of the thyroid gland to produce sufficient thyroid hormone (i.e., thyroxine (T4) or triiodothyronine (T3)) to meet the metabolic demand of the body.[1] Hypothyroidism may develop from primary thyroid gland failure (primary hypothyroidism) or insufficient thyroid gland stimulation by the hypothalamus or pituitary gland (secondary hypothyroidism). Hyperthyroidism (overactive thyroid) occurs when thyroid gland produces too much of T4/T3 or both. For the diagnosis of thyroid disorders, most guidelines recommend a two-step testing approach: TSH should be measured first, only followed by a measurement of (free(f)) T4/(f)T3, if TSH is out of the reference range or if there is clinical suspicion of abnormal TSH secretion. However, it is observed that TSH and (f)T4/(f)T3 are often measured together in a one-step test approach. The argument in favour of the one-step approach is prevention of missed diagnosis of patients with normal TSH levels and thyroid hormone levels outside the reference range. Therefore, the objective of this HTA report is to evaluate the available clinical and economic evidence associated with the one-step and the two-step test approaches in adults with suspected primary or secondary thyroid dysfunction.

2 Research question

The classical HTA approach is not suited for the objective of this HTA report. The central research questions focus mainly on the number of missed adults with suspected primary or secondary thyroid dysfunction when applying the two-step approach and the clinical and economic consequences of the one-step compared to the two-step approach. In general, systematic reviews (SRs) on the efficacy/effectiveness/safety, cost-effectiveness/budget impact, and non-systematic reviews on the legal/social/ethical and organisational issues provide the first step towards the classical HTA. However, this standard HTA approach is not suited for the objective of this HTA. Therefore, an adapted HTA approach with the following research questions will be used:

1a) What is the difference in number of cases missed (i.e., with a normal TSH level and a thyroid hormone level of (f)T4/(f)T3 outside the reference range) when adults with suspected primary or sec-

ondary thyroid dysfunction are tested with the two-step (i.e., testing TSH first, followed by the measurement of thyroid hormone (f)T4/(f)T3[a] only if TSH is out of the reference range) compared to the one-step (i.e., measurement of TSH and thyroid hormone (f)T4/(f)T3 simultaneously) test approach?

1b) Which possible diagnoses are missed in cases with a normal TSH level and a thyroid hormone level of (f)T4/(f)T3 outside the reference range, when adults with suspected primary or secondary thyroid dysfunction are tested with the two-step compared to the one-step test approach?

2. What are the costs and budget impact of thyroid function tests performed with the one-step test approach compared to the two-step test approach in adults with suspected primary or secondary thyroid dysfunction?

3 Medical background

The thyroid is a butterfly-shaped gland, located in the front of the neck. Major thyroid hormones secreted by the thyroid gland are thyroxine (T4) and triiodothyronine (T3), which mainly regulate body metabolism. After its release from the thyroid gland, T4 is converted to T3, which is an active thyroid hormone, or to reverse T3, an inactive form. The amount of T4 and T3 produced by the thyroid gland is controlled by the thyroid stimulating hormone (TSH), which is secreted in the bloodstream by the pituitary gland. Total T4/T3 level includes T4/T3 that is bound to protein, as well as T4/T3 that is not, known as free (f)T4/(f)T3. Almost all of the T4/T3 are bound while fT4/fT3 is able to enter and affect body tissue. The release of TSH by the pituitary gland is regulated by the concentration of T4 and T3 in the blood in a conversive pathway; when T4/T3 concentrations are low, the TSH production is increased, and when T4/T3 concentrations are high, the TSH production is decreased. TSH is secreted during lifetime and increases significantly in periods of rapid growth, therefore TSH levels vary widely based on age, sex, and stage of life. In addition external factors, such as stress, diet, and medication can result in fluctuation

[a] In the NICE guideline[11], the following two-step test approach is outlined: first test TSH alone and then a) if the TSH is above the reference range, test free T4 in the same sample, or b) if the TSH is below the reference range, test free T4 and free T3 in the same sample. In this report this is shortly denoted as (f)T4/(f)T3.

of TSH levels. A normal thyroid function with serum levels of TSH and (f)T4/(f)T3 within the reference ranges is referred as euthyroidism.[2,3]

Thyroid disorder is a medical condition that affects the function of the thyroid gland. The most common thyroid problems involve abnormal production of thyroid hormones. Disorders of thyroid function are frequently diagnosed, with prevalence in Europe varying between 2% and 6% in large population-based studies and the incidence calculated in a meta-analysis being 259 per 100'000 per year.[4–7] The prevalence of thyroid disorders depends on a large number of factors, of which the most important include: sex, age, geographic factors, and ethnicity. There is a clear female preponderance; women are five to eight times more likely than men to have thyroid problems.[3–6] Thyroid disorders, especially subclinical thyroid dysfunction, are common in the elderly, as the prevalence of thyroid disorders increases over age.[8,9] Also, the prevalence of thyroid disorders is higher in countries with iodine deficiency.[10] An overview of thyroid disorders based on TSH and (f)T4/(f)T3 serum levels is listed in **Table 1**. Thyroid disorders affect many body systems, the symptoms are often non-specific, and most single symptoms alone are not predictive of thyroid dysfunction, therefore there is a wide range of possible indications for testing.[11,12] The best way to diagnose thyroid dysfunction is through blood tests. The amount of thyroid hormones in the blood (TSH, total and free T4/T3 and reverse T3) indicates if the thyroid gland is functioning properly. Not all combinations are even appropriate and useful in every diagnosis. TSH level in a blood sample is usually the first test performed. Both T4 and fT4 can be measured but fT4 more accurately reflects how the thyroid gland is functioning. Because T4 is converted into T3, free T4 is the more important hormone to measure. Any changes show up in T4 first. T3 tests are useful to identify hyperthyroidism or to determine the severity of hyperthyroidism. T3 testing is rarely helpful to identify hypothyroidism. Measurement of free T3 is possible, but is often not reliable and therefore seen as not helpful. Blood test for reverse T3 is not clinically useful.

Table 1: Thyroid disorders based on TSH and (f)T4/(f)T3 serum levels

Thyroid Disorders		**TSH**		
		Low	**Normal**	**High**
(f)T4 or (f)T3	**High**	Primary hyperthyroidism	Central hyperthyroidism	Secondary hyperthyroidism
	Normal	Subclinical hyperthyroidism	Euthyroidism	Subclinical hypothyroidism
	Low	Secondary hypothyroidism	Central hypothyroidism	Primary hypothyroidism

(f)T3 = (free) triiodothyronine, (f)T4 = (free) thyroxine, TSH = thyroid-stimulating hormone.

Hyper- or hypothyroidism

Hyperthyroidism (overactive thyroid) occurs when thyroid gland produces too much of the hormone thyroxine. Hypothyroidism (underactive thyroid) occurs when thyroid gland does not produce enough hormones. Hyperthyroidism is characterised by increased thyroid hormone synthesis and secretion from the thyroid gland, whereas thyrotoxicosis refers to the clinical syndrome of excess circulating thyroid hormones, irrespective of the source.[13]

Primary and secondary hyper- or hypothyroidism

Primary hyper- or hypothyroidism originates within the thyroid gland (e.g., caused by autoimmune thyroid disease), whereas secondary hyper- or hypothyroidism refers to a change in the thyroid gland as a result of a disease in another organ (e.g., due to a TSH-secreting pituitary adenoma).[14]

Subclinical hyper- or hypothyroidism

Subclinical hyper- or hypothyroidism occurs when TSH levels are not within the reference range, but concentrations of (f)T4/(f)T3 are within the reference range.[15]

Central hyper- or hypothyroidism

A normal TSH value usually means that the thyroid is functioning properly.[3] However, central hyper- or hypothyroidism can be seldomly diagnosed when TSH levels are within the reference range and concentrations of (f)T4/(f)T3 are outside the reference range.[16] This is characterised by a normal functioning thyroid gland and a defect of thyroid hormone production due to insufficient stimulation by TSH caused by a pituitary gland or hypothalamus disorder.[17] TSH levels in these patients are within the reference range, therefore testing TSH first (i.e., with the two-step approach) may miss diagnosis of central hyper- or hypothyroidism.

Thyroid function tests are used for the diagnosis, screening, and monitoring treatment of thyroid disorders. The focus of this HTA is on the application of thyroid function tests for diagnosis in adults with suspected primary or secondary thyroid dysfunction. Screening (i.e., presumptive identification of unrecognised disease with tests in people who do not have the symptoms of early disease) and monitoring or follow-up of thyroid dysfunction are out of the scope.[18]

4 Technology description

A thyroid function test quantifies TSH and the circulating thyroid hormones in serum (f)T4/(f)T3, to assess the ability of the thyroid gland to produce and regulate thyroid hormone production.[11,12] Thyroid function tests have high sensitivity and specificity, therefore these tests are amongst the most widely used blood tests when a thyroid disorder is suspected.[11] Since 1995, third-generation assays, which represent an extra 10-fold increase in sensitivity compared to second-generation assays, have been used.[6] Two possible test approaches can be applied: the one-step test approach (i.e., measurement of TSH and thyroid hormone (f)T4/(f)T3 simultaneously) or two-step test approach (i.e., testing TSH first, followed by the measurement of thyroid hormone (f)T4/(f)T3 only if TSH is out of the reference range). When hypo- and hyperthyroidism is suspected in adults, clinical guidelines recommend the two-step test approach: first test TSH alone and then a) if the TSH is above the reference range, test free T4 in the same sample, or b) if the TSH is below the reference range, test fT4 and fT3 in the same sample.[11] For the one-step test approach only one sample is taken for analysis, whereas in the two step approach there are to different time points to be assessed, making the interpretation difficult. For the two-step approach, the second testing can be done from a second blood sample or determined from the first sample. Because thyroid hormones and TSH test results are variable over time, current guidelines emphasize to have an interval of two to three months before the diagnosis. To control the time points of blood sampling, two complete sets of thyroid hormone tests should be analysed for comparable results.[19]

Thyroid function tests are not standardized or harmonized. The reference intervals of thyroid function tests are hampered by wide variability of commercial immunoassays. Clinical laboratories may address this issue by using assay-specific reference intervals. Some laboratories may use Liquid Chromatography with tandem Mass Spectrometry (LC-MS/MS) instead of immunometric assays for thyroid function tests.[20] LC-MS/MS analysis is selective and enables the determination of several compounds in a single analysis.

The thyroid function tests needed for the initial diagnosis may differ from the tests needed for follow-up. For example, for the characterization of (subclinical) hypothyroidism, there is a need to have both TSH and fT4. In contrast, the longitudinal follow-up can be performed in many cases with TSH only.

5 PICO

The Population Intervention Comparator Outcome (PICO) method is used to specify the research question. The PICO is described in more detail in **Table 2** and **Table 3**.

Table 2: PICO 1 (Population - Intervention - Comparator - Outcome) for clinical systematic review

P: Adults with suspected primary or secondary thyroid dysfunction

I: Thyroid function tests performed with the one-step approach (i.e., measurement of TSH and thyroid hormone (f)T4/(f)T3 simultaneously)

C: Not applicable

O: Primary clinical outcome:
Number and percentage of patients classified in nine categories :

- Low TSH – Low (f)T4/(f)T3
- Low TSH – Normal (f)T4/(f)T3
- Low TSH – High (f)T4/(f)T3
- Normal TSH – Low (f)T4/(f)T3
- Normal TSH – Normal (f)T4/(f)T3
- Normal TSH – High (f)T4/(f)T3
- High TSH – Low (f)T4/(f)T3
- High TSH – Normal (f)T4/(f)T3
- High TSH – High (f)T4/(f)T3

		TSH		
		Low	**Normal**	**High**
(f)T4/(f)T3	**Low**	N (%)	N (%)	N (%)
	Normal	N (%)	N (%)	N (%)
	High	N (%)	N (%)	N (%)

Secondary clinical outcome:
In studies included reporting data on our primary outcome of interest, what is described on which possible diagnoses could have been missed in adults with a normal TSH level and a thyroid hormone level of (f)T4/(f)T3 outside the reference range when tested only with the first test of the two-step-approach?

Table 3: PICO 2 for systematic review of costs and budget impact analysis

P: Adults with suspected primary or secondary thyroid dysfunction

I: Thyroid function tests performed with the one-step approach (i.e., measurement of TSH and thyroid hormone (f)T4/(f)T3 simultaneously)

C: Thyroid function tests performed with the two-step approach (i.e., testing TSH first, followed by the measurement of thyroid hormone (f)T4/(f)T3 only if TSH is out of the reference range)

O: Economic outcomes

- Medical costs
- Budget impact analysis

6 HTA key questions

For the evaluation of the thyroid function tests the following key questions covering the adapted HTA domains are addressed:

1a) What is the difference in number of cases missed (i.e., with a normal TSH level and a thyroid hormone level of (f)T4/(f)T3 outside the reference range) when adults with suspected primary or secondary thyroid dysfunction are tested with the two-step (i.e., testing TSH first, followed by the measurement of thyroid hormone (f)T4/(f)T3 only if TSH is out of the reference range) compared to the one-step (i.e., measurement of TSH and thyroid hormone (f)T4/(f)T3 simultaneously) test approach?

1b) Which possible diagnoses are missed in cases with a normal TSH level and a thyroid hormone level of (f)T4/(f)T3 outside the reference range, when adults with suspected primary or secondary thyroid dysfunction are tested with the two-step compared to the one-step test approach?

2. What are the costs and budget impact of thyroid function tests performed with the one-step test approach compared to the two-step test approach in adults with suspected primary or secondary thyroid dysfunction?

3. Are there any ethical, legal, social, or organisational issues related to the one-step versus two-step thyroid function test approach in adults with suspected primary or secondary thyroid dysfunction?

7 Methodology

For the adapted HTA approach, the following methodology will be applied: one SR on the clinical outcomes (Table 2) and one on costs and budget impact analyses (Table 3), and non-systematic reviews will be performed for the legal/social/ethical and organisational issues of the HTA. In the following sections the SR methodology is described in detail.

7.1 Clinical systematic review

A SR is a method to collect, critically appraise, and summarise the best available evidence in a transparent and systematic way using generally accepted evidence-based principles. The SR is designed to

search for up-to-date and high-quality evidence, according to current standards and clinical practice. The applied methodology follows international standards, such as the Cochrane Collaboration guidelines for performing SRs, and the reporting of the SR follows the recommendations of the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA).[21,22]

The SR process consists of the following fundamental steps:

1. Formulation of the research questions
2. Comprehensive information search, including defining data sources and search strategy
3. Selection procedure, applying pre-determined clear inclusion and exclusion criteria
4. Critical appraisal (quality and risk of bias assessment)
5. Data extraction
6. Data synthesis
7. Quality control

7.1.1 Databases and search strategy

7.1.1.1 Search strategy

PubMed (MEDLINE) and Embase.com databases will be searched for peer-reviewed scientific literature. Since there is considerable overlap in studies included in other literature databases (such as Cochrane Library), it is decided to search in these two main databases. The search strategy will be built using the PICO-framework (see Chapter 5). Only a search string will be made for the 'intervention' (i.e., the one-step test approach). We will not include search strings for 'population' and 'outcome', because the objectives and outcomes of studies on thyroid function tests are very diverse (e.g., studies focussing on the necessity of additional (f)T4/(f)T3 testing; studies on the effectiveness of interventions to reduce ordering of (f)T4/(f)T3 tests in addition to TSH tests) and population details are not always described in the title or abstract of these articles. Adding search strings for 'population' and 'outcome' to the search strategy will lower the number of hits but will increase the chance of missing pertinent articles. Animal studies, case reports, irrelevant publication types (e.g., editorials, expert opinions), and congress abstracts will be excluded with additional search strings. We will search in four languages: English, German, French, and Italian. The publication period will be limited from 1990 to 2021. This time period is based on a combination of publication dates found with a preliminary search for key articles and expert input. The details of the search strategies are included in Appendix 9.1. The literature database output, including all indexed fields per record (e.g., title, authors, and abstract), will be exported to Endnote version 20. Duplicates in Endnote will be automatically identified and manually deleted.

7.1.1.2 Selection procedure

From the articles retrieved from PubMed (MEDLINE) and Embase.com the relevant references will be selected by a three-step selection procedure, based on:

1. Screening of title and abstract: this step will yield the articles that will be assessed in full-text. The major topics of the articles will be assessed on relevancy for the research question by the title and abstract. In this step, articles that seem to contain relevant data for the research question will be selected for full-text screening, while articles that do not seem to contain relevant data will not be selected for full-text assessment. In case of doubt, the study will be assessed in full-text.
2. Screening of full article: the articles selected during the first phase will be assessed in full-text. Articles will be included if the reported information is relevant for the research question and the methodological description and results section are of sufficient quality, based on the predefined inclusion and exclusion criteria (Section 7.1.1.3).
3. Screening during data extraction phase: further scrutiny of the article during the data extraction phase might lead to exclusion, for example for studies with unexplained errors in their patient flow or studies based on duplicate data.

The process of selection and inclusion and exclusion of articles will be registered in Microsoft Excel and an Endnote library. The exclusion criteria applied during the full-text screening phase will be reported in a PRISMA flow chart. The implemented quality control during the selection process is described in Section 7.1.1.4.

7.1.1.3 Inclusion and exclusion criteria

The inclusion and exclusion criteria which will be applied during the selection processes are presented in **Table 4.** If relevant other criteria emerge during the screening phase, the list will be adapted in close collaboration with the FOPH. The final list of applied inclusion and exclusion criteria will be presented in the HTA report.

Table 4. Inclusion and exclusion criteria

	Inclusion	**Exclusion**
Period publication	≥1990	<1990
Language of publication	• English • German • French • Italian	All other languages
Country of study	Worldwide	-
Study design/type	• Comparative non-randomised studies (e.g., cohort studies, retrospective	• Systematic reviews† • Narrative reviews

	observational studies) • Randomised controlled trials	• Case reports • Abstract only (e.g., conference abstract) • Non-pertinent publication types (e.g., expert opinion, letter to editor, editorial, comment)
Study quality	Sufficient methodological quality and coherent reporting of the results	Major insufficient methodological quality or incoherent reporting of the results (e.g., unexplained errors in patient flow)
Study population	Adults with suspected primary or secondary thyroid dysfunction	• Children • Populations screened for thyroid disorders • Populations monitored for the treatment of thyroid disorders • Studies only including women with specific female sex hormonal states, e.g., pregnant women, non-pregnant women on fertility-related treatment, or menopausal women
Study intervention	Thyroid function tests performed with the one-step approach (i.e., measurement of TSH and thyroid hormone (f)T4/(f)T3 simultaneously)	All other study interventions/comparators
Comparison	Not applicable	Not applicable
Study outcomes	See PICO 1‡	Other outcomes

Keys: (f)T3 = (free) triiodothyronine, (f)T4 = (free) thyroxine, PICO = Population Intervention Comparator Outcome, TSH = thyroid-stimulating hormone. †Relevant SRs will be selected during the screening of title and abstract phase. During the full-text screening phase, reference lists of good quality SRs will be checked for possibly missed relevant individual articles. No data extraction will be performed for SRs, only for relevant individual articles; ‡See PICO 1 in Chapter 5.

7.1.1.4 Quality control

The following quality control measures will be applied during the systematic literature search:

- Search strategy
    - An information specialist will be consulted during the development of the search strategy. Quality checks will be implemented, for example the search strategy will be checked by a second researcher and run multiple times at separate days.
    - The supplementary search technique citation chasing (i.e., backward by finding other studies cited within the selected articles) will be applied in addition to the database searches. Additional studies will be enclosed in the selection process. Relevant SRs will be selected

during the screening of title and abstract phase. During the full-text screening phase, reference lists of good quality SRs will be checked for possibly missed relevant individual articles.

- Selection process
    - The first 30% of titles and abstracts from the peer-reviewed literature will be screened in duplicate by two independent researchers. The results will be compared and discussed before the remaining references are assessed by one researcher. Both researchers categorise the titles as 'include for full-text assessment', 'exclude for full-text assessment', or 'doubt'. If there are differences between the two researchers regarding more than 2% of the articles selected as 'include for full-text assessment', another 10% of the articles will be screened in duplicate. This will be repeated if necessary. If there is still more than 2% discrepancy at 50% of the duplicate selection, the screening of title and abstracts will be done fully in duplicate by two independent researchers. If the two reviewers disagree on the relevance of a study, this will be discussed. If the differences remain after discussion, the study will be assessed in full-text.
    - The full-text articles from the peer-reviewed literature will be assessed for relevancy by one researcher in close collaboration with a second researcher; any doubts are discussed in detail. In case of discrepancy or disagreements during the selection phase, a third researcher is consulted. The study will be discussed until consensus is reached. To double check whether the relevant articles are included, all full-text articles categorised as excluded will be assessed in duplicate by a second researcher.

7.1.2 Data extraction, analysis, and synthesis

7.1.2.1 Data extraction

Relevant data from the included studies found in the peer-reviewed literature will be summarised using a standardised data-extraction spreadsheet in Excel to present relevant information for the review objectives per included article (i.e., baseline characteristics, study results, and risk of bias).

7.1.2.2 Critical appraisal

The risk of bias of the study designs of the individual included studies will be assessed. For this HTA we expect to include a mix of study designs, and it might be that no standardised checklist is available for each study design (e.g., for retrospective observational studies). Based on the type of studies which will be included, we will choose the most appropriate method to assess the risk of bias. For example,

the National Institute for Health and Care Excellence (NICE) has a series of methodology checklists,[11] including checklists for cohort studies and diagnostic test accuracy studies (i.e., the QUADAS-2 tool), which might be used. This will be discussed with the FOPH during the project and the preferred critical appraisal method will be reported in the methodology section of the HTA report.

For the clinical research question (i.e., a comparator is not applicable), the type of studies included, and outcomes of interest, it is not possible to apply the Grading of Recommendations, Assessment, Development, and Evaluations (GRADE) approach for diagnostic tests and strategies nor the GRADE approach for therapy.[23]

7.1.2.3 Data synthesis

The options for clinically relevant data merging/stratification will be discussed with clinical experts, based on the level of detail of data reporting in the included studies. The clinical experts will be blinded for the study results during this stratification process. The details of stratification will be reported in the methodology section of the HTA report.

Pooled estimates will be calculated when 1) two or more studies report on the same outcome and assessed in the same way 2) sufficient data is reported in the studies. Considering the variability in the baseline characteristics of the studies, a random-effects model will be considered for the analyses.[24] All analyses will be conducted using the MetaXL (www.epigear.com) add-in for Microsoft Excel. The data will be summarised in summary tables, and forest plots will be created to visualise the results if possible.

Outcomes for which it is not possible to calculate pooled estimates will be presented narratively in summary tables and accompanying text. The summary tables will present the key features (e.g., study types, setting, populations, interventions used) and main outcomes of the selected studies and will be organised in a logical manner (e.g., by outcome measure or subpopulation).

7.1.2.4 Quality control

The following quality control measures will be applied during the data extraction and synthesis:

- The critical appraisal of included studies will be done in duplicate. In case of discrepancy a third researcher will be consulted to reach consensus.
- The data extraction spreadsheet will be fully checked with the original articles by a second researcher.
- The data synthesis files and evidence profiles/summary tables will be fully reviewed by a second researcher.

7.2 Costs and budget impact analyses

7.2.1 Databases and search strategy

7.2.1.1 Search strategy

The budget impact and cost analyses systematic literature searches follow the principles of the systematic literature search outlined in Chapter 7.1. PubMed (MEDLINE) and Embase.com databases will be searched for peer-reviewed scientific literature. In addition, NHS EED economic database will be searched. The searches are built using the PICO-framework (see Chapter 5). In PubMed (MEDLINE) and Embase.com, the search terms of the clinical outcome literature search are combined with budget impact and cost search terms. The details of the search strategy are presented in Appendix 9.2.

7.2.1.2 Selection procedure

The full-text articles, which will be selected in the first step, will then be assessed based on the inclusion and exclusion criteria as defined in this HTA protocol (Table 5). Articles will be included if they fulfil the eligibility criteria. The process of selection and inclusion and exclusion of articles will be recorded in Microsoft Excel and Rayyan (www.rayyan.ai). This method will provide transparency regarding all selection steps and assures reproducibility. The selection procedure applied during the full-text screening phase will be reported in a PRISMA flow chart.

7.2.1.3 Inclusion and exclusion criteria

The inclusion and exclusion criteria which will be applied during the selection processes are listed in **Table 5**. The final list of the inclusion and exclusion criteria will be presented in the HTA report.

Table 5. Inclusion and exclusion criteria for costs and budget impact analyses studies

	Inclusion	**Exclusion**
Period publication	No restriction/ from database inception	-
Country of study	Worldwide	-
Study design/type	Economic evaluations: • Budget impact analysis • Costs analysis	Resource use measurement
Study population	Adults with suspected primary or secondary thyroid dysfunction	• Children

		• Populations screened for thyroid disorders • Populations monitored for the treatment of thyroid disorders • Studies only including women with specific female sex hormonal states, e.g., pregnant women, non-pregnant women on fertility-related treatment, or menopausal women
Study intervention	Thyroid function tests performed with the one-step approach (i.e., measurement of TSH and (f)T4/(f)T3 simultaneously)	All other study interventions
Study comparison	Thyroid function tests performed with the two-step approach (i.e., testing TSH first, followed by the measurement of (f)T4/(f)T3 only if TSH is out of the reference range)	• All other comparators • No comparator
Study outcomes	See PICO 2‡	Other outcomes

Keys: (f)T3 = (free) triiodothyronine, (f)T4 = (free) thyroxine, PICO = Population Intervention Comparator Outcome, TSH = Thyroid-stimulating hormone. ‡See PICO 2 in Chapter 5.

7.2.1.4 Quality control

- Search strategy

Search strategy will be checked by a second researcher. The supplementary search technique citation chasing (i.e., studies cited within the included articles) will be applied.

- Selection process

The full-text articles from the peer-reviewed literature will be assessed for relevancy in duplicate by two independent researchers. If there are differences between the findings of the two researchers, these differences will be identified and discussed. In case of discrepancy or disagreements a third researcher will be consulted.

7.2.2 Data extraction, analysis, and synthesis

7.2.2.1 Data extraction

Relevant data from the included articles found in the peer-reviewed literature will be summarised using a preliminary data-extraction spreadsheet in Microsoft Excel. This spreadsheet will include:

- First author, year

- Country
- Type of study (cost and/or budget impact)
- Study population
- Intervention
- Comparator
- Economic outcomes
- Time Horizon
- Primary sources for the resource use/cost inputs
- Conflict of interest and study funding
- Discount rate
- Study perspective
- Key assumptions
- Conclusions of the study authors

7.2.2.2 Data synthesis

Data synthesis will be done using descriptive comparisons of the study question, methods, and results. Summary tables will be included which will present key information described in the data extraction chapter 7.2.2.1. The analytical approaches used in the studies will be compared and their robustness will be discussed. If the published studies do not provide sufficient information on the costs of thyroid function tests, Swiss databases and publicly available sources will be considered. A new budget impact model will be developed for Switzerland in the HTA phase. The general approach to the budget impact modelling and analysis will be further specified in the HTA report and is summarised in Chapter 7.2.2.4.

7.2.2.3 Quality Control

The following quality control measures will be applied during the data extraction and synthesis:

- The information filled into the data extraction spreadsheet will be checked by two reviewers.
- The critical appraisal of included studies will be done in duplicate. In case of discrepancy, a third researcher will be consulted.
- The data synthesis files and summary tables will be checked by two reviewers.

7.2.2.4 Budget impact analysis

In this section, general steps in developing a budget impact model are described, including conceptual model development, main characteristics and model inputs.

- ***Conceptual budget impact model development***

Research questions and the PICO described in Chapter 5 will be used for the development of the conceptual budget impact model. A conceptual model will be essential to identify necessary inputs and will help to structure data collection efforts. The conceptual model will address the key parameters that drive both specific costs and budget impact. Published models will be used as a starting point for the development of the conceptual budget impact model. The draft conceptual model will be discussed with a clinical expert to ensure that the model inputs reflect Swiss clinical practice.

- ***Main characteristics of the budget impact model***

The conceptual model and relevant data from the included articles found in the peer-reviewed literature will be used for the basis for the budget impact model. The main model characteristics include:
Setting: The analysis will be performed for the Swiss healthcare setting, and where possible, relevant input parameters will be based on data from Switzerland (e.g., Swiss sources for healthcare costs).

Perspective: The analysis will be performed from the Swiss healthcare payer perspective, and only direct healthcare costs will be included. Societal costs, such as informal care and productivity costs, will not be included.

Time horizon: The time horizon of the budget impact model will be five years.

Discount rate: In the base-case analysis, healthcare costs will be discounted at 3.0%. In scenario analyses, the impact of not using a discount rate or a discount rate of 6.0% will be explored.

Currency, price data, and conversion: Costs will be reported in Swiss Franc (CHF) adjusted for inflation to current price levels using inflation rates from the Swiss Federal Statistical Office, which will be accessed from the Organisation for Economic Co-operation and Development (OECD) website (https://data.oecd.org).

- ***Model inputs***

Model input parameters on clinical outcomes will be informed mainly from the results of the data extraction of the clinical outcomes systematic literature search. Clinical expert opinion will be used whenever data will not be available from the literature. Similarly, model input parameters on costs/budget analysis and resource use outcomes will be informed mainly from the results of the data extraction of the budget impact systematic literature search. In the case of no Swiss-specific data on costs and related resource use, as identified in the systematic literature searches described in this protocol, additional searches will be performed. Additional searches on publicly available databases will then be performed in collaboration with the FOPH to determine healthcare resource use, and unit costs. If the required data cannot be identified from public sources and systematic literature searches, assumptions will be made based on data from other comparable countries and/or expert opinion.

- ***Budget impact analysis***

Following the conceptual model development and data collection, thyroid function test budget impact model will be populated with the extracted clinical, epidemiological, and cost parameters. The budget impact model will be developed to calculate the projected population-level costs of thyroid function tests for the diagnosis of suspected primary or secondary thyroid dysfunction. The model will be programmed in Microsoft Excel. Subgroup analyses will be performed, when possible, to differentiate important population (e.g., primary care) and/or disease/indication settings. The time horizon of the budget impact analysis will be five years, with possibilities to extend model for longer durations. For this analysis, clinical and epidemiological data will be required about the current use of thyroid function tests for the diagnosis of suspected primary or secondary thyroid dysfunction in Switzerland. If this data is not readily available, then assumptions will be made based on data from other comparable countries and/or expert opinion.

7.2.2.5 Legal, social, ethical, and organisational issues

In the HTA report, we will identify and address the main legal, social, ethical, and organisational issues of the studies included in the clinical outcomes, costs, and budget impact systematic literature searches. In addition, we will perform grey literature searches on these HTA domains.

8 References

1 Ortiga-Carvalho TM, Chiamolera MI, Pazos-Moura CC, *et al.* Hypothalamus-Pituitary-Thyroid Axis. In: *Comprehensive Physiology*. Wiley 2016. 1387–428. doi:10.1002/cphy.c150027

2 (ETA). European Thyroid Association. https://www.eurothyroid.com (accessed 31 May 2021).

3 (ATA). American Thyroid Association. https://www.thyroid.org (accessed 31 May 2021).

4 Virta LJ, Eskelinen SI. Prevalence of hypothyroidism in Finland—a nationwide prescription study. *Eur J Clin Pharmacol* 2011;67:73–7. doi:10.1007/s00228-010-0884-4

5 Åsvold BO, Vatten LJ, Bjøro T. Changes in the prevalence of hypothyroidism: the HUNT Study in Norway. *Eur J Endocrinol* 2013;169:613–20. doi:10.1530/EJE-13-0459

6 Garmendia Madariaga A, Santos Palacios S, Guillén-Grima F, *et al.* The Incidence and Prevalence of Thyroid Dysfunction in Europe: A Meta-Analysis. *J Clin Endocrinol Metab* 2014;99:923–31. doi:10.1210/jc.2013-2409

7 Wouters HJCM, Slagter SN, Muller Kobold AC, *et al.* Epidemiology of thyroid disorders in the Lifelines Cohort Study (the Netherlands). *PLoS One* 2020;15:e0242795. doi:10.1371/journal.pone.0242795

8 British Thyroid Foundation. Older patients and thyroid disease. https://www.btf-thyroid.org/older-patients-and-thyroid-disease (accessed 31 May 2021).

9 Gesing A. The thyroid gland and the process of aging. *Thyroid Res* 2015;8:A8.

doi:10.1186/1756-6614-8-S1-A8

10 Zimmermann MB, Boelaert K. Iodine deficiency and thyroid disorders. *Lancet Diabetes Endocrinol* 2015;3:286–95. doi:10.1016/S2213-8587(14)70225-6

11 National Institute for Health and Care Excellence (NICE). The Guidelines manual: appendices B-1. https://www.nice.org.uk/process/pmg6/resources/the-guidelines-manual-pdf-3304416006853

12 National Institute for Health and Care Excellence. Thyroid disease: assessment and management [B] Indications for testing NICE guideline NG145 Prognostic evidence review underpinning recommendations. 2019.

13 De Leo S, Lee SY, Braverman LE. Hyperthyroidism. *Lancet* 2016;388:906–18. doi:10.1016/S0140-6736(16)00278-6

14 Hammett-Stabler C, Maygarden S. Pathology: A Modern Case Study. Chapter 15 Pathology of the Endocrine System. https://accessmedicine.mhmedical.com/content.aspx?sectionid=95970959&bookid=1569

15 Vasileiou M, Gilbert J, Fishburn S, *et al.* Thyroid disease assessment and management: summary of NICE guidance. *BMJ* 2020;:m41. doi:10.1136/bmj.m41

16 Wardle CA, Fraser WD, Squire CR. Pitfalls in the use of thyrotropin concentration as a first-line thyroid-function test. *Lancet* 2001;357:1013–4. doi:10.1016/S0140-6736(00)04248-3

17 Persani L. Central Hypothyroidism: Pathogenic, Diagnostic, and Therapeutic Challenges. *J Clin Endocrinol Metab* 2012;97:3068–78. doi:10.1210/jc.2012-1616

18 Hennessey J V., Garber JR, Woeber KA, *et al.* American Association Of Clinical Endocrinologists And American College Of Endocrinology Position Statement On Thyroid Dysfunction Case Finding. *Endocr Pract* 2016;22:262–70. doi:10.4158/EP151038.PS

19 Pearce SHS, Brabant G, Duntas LH, *et al.* 2013 ETA Guideline: Management of Subclinical Hypothyroidism. *Eur Thyroid J* 2013;2:215–28. doi:10.1159/000356507

20 Van Houcke SK, Van Uytfanghe K, Shimizu E, *et al.* IFCC international conventional reference procedure for the measurement of free thyroxine in serum. *Clin Chem Lab Med* 2011;49. doi:10.1515/CCLM.2011.639

21 Higgins J, Thomas J, Chandler J. Cochrane Handbook for Systematic Reviews of Interventions version 6.0. Updated July 2019. *Cochrane 2019*https://training.cochrane.org/cochrane-handbook-systematic-reviews-interventions

22 Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. http://www.prisma-statement.org

23 Schünemann H, Brożek J, Guyatt G, *et al.* GRADE handbook. GRADE handbook for grading quality of evidence and strength of recommendations. Updated October 2013. GRADE Work. https://gdt.gradepro.org/app/handbook/handbook.html

24 DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials* 1986;7:177–88. doi:10.1016/0197-2456(86)90046-2

9 Appendices

Appendix 9.1. Search strategy for the clinical outcomes

Table I. Search strategy for the clinical systematic literature search: PubMed (MEDLINE)

Population	No search string
Intervention: thyroid function test	("Thyroid function tests"[Mesh] OR test*[tiab] OR assay*[tiab]) AND (("Thyrotropin"[Mesh] OR thyrotropin[tiab] OR thyroid stimulating hormone*[tiab] OR thyroid-stimulating hormone*[tiab] OR TSH[tiab]) AND ("Thyroxine"[Mesh] OR thyroxine[tiab] OR T4[tiab] OR freeT4[tiab] OR free-T4[tiab] OR FT4[tiab] OR "Triiodothyronine"[Mesh] OR Triiodothyronine[tiab] OR T3[tiab] OR freeT3[tiab] OR free-T3[tiab] OR FT3[tiab]))
Comparator	No search string
Outcomes	No search string
Limits	*No animal studies:* NOT (Animals[Mesh] NOT (Humans[Mesh] AND Animals[Mesh]))
	No case reports, irrelevant publication types, congress abstracts: NOT (case reports[pt] OR editorial[pt] OR letter[pt] OR news[pt] OR comment[pt] OR congress[pt])
	Publication period: 1990 – current
	Language: English, German, French, Italian

Table II. Search strategy for the clinical systematic literature search: Embase.com

Population	No search string
Intervention: thyroid function test	('thyroid function test'/exp OR test*:ti,ab OR assay*:ti,ab) AND (('thyrotropin'/exp OR thyrotropin:ti,ab OR "thyroid stimulating hormone*":ti,ab OR "thyroid-stimulating hormone*":ti,ab OR TSH:ti,ab) AND ('thyroxine'/exp OR thyroxine:ti,ab OR T4:ti,ab OR freeT4:ti,ab OR free-T4:ti,ab OR FT4:ti,ab OR Triiodothyronine:ti,ab OR T3:ti,ab OR freeT3:ti,ab OR free-T3:ti,ab OR FT3:ti,ab))
Comparator	No search string
Outcomes	No search string
Limits	*No animal studies:* NOT ('animal'/exp OR 'nonhuman'/exp NOT ('animal'/exp OR 'nonhuman'/exp AND 'human'/exp))
	Relevant publication types: [article]/lim OR [article in press]/lim OR [review]/lim
	Publication period: 1990 – current
	Language: English, German, French, Italian

Appendix 9.2. Search strategy costs and budget impact analyses

Table I. Search strategy for costs and budget impact literature: PubMed (MEDLINE)

Population	No search string
Intervention/Comparator: thyroid function tests	("Thyroid function tests"[Mesh] OR test*[tiab] OR assay*[tiab]) AND (("Thyrotropin"[Mesh] OR thyrotropin[tiab] OR thyroid stimulating hormone*[tiab] OR thyroid-stimulating hormone*[tiab] OR TSH[tiab]) AND ("Thyroxine"[Mesh] OR thyroxine[tiab] OR T4[tiab] OR freeT4[tiab] OR free-T4[tiab] OR FT4[tiab] OR "Triiodothyronine"[Mesh] OR Triiodothyronine[tiab] OR T3[tiab] OR freeT3[tiab] OR free-T3[tiab] OR FT3[tiab]))
Outcomes	No search string
Limits	*Study design:* (“budget impact ”[Mesh] OR “cost analysis”[Mesh] OR “budget analysis”[Mesh] OR “cost assessment” [tiab] OR “economic” [tiab] OR “economic value” [tiab] OR “cost” [tiab] OR “budget” [tiab] OR “budget impact” [tiab] OR “budget analysis” [tiab])
Limits	*Publication period:* No restrictions
	Language: No restrictions

Table II. Search strategy for costs and budget impact literature: Embase.com

Population	No search string
Intervention/Comparator: thyroid function tests	('thyroid function test'/exp OR test*:ti,ab OR assay*:ti,ab) AND (('thyrotropin'/exp OR thyrotropin:ti,ab OR "thyroid stimulating hormone*":ti,ab OR "thyroid-stimulating hormone*":ti,ab OR TSH:ti,ab) AND ('thyroxine'/exp OR thyroxine:ti,ab OR T4:ti,ab OR freeT4:ti,ab OR free-T4:ti,ab OR FT4:ti,ab OR Triiodothyronine:ti,ab OR T3:ti,ab OR freeT3:ti,ab OR free-T3:ti,ab OR FT3:ti,ab))
Outcomes	No search string
Limits	*Study design:* ('budget impact /exp OR 'costs'/exp OR 'cost analysis'/exp OR 'budget analysis'/de OR ((cost NEAR/3 analysis*) OR ((cost OR costs) NEAR/3 (budget* OR economic* OR value *)):ab,ti)
Limits	*Publication period:* No restrictions
	Language: No restrictions

Table III. Search strategy for costs and budget impact literature: NHS EED

Database	NHS EED
Population	(thyroid dysfunction) OR (hypothyroidism) OR (hyperthyroidism)
Intervention	(*triiodothyronine*) OR (*thyroxine*) OR (*thyroid stimulating hormone*)
Comparator	No search string
Outcomes	No search string
Limits	No limits

NHS EED: https://www.crd.york.ac.uk/CRDWeb/