Health Technology Assessment (HTA)

HTA Protocol

Title	Levothyroxine for patients diagnosed with subclinical hypothyroidism
Author/Affiliation	**Remziye Zaim, Frederick Thielen, Philip Klein, 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	Levothyroxine
Date	19-08-2021

Executive Summary

Subclinical hypothyroidism (SCH) is a hormonal disorder in which the serum thyroid-stimulating hormone levels (TSH) are elevated while the thyroid hormone levels are within the normal reference range. SCH can be treated with hormone replacement therapy using levothyroxine, a synthetic version of the endogenous thyroid hormone thyroxine (T4), to raise low levels of endogenous thyroid hormones. However, the need for treatment of this subclinical disorder is still under debate. In Switzerland, levothyroxine is approved by the Swiss Agency for Therapeutic Products (swissmedic) for the treatment of hypothyroidism. There is no standard alternative treatment option for patients diagnosed with SCH. In this Health Technology Assessment (HTA) protocol, the central research question and the operational key questions covering the HTA domains clinical effectiveness/safety, cost-effectiveness/budget impact, ethical/legal/social/organisational issues are formulated and the methodological approach to conduct the HTA is described.

For the clinical review, a systematic literature search of the Pubmed (MEDLINE) and Embase databases will be conducted adhering to international methodological standards. Randomized controlled trials (RCTs) and possibly comparative non-randomised studies if less than ≤1 RCT is found, including patients with SCH who are treated with levothyroxine compared with no treatment or placebo, will be included. We will search worldwide, in English, German, French and Italian language journals. Studies only including women with specific female sex hormonal states and congenital hypothyroidism are excluded. The options for clinically relevant data merging/stratification will be investigated after data-extraction. Outcomes for which there is no possibility to calculate pooled estimates will be presented narratively using summary tables and accompanying text.

For the economic review, cost-effectiveness searches will follow the principles of the clinical systematic literature search. After applying the quality control measures, data synthesis will be performed using descriptive comparisons of the study question, methods and results. If the published studies on cost-effectiveness of levothyroxine for patients with SCH, compared to placebo or no treatment, do not provide sufficient information in Switzerland, cost-effectiveness and budget impact models will be developed in the HTA phase. Subgroup analyses will be performed, where possible, to differentiate cost-effectiveness by population or disease/indication area. Parameter and structural uncertainty of the cost-effectiveness model will be explored.

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

Die subklinische Hypothyreose (SKH) oder latente Schilddrüsenunterfunktion ist eine hormonelle Störung, bei der die Serumkonzentration des schilddrüsenstimulierenden Hormons (Thyreoidea-stimulierendes Hormon, TSH) erhöht ist, während die Spiegel der Schilddrüsenhormone im normalen Referenzbereich liegen. Die SKH kann durch eine Hormonersatztherapie behandelt werden, bei der Levothyroxin, eine künstlich hergestellte Version des körpereigenen Schilddrüsenhormons Thyroxin (T4), eingesetzt wird, um niedrige Spiegel der körpereigenen Schilddrüsenhormone anzuheben. Die Notwendigkeit einer Behandlung dieser subklinischen Störung wird jedoch noch diskutiert. In der Schweiz ist Levothyroxin vom Schweizerischen Heilmittelinstitut (Swissmedic) für die Behandlung der Hypothyreose zugelassen. Für Patientinnen und Patienten mit diagnostizierter SKH gibt es keine standardmässige alternative Behandlungsmöglichkeit. In diesem Health Technology Assessment (HTA) werden die zentrale Forschungsfrage und die operativen Schlüsselfragen in den HTA-Bereichen klinische Wirksamkeit/Sicherheit, Kosteneffektivität/Budgetauswirkungen und ethische/rechtliche/soziale/organisatorische Fragen formuliert. Des Weiteren wird der methodische Ansatz zur Durchführung des HTA beschrieben.

Für die klinische Prüfung wird eine systematische Literaturrecherche in den Datenbanken Pubmed (MEDLINE) und Embase unter Einhaltung internationaler methodischer Standards durchgeführt. Einbezogen werden randomisierte kontrollierte Studien (RCTs) und möglicherweise vergleichende nichtrandomisierte Studien bei ≤1 RCT, die mit Levothyroxin behandelte Patientinnen und Patienten mit SKH im Vergleich zu keiner Behandlung oder zu Placebo einschliessen. Gesucht wird weltweit in englisch-, deutsch-, französisch- und italienischsprachigen Zeitschriften. Studien, die nur Frauen mit spezifischen Geschlechtshormonstörungen und angeborener Hypothyreose einbeziehen, werden ausgeschlossen. Nach der Datenextraktion werden die Möglichkeiten für die 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 zur Kosteneffektivität nach den Grundsätzen der klinischen systematischen Literaturrecherche. Nach Anwendung der Qualitätskontrollmassnahmen wird die Datensynthese anhand deskriptiver Vergleiche der Studienfragen, Methoden und Ergebnisse durchgeführt. Falls die veröffentlichten Studien zur Kosteneffektivität von Levothyroxin bei Patientinnen und Patienten mit SKH im Vergleich zu Placebo oder zu keiner Behandlung in der Schweiz nicht genügend Informationen liefern, werden in der HTA-Phase Kosteneffektivitäts- und Budgetauswirkungsmodelle entwickelt. Wo möglich, werden Untergruppenanalysen durchgeführt, um die Kosteneffektivität nach Population oder Krankheit/Indikationsbereich zu differenzieren. Die

Parameter- und Strukturunsicherheit des Kosteneffektivitätsmodells wird untersucht.

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

Résumé

L'hypothyroïdie subclinique (HSC) est un trouble hormonal se caractérisant par des taux sériques élevés de thyréostimuline (TSH), tandis que les taux d'hormones thyroïdiennes se situent dans l'intervalle normal de référence. Elle peut faire l'objet d'un traitement hormonal de substitution à base de lévothyroxine, une version synthétique de la thyroxine (T4). L'objectif est d'augmenter les faibles niveaux d'hormones thyroïdiennes endogènes, dont la thyroxine fait partie. Cependant, la nécessité de traiter ce trouble subclinique est toujours sujette à débat. En Suisse, la lévothyroxine est autorisée par l'Institut suisse des produits thérapeutiques (Swissmedic) pour le traitement de l'hypothyroïdie. Il n'existe pas d'autre traitement standard pour les patients ayant reçu un diagnostic d'HSC. Le présent rapport d'ETS (évaluation des technologies de la santé) formule la question de recherche centrale et les questions-clés opérationnelles dans les domaines d'ETS suivants : efficacité/sécurité clinique, rapport coût-efficacité/impact budgétaire, aspects éthiques/juridiques/sociaux/organisationnels. L'approche méthodologique adoptée y est également décrite.

Pour l'évaluation clinique, une recherche systématique de la littérature sera effectuée dans les bases de données Pubmed (MEDLINE) et Embase, en suivant les normes méthodologiques internationales. Elle inclura des essais cliniques randomisés (ECR) – et des études comparatives non randomisées si le nombre d'ECR identifiés est inférieur ou égal à 1 – comparant des patients atteints d'HSC traités par la lévothyroxine à d'autres non traités ou recevant un placebo. Nous rechercherons dans les revues du monde entier – en anglais, en allemand, en français et en italien. Les études portant uniquement sur des femmes atteintes d'hypothyroïdie congénitale et dont les hormones sexuelles féminines présentent des états spécifiques sont exclues. 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 portant sur le rapport coût-efficacité seront menées selon les principes de la recherche systématique de la littérature clinique. Une fois les mesures de contrôle de la qualité appliquées, une synthèse des données sera effectuée par le biais de comparaisons descriptives de la problématique, de méthodes et des résultats de chaque étude. Si les études

publiées sur le rapport coût-efficacité de la lévothyroxine par rapport à un placebo ou à l'absence de traitement chez les patients présentant une HSC ne fournissent pas suffisamment d'informations pour la Suisse, des modèles coût-efficacité et d'impact budgétaire seront mis au point lors de la phase d'ETS. Des analyses par sous-groupes seront effectuées dans la mesure du possible, afin d'obtenir des rapports coût-efficacité différenciés par populations ou par familles de maladies/d'indications. L'incertitude liée aux paramètres et à la structure du modèle coût-efficacité sera examinée.

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

L'ipotiroidismo subclinico (*Subclinical Hypothyroidism*, SCH) è un disturbo ormonale caratterizzato da elevati livelli sierici di ormone stimolante la tiroide (TSH) e valori normali dell'ormone tiroideo. L'SCH può essere trattato con una terapia ormonale sostitutiva a base di levotiroxina, una versione sintetica della tiroxina (T4), ormone tiroideo endogeno, per ripristinare i livelli normali di ormoni tiroidei endogeni. Ciononostante, la necessità di trattare questo disturbo subclinico è ancora oggetto di discussione. In Svizzera, la levotiroxina è approvata dall'Istituto svizzero per gli agenti terapeutici (Swissmedic) per il trattamento dell'ipotiroidismo. Non vi è un'alternativa terapeutica standard per i pazienti cui è stato diagnosticato l'SCH. Nel presente protocollo di *Health Technology Assessment* (HTA) vengono formulati il quesito principale dello studio e i quesiti chiave operativi che coprono i campi efficacia/sicurezza clinica, efficacia dei costi/impatto sul bilancio, questioni etiche/legali/sociali/organizzative; inoltre, viene illustrato l'approccio metodologico alla base dell'HTA.

Per la revisione clinica sarà condotto uno studio sistematico della letteratura nelle banche dati di Pubmed (MEDLINE) ed Embase, in linea con gli standard metodologici internazionali. Saranno inclusi studi controllati randomizzati (RCT), ed eventualmente studi comparativi non randomizzati se il numero di RCT reperiti è ≤1, che confrontano i pazienti affetti da SCH trattati con levotiroxina a quelli non trattati o trattati con placebo. Lo studio, condotto su scala mondiale, prenderà in esame riviste in lingua inglese, tedesca, francese e italiana. Non saranno presi in considerazione gli studi che includono esclusivamente donne con specifici stati ormonali sessuali femminili e ipotiroidismo congenito. 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 per il rapporto costo-efficacia avverranno sulla scorta dei principi dello studio sistematico della letteratura clinica. Dopo aver compiuto le misurazioni di controllo della qualità, la sintesi dei dati sarà effettuata confrontando in modo descrittivo il quesito, i metodi e i risultati dello studio. Se gli studi pubblicati sul rapporto costo-efficacia della levotiroxina per i pazienti con SCH rispetto al trattamento a base di placebo o a nessun trattamento non forniscono informazioni sufficienti in Svizzera, il rapporto costo-efficacia e i modelli di impatto sul bilancio saranno sviluppati in sede di HTA. Se possibile, saranno condotte analisi di sottogruppo per distinguere il rapporto costo-efficacia per popolazione o gruppo di malattie/indicazioni. Sarà indagata l'incertezza legata ai parametri e alla struttura del modello costo-efficacia.

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

CEA registry	Cost Effectiveness Analysis registry
CHEC	Consensus Health Economic Criteria
CHF	Swiss Franc
CHD	Coronary Heart Disease
DARTH	Decision Analysis in R for Technologies in Health
ETA	European Thyroid Association
FOPH	Federal Office of Public Health
GRADE	Grading of Recommendations, Assessment, Development and Evaluations
HTA	Health Technology Assessment
ICER	Incremental Cost Effectiveness Ratio
ISPOR	International Society for Pharmacoeconomics and Outcomes Research
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
QALY	Quality Adjusted Life Year
RCT	Randomised Controlled Trial
ROBINS-I	Risk of Bias In Non-Randomised Studies of Interventions
SCH	Subclinical hypothyroidism
SMDM	Society for Medical Decision Making
SR	Systematic Review
TSH	Thyroid-Stimulating Hormone

Objective of the HTA protocol

The objective of the health technology assessment (HTA) protocol is to formulate the research question, to define the Population Intervention Comparator Outcome (PICO) and describe the methodology to conduct a systematic literature search, extract, analyse, and synthesise the data in the HTA report. Key questions are defined, addressing the main HTA domains, i.e., efficacy/effectiveness/safety, cost-effectiveness/budget impact, legal/social/ethical and organisational issues.

1 Policy question

Subclinical hypothyroidism (SCH) is a hormonal disorder in which the serum thyroid-stimulating hormone (TSH) levels are elevated but the thyroid hormone levels are within the normal reference range. Epidemiological evidence suggests increased cardiovascular risk in patients with SCH and clinical data have demonstrated some benefits of levothyroxine treatment in reducing these events.[1,2] However, evidence on the association of SCH and musculoskeletal system, cognitive dysfunction, mood disorders, dyslipidaemia, diabetes and goitre is conflicting.[1] Clinical guidelines have suggested that decisions to treat SCH with levothyroxine should be made taking into account the age of the patient, associated risk factors and comorbid conditions.[1,3,4]

There are currently no limitations imposed on mandatory insurance coverage of levothyroxine treatment in SCH patients in Switzerland. Based on the controversial evidence in the literature, the effectiveness, appropriateness and economic efficiency (so-called EAE criteria) of levothyroxine treatment in SCH patients has been questioned by the applicant of this HTA topic. A limitation of coverage (e.g. for a specific subpopulation of patients) could be considered if the EAE criteria are not fulfilled.

2 Research question

What are the efficacy[1], effectiveness[2], safety[3], cost-effectiveness, and budget impact implications of levothyroxine treatment in populations diagnosed with SCH?

---

[1]Efficacy is the extent to which a specific health technology produces a beneficial, reproducible result under study conditions compared with alternative technologies (i.e., internal validity).

[2] Effectiveness is the extent to which a specific health technology, when applied in real world circumstances in the target group, does what it is intended to do for a diagnostic or therapeutic purpose regarding the benefits compared with alternative technologies (i.e., external validity).

[3]Safety is a judgement of the harmful effects and their severity using the health technology. Relevant adverse events are those that result in death, are life-threatening, require inpatient hospitalisation or cause prolongation of existing hospitalisation (i.e., serious adverse events) and those that occur repetitively and the most frequent (highest rate).

3 Medical background

The thyroid gland secrets the thyroid hormones thyroxine (T4) and triiodothyronine (T3). 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. 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 T3/T4 concentrations are low, the TSH production is increased, and when T3/T4 concentrations are high, the TSH production is decreased. Thyroid disorder is a medical condition that involves abnormal production of thyroid hormones and can be classified as hypothyroidism or hyperthyroidism. Hypothyroidism results from an impairment of the thyroid gland to produce sufficient thyroid hormone (i.e., T4 or T3), resulting in elevated levels of TSH.[5] Hyperthyroidism is caused by an overactive thyroid gland producing too much thyroid hormones, leading to decreased TSH levels.[6] 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).[7]

Thyroid function tests are used for the diagnosis and to monitor treatment of thyroid disorders. A thyroid function test quantifies TSH and the circulating thyroid hormones in serum, to assess the ability of the thyroid gland to produce and regulate thyroid hormone production.[8] When TSH levels are above the reference range (i.e., >4.5mlU/L),[9] and concentrations of free thyroxine (fT4) are within the reference range (i.e., 5.0 to 12.0 ug/L),[10] the condition is called subclinical hypothyroidism (SCH).

SCH is a commonly diagnosed condition and ranges between 3% and 18% in the adult population; with aging, TSH levels increase and elderly are more often diagnosed with SCH.[1,11,12] Most patients with SCH do not have symptoms, have symptoms that are non-specific, or have symptoms similar to those observed in overt hypothyroidism.[13] Among these symptoms, global fatigue is the symptoms that triggers most of the thyroid hormone testing in general practice.[14–16] Over time, the condition of SCH may progress to overt hypothyroidism.[17] SCH has various causes. In up to 80% of the cases, the disorder is associated with antithyroid peroxidase antibodies, a marker of chronic lymphocytic (Hashimoto's) thyroiditis. Moreover, patients with treated overt thyroid failure often have SCH, caused by inadequate thyroid hormone supplementation, poor adherence, drug interactions, or inadequate monitoring of treatment.[12,18,19]

SCH can be treated with hormone replacement therapy using levothyroxine. Because SCH is often regarded as a solely biochemical abnormality, the need for lifelong levothyroxine replacement therapy is still under debate. Levothyroxine treatment is generally considered 1) to prevent progression to overt hypothyroidism, and 2) to reduce symptoms of thyroid hormone deficiency (which appears to be reversible in at least 25% of patients suffering from SCH) [20–23]

Several studies have revealed an increased risk for cardiovascular events in people with SCH, especially in younger SCH patients.[24–26] A meta-analysis of 15 observational studies showed that only

younger (i.e., <65 years) SCH patients had an increased risk of coronary heart disease (CHD) and cardiovascular mortality and an increased incidence of CHD was found also in those with serum TSH value <10 mU/l.[27] Beneficial effects of levothyroxine treatment in reducing these events have been reported in SCH patients.[1,11,23,24] However, other studies showed that the degree of serum TSH elevation and the age of the patient affect the levothyroxine treatment benefits.[3,4] Hence, the available evidence on the risk and benefits of levothyroxine treatment for SCH has not clearly been established.[1,4]

Guidelines recommend treating younger SCH patients with TSH levels >10 mU/L, even in the absence of symptoms. 'Young SCH patients' are defined as those aged less than 65 or 70, based on the clinical guideline.[3,14,28] In young SCH patients with mildly raised TSH levels (4.0-10.0 Um/L) and with thyroid-related symptoms treatment is considered.[3,14,28] In older SCH patients (>70 years of age), levothyroxine is only prescribed when clear symptoms of hypothyroidism and/or cardiovascular risk factors are present.[3,14,28] In randomised controlled trials (RCTs) evaluating the effects of levothyroxine treatment, the age and THS level thresholds applied as eligibility criteria vary and do not necessarily correspond with those in the guidelines.

4 Technology description

Levothyroxine is a synthetic version of the endogenous thyroid hormone thyroxine (T4). It is used as a hormone-replacement therapy to raise low levels of endogenous thyroid hormones.

In younger SCH patients with mildly raised TSH levels (4.0-10.0 Um/L) and with thyroid-related symptoms treatment is considered.[3,14,28] In older SCH patients, levothyroxine is only prescribed when clear symptoms of hypothyroidism and/or cardiovascular risk factors are present.[3,14,28] The usual starting dose is 25 to 50 mcg daily, with upward dose titration at 6-8 week intervals, until TSH levels are in the lower half of the reference range.[29] The average levothyroxine dose requirement in subclinical hypothyroidism is 0.5 mcg/kg/day.[29] For patients who are being monitored with sequential testing, TSH levels should be obtained at 6-12 month interval, or more frequently if new symptoms develop earlier.[29] Levothyroxine should be taken orally in the morning, ideally 30 minutes before breakfast.

In Switzerland, levothyroxine is approved by the Swiss Agency for Therapeutic Products (Swissmedic) for the treatment of hypothyroidism in adults and children. There is no standard alternative treatment option for patients diagnosed with SCH. Levothyroxine-containing medicines such as Eltroxin LF®, Euthyrox® and Tirosint® are listed on the "Spezialitätenliste"[30] and are reimbursed by the mandatory health insurance.

5 PICO

The Population Intervention Comparator Outcome (PICO) method is used to specify the research question. The PICO as pre-defined by the Federal Office of Public Health (FOPH) is further specified in this HTA protocol.

Table 1: PICO (Population - Intervention - Comparator - Outcome)

P:	Patients diagnosed with SCH
I:	Treatment with levothyroxine
C:	• Placebo • No treatment

O	**Efficacy/effectiveness outcomes:**
	• TSH levels
	• Symptoms of hypothyroidism (e.g., fatigue); measured as change in occurrence of symptoms or with a validated score (e.g., Hypothyroid Symptom Score, Tiredness Score)
	• Hypothyroidism-associated symptoms:
	  ○ Cardiovascular disorders (e.g., atrial fibrillation, heart failure, coronary artery disease)
	  ○ Musculoskeletal disorders (e.g., muscle cramps, bone fractures, osteoporosis)
	  ○ Neurological disorders (e.g., cognitive decline, depression)
	• Withdrawal of treatment due to lack of efficacy of levothyroxine
	• Health-related quality of life
	Safety outcomes:
	• Mortality
	• Serious and clinically important levothyroxine-associated adverse events
	Economic outcomes:
	• Life years [total and incremental]
	• Quality-adjusted life years [total and incremental]
	• Medical costs [total and incremental]
	• Incremental Cost-Effectiveness Ratio [ICER]
	• Budget impact

6 HTA key questions

For the evaluation of levothyroxine, the following key questions covering the central HTA domains will be addressed:

1. Is levothyroxine efficacious/effective compared to the placebo or no treatment in patients diagnosed with SCH?

2. Is levothyroxine safe compared to the placebo or no treatment in patients diagnosed with SCH?

3. What are the costs of levothyroxine treatment in patients diagnosed with SCH?

4. Is levothyroxine cost-effective compared to placebo or no treatment in patients diagnosed with SCH?

5. What is the budget impact of levothyroxine treatment in patients diagnosed with SCH?

6. Are there ethical, legal, social, or organisational issues related to levothyroxine treatment in patients diagnosed with SCH?

7 Methodology

In the following sections the systematic review (SR) methodology of both the efficacy/effectiveness/safety and the cost-effectiveness/budget impact of levothyroxine treatment in populations diagnosed with SCH are described in detail.

7.1 Efficacy, effectiveness, and safety

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). [31,32]

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

Within these fundamental steps, a stepwise approach will be implemented too:

I. To conduct a systematic literature search for original RCTs;

II. In case no RCT or only one RCT is found, an additional systematic literature search will be conducted for comparative non-randomised studies.

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 will be considerable overlap in studies included in other literature databases (such as Cochrane Library), the decision was made to search in these two main databases. The searches are built using the PICO-framework (see Chapter 5). Given the various outcomes of interest, it was decided to keep the search broad. Only search strings on 'patient' (i.e., patients diagnosed with SCH) and 'intervention' (i.e., treatment with levothyroxine) will be applied in combination with a search string for study design; RCTs (supplemented with a SR search string for finding possible missed relevant individual references, see section 7.1.1.4 Quality control) or comparative non-randomised studies. Furthermore, animal studies will be excluded with an additional search string. We will search in four languages: English, German, French, and Dutch. The publication period will be limited for 2000 to 2021. This time period is based on the publication dates found with a preliminary search for key articles. 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 objectives by the title and abstract. In this step, articles that seem to contain relevant data for the objectives 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 objectives and the methodological description and results section are of sufficient quality, based on the predefined inclusion and exclusion criteria (Chapter 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 PRISMA flow charts. The implemented quality control during the selection process is described in Chapter 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 for articles on levothyroxine treatment in patients diagnosed with SCH are presented in Table 2. The final list of applied inclusion and exclusion criteria will be presented in the HTA report.

Table 2. Inclusion and exclusion criteria for RCTs and comparative non-randomised studies

<table>
<tr><th></th><th colspan="2">Inclusion</th><th colspan="2">Exclusion</th></tr>
<tr><td>Period publication</td><td colspan="2">≥2000</td><td colspan="2"><2000</td></tr>
<tr><td>Language of publication</td><td colspan="2">• English
• German
• French
• Dutch</td><td colspan="2">All other languages</td></tr>
<tr><td>Country of study</td><td colspan="2">All countries</td><td colspan="2"></td></tr>
<tr><td rowspan="2">Study design/type</td><td rowspan="2">Step 1
• RCTs
• Open-label extension studies (i.e., of an RCT included with our systematic literature search)</td><td rowspan="2">Step 2
• Comparative non-randomised studies (e.g., cohort studies, case-control studies)</td><td colspan="2">• Systematic reviews
• Narrative reviews
• Non-comparative studies
• Case reports
• Abstract only (e.g., conference abstract)
• Non-pertinent publication types (e.g., expert opinion, letter to editor, editorial, comment)</td></tr>
<tr><td>Step 1
• Comparative non-randomised studies
• (Irrelevant) post-hoc/subgroup analysis of another RCT included in the systematic literature search</td><td>Step 2
• RCTs
• Open-label extension studies</td></tr>
<tr><td>Study quality</td><td colspan="2">• Sufficient methodological quality and coherent reporting of the results</td><td colspan="2">• Insufficient methodological quality or inco-</td></tr>
</table>

		herent reporting of the results (e.g., unexplained errors in patient flow)
		<u>Step 1</u> • Cross-over trial without washout period
Study population	Patients diagnosed with SCH	• Studies only including women with specific female sex hormonal states, e.g. pregnant women, non-pregnant women on fertility-related treatment, or menopausal women • Congenital hypothyroidism • Studies that include only patients diagnosed with SCH in combination with another specific condition (e.g., patients with diabetic nephropathy)
Study intervention	• Levothyroxine treatment	• Any other intervention than levothyroxine • Short treatment or follow-up duration (to be determined in further detail during the project)
Study comparison	• Placebo • No treatment	• Comparison of different brands or doses of levothyroxine • Comparison with other drug-treatment • No comparison
Study outcomes	See PICO table†	• Other outcomes

Keys: PICO = Population-Intervention-Comparator-Outcome, RCT = randomised controlled trial, SCH = subclinical hypothyroidism, 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 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 of the Erasmus Medical Centre Rotterdam was consulted during the development of the search strategies. Quality checks were implemented and the search strategy was checked by a second researcher.
    - 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 first 10% of the full-text articles from the peer-reviewed literature will be assessed for relevancy and critically appraised in duplicate by two independent researchers. The results will be compared and discussed early in the process. If there are differences between the two researchers with regard to more than 5% of the articles screened in duplicate, another 10% of the articles will be screened in duplicate. This will be repeated if necessary. If there is still more than 5% discrepancy at 50% of the duplicate selection, the screening of full-text articles will be done fully in duplicate by two independent researchers. The remaining full-text selection is done by one researcher in close collaboration with a second reviewer; any doubts are discussed in detail. In case of discrepancy or disagreements during the selection phase, a third researcher is consulted. The study is discussed until consensus was reached.

7.1.2 Data extraction, analysis, and synthesis

7.1.2.1 Data extraction

Relevant data from the included RCTs 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

Based on the key risk of bias criteria used in the Grading of Recommendations, Assessment, Development and Evaluations (GRADE) approach, the risk of bias of the study designs of the included RCTs will be assessed.[34] These key study limitations or risk of bias of RCTs include:

- Lack of allocation concealment (i.e., those enrolling patients are aware of the study arm or period to which the next enrolled patient will be allocated, e.g., based on birth date or chart number)
- Lack of blinding (i.e., patient, caregivers, those recording outcomes, those adjudicating outcomes, or data analysts are aware of the arm to which patients are allocated)
- Incomplete accounting of patients and outcome events:
  - Loss to follow-up (i.e., the significance of particular rates of loss to follow-up varies widely and is dependent on the relation between loss to follow-up and number of events; the higher the proportion lost to follow-up in relation to intervention and control arm event rates, and differences between intervention and control arm, the greater the threat of bias)
  - Intention to treat (i.e., failure to adhere to the intention-to-treat principle)
- Selective outcome reporting (i.e., incomplete or absent reporting of some outcomes and not others on the basis of the results)
- Other limitations (e.g., use of unvalidated outcome measures; carryover effects in crossover trial)

Each risk of bias criterion of the included RCTs will be rated as low risk of bias, moderate or unclear (i.e., not reported in the article) risk of bias, or high risk of bias. Based on the crucial limitations for one or more of these criteria, the risk of bias of the study design within the whole study will be rated in one of the three categories: low risk of bias, moderate risk of bias, or high risk of bias.

As suggested by GRADE, for comparative non-randomised studies the ROBINS-I tool (i.e., the risk of bias in non-randomised studies of interventions tool) will be applied. The key study limitations or risk of bias for comparative non-randomised studies include:

- Bias due to confounding
- Bias in selection of participants into the study
- Bias in classification of interventions
- Bias due to deviations from intended interventions
- Bias due to missing data
- Bias in measurement of the outcome
- Bias in selection of the reported result

7.1.2.3 Data Synthesis

The options for clinically relevant stratification of the data by gender, age and/or TSH levels will be investigated and discussed with clinical experts. 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 and 2) sufficient data is reported in the studies. Considering the heterogeneity in the data, a random-effects model (DerSimonian & Laird) will be considered for the analyses. 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, when possible.

For outcomes for which it is possible to calculate pooled estimates, a GRADE assessment for the level of the quality or certainty of the evidence on outcome level will be implemented. The GRADE approach is a system for rating the certainty of a body of evidence in SRs, which for a specific outcome is rated across studies instead of a quality assessment of individual studies.[34] The certainty of the evidence is assessed by looking at the following features of the evidence found for each outcome:

- **Study limitations (risk of bias)** – the 'internal validity' of the evidence
- **Inconsistency** – the heterogeneity or variability in the estimates of treatment effect across studies
- **Indirectness** – the degree of differences between the population, intervention, comparator for the intervention, and outcome of interest across studies
- **Imprecision** (random error) – the extent to which confidence in the effect estimate is adequate to support a particular decision
- **Publication bias/other considerations** – the degree of selective publication of studies

The certainty of the evidence is classified as high, moderate, low, or very low:

- **High** – further research is very unlikely to change our confidence in the estimate of effect
- **Moderate** – further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate
- **Low** – further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate
- **Very low** – any estimate of effect is very uncertain

The evidence on selected outcomes will be summarised in GRADE evidence profiles. Evidence profiles will include information about the body of evidence (i.e., number of studies including references, study design), the quality assessment (i.e., risk of bias, inconsistency, indirectness, imprecision, publication bias/other considerations), number of patients in the intervention and control group, effect, and a grade for the quality of evidence.

Outcomes for which it is not possible to calculate pooled estimates and implement a GRADE assessment for the level of evidence on outcome level will be presented narratively in summary tables, grouped by outcomes, and accompanying text, to provide insight into the direction of the treatment effects found in the clinical literature. 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 Cost-effectiveness

7.2.1 Databases and search strategy

7.2.1.1 Search strategy

The cost-effectiveness systematic literature search follows 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, economic databases, using the Cost Effectiveness Analysis (CEA) Registry and NHS EED, 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 efficacy, effectiveness, and safety literature search were combined with cost-effectiveness search terms. The details of the search strategy are presented in Appendix 9.2.

7.2.1.2 Selection procedure

All articles retrieved from PubMed (MEDLINE), Embase.com, NHS EED and the CEA registry databases and relevant references will be selected by a two-step selection procedure, based on:

1. Screening of title and abstract:

The major topics of the articles will be assessed based on relevancy for the objectives by the title and abstract. In this step, articles that seem to contain relevant data for the HTA objectives 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.

2. Screening of full-text article:

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 the HTA protocol. 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 Endnote version 20. 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 3. The final list of the inclusion and exclusion criteria will be presented in the HTA report.

Table 3. Inclusion and exclusion criteria for cost-effectiveness studies

	Inclusion	**Exclusion**
Period publication	No restriction/ from database inception	
Country of study	All countries	
Study design/type	Economic evaluations • Cost-utility analysis • Cost-effectiveness analysis • Cost-minimisation analysis • Cost-benefit analysis	Resource use measurement

	• Budget impact analysis • Costing studies	
Study population	Patients diagnosed with SCH	• Studies only including women with specific female sex hormonal states, e.g. pregnant women, non-pregnant women on fertility-related treatment, or menopausal women • Congenital hypothyroidism • Studies that include only patients diagnosed with SCH in combination with another specific condition (e.g., patients with diabetic nephropathy)
Study intervention	Levothyroxine	Any other intervention than levothyroxine
Study comparison	Placebo or no treatment	Comparison of different brands or doses of levothyroxine
Study outcomes	See PICO table‡	Other outcomes

Keys: PICO = Population-Intervention-Comparator-Outcome, SCH = subclinical hypothyroidism, TSH = thyroid-stimulating hormone. ‡See PICO in Chapter 5.

7.2.1.4 Quality control

- Search strategy

Search strategy was checked by a second researcher. The supplementary search technique citation chasing (i.e., studies cited within the included articles) was 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
- Study population
- Intervention
- Comparator
- Outcome measures
- Total/Incremental costs and QALYs
- Time Horizon
- Model used (Yes/No)
    - Type of model
    - Health states
- Primary sources for the resource use/cost inputs
- Primary sources for the health-related quality of life inputs
- Discount rates, study perspective, cycle length, half-cycle correction, study funding source

7.2.2.2 Critical appraisal

Critical appraisal of the included economic evaluations will be performed in the HTA phase using the Consensus Health Economic Criteria (CHEC).[35,36] The aim of the CHEC is to assess the methodological quality of economic evaluations in systematic reviews.[35,36] In addition to the CHEC checklist, questions specific to the decision problem (e.g., adverse events) will be considered.

7.2.2.3 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 incremental cost-effectiveness ratios will be presented and the reliability (internal validity) and relevance (generalisability) of the estimates will be explored. The analytical approaches used in the studies will be compared and their robustness will be discussed. If the published studies on cost-effectiveness of levothyroxine compared to placebo or no treatment for patients with SCH do not provide sufficient information to draw firm conclusions on the cost-effectiveness in Switzerland, a *de novo* health economic model will be developed in the HTA phase. The general approach to health economic modelling will be further specified in the HTA report and is summarised in Chapter 7.2.2.5.

7.2.2.4 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 independently 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.

7.2.2.5 Cost-effectiveness and budget impact

In this section, general steps in developing a *de novo* health economic model are described, including conceptual model development, selection of appropriate model type, main characteristics of the health economic model, model input parameters, and budget impact analyses.

- ***Conceptual health economic model development***

The decision problem and the PICO that were described in Chapter 5 will be used for the development of the conceptual health economic 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 costs and health effects as well as their relation. According to the ISPOR-SMDM good modelling research practice,[37] a careful review of existing health economic models, when available, is a prerequisite to develop the underlying conceptual model. Included economic evaluations will be used as a starting point for the development of the conceptual health economic model. The draft conceptual model will be discussed with a clinical expert to ensure that the model structure reflects clinical practice.

- ***Model selection***

The design of the conceptual model will derive the choices between different model types (e.g., a decision tree, cohort level Markov model or patient level simulation model). For example, a decision tree is suitable for decision problems with a relatively short and fixed time horizons, with no recurring events. For longer time horizons, when timing of events and/or their recurrence need to be considered, a cohort level Markov model is more suitable. If it is important to consider a patient's event history, a patient level simulation model is considered. A patient simulation model can incorporate recurrence of events and account for patient's event history, without leading to excessive number of health states. However, patient level simulation models are also more demanding in terms of data requirements. Hence, the translation of the conceptual model to a specific type of health economic model depends on the available data that will be extracted during the HTA phase.

- ***Main characteristics of the health economic model***

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 lifetables for background mortality and Swiss sources for healthcare costs).

Perspective: The analysis will be performed from a 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 preferred time horizon of the base-case analysis is lifetime. A lifetime horizon will depend on the feasibility and availability of data. Shorter time horizons will be considered in scenario analyses, where possible or relevant.

Discount rate: In the base-case analysis, healthcare costs and effects will be discounted at 3.0%. In scenario analyses, the impact of not using a discount rate or a discount rate of 5.0% will be explored.

Health outcomes: Health outcomes will be reported in life years and quality-adjusted life years (QALYs).

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 OECD website (https://data.oecd.org).

- ***Model input parameters***

Model input parameters on clinical outcomes will be informed mainly from the results of the data extraction of the systematic literature search of efficacy, effectiveness, safety and cost-effectiveness. Clinical expert opinion will be used whenever data will not be available from the literature. In the case of no Swiss-specific data on costs, resource use and utilities are identified in the systematic literature searches for efficacy, effectiveness and safety and cost-effectiveness described in this report, additional searches will be performed. Additional searches on medical databases will then be performed in collaboration with the FOPH to determine medication use, healthcare resource use, and unit costs. If the required data cannot be identified from the medical database and systematic literature searches, assumptions will be made based on data from other comparable countries and/or expert opinion.

- ***Health economic model and cost-effectiveness analysis***

Following the conceptual model development and data collection, the health economic model will be populated with the obtained parameters. The model will be programmed in statistical programming language R based on the modelling framework developed by the Decision Analysis in R for Technologies in Health (DARTH) workgroup[38–40] Parametric uncertainty will be explored through one-way and probabilistic sensitivity analyses and structural uncertainty will be explored in several scenario analyses. Subgroup analyses will be performed, where possible, to differentiate cost-effectiveness by population or disease/indication area.

- ***Budget impact analysis***

In addition to the cost-effectiveness model, a budget impact model (BI model) will be developed to calculate the projected population-level overall costs of levothyroxine for patients diagnosed with SCH. The BI model will be built as an extension to the cost-effectiveness model, described above. Hence, the core model characteristics for the BI model will be dependent on the cost-effectiveness model. The time horizon of the BI model will be restricted to 5 years. For the BI model, epidemiological data will be required about the current use of levothyroxine for patients with SCH in Switzerland. If this data will not be available, assumptions will be made based on data from other comparable countries and/or expert opinion.

7.2.2.6 Legal, social, ethical, and organisational issues

In the HTA report, we will identify and report the main legal, social, ethical, and organisational issues of the studies included in the efficacy, effectiveness, and safety and cost-effectiveness systematic literature searches. In addition, we will perform grey literature searches on these HTA domains.

8 References

1 Javed Z, Sathyapalan T. Levothyroxine treatment of mild subclinical hypothyroidism: a review of potential risks and benefits. *Ther Adv Endocrinol Metab* 2016;7:12–23. doi:10.1177/2042018815616543

2 Rodondi N, den Elzen WPJ, Bauer DC, *et al.* Subclinical hypothyroidism and the risk of coronary heart disease and mortality. *JAMA* 2010;304:1365–74. doi:10.1001/jama.2010.1361

3 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

4 Stott DJ, Rodondi N, Kearney PM, *et al.* Thyroid hormone therapy for older adults with subclinical hypothyroidism. *New England Journal of Medicine* 2017;376:2534–44.

5 Ortiga-Carvalho TM, Chiamolera MI, Pazos-Moura CC, *et al.* Hypothalamus-Pituitary-Thyroid Axis. *Compr Physiol* 2016;6:1387–428. doi:10.1002/cphy.c150027

6 Birtwhistle R, Morissette K, Dickinson JA, *et al.* Recommendation on screening adults for asymptomatic thyroid dysfunction in primary care. *CMAJ* 2019;191:E1274–80. doi:10.1503/cmaj.190395

7 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

8 NICE Guideline. Thyroid Disease: assessment and management. 2019.https://www.nice.org.uk/guidance/ng145/resources/thyroid-disease-assessment-and-management-pdf-66141781496773

9 Rugge B, Balshem H, Sehgal R. Screening and Treatment of Subclinical Hypothyroidism or Hyperthyroidism. Comparative Effectiveness Review No. 24. Agency for Healthcare Research and Quality. 2011. www.effectivehealthcare.ahrq.gov/reports/final.cfm.

10 UCLA H. UCLA Endocrine Center Thyroid Hormone Levels. https://www.uclahealth.org/endocrine-center/

11 Baumgartner C, Blum MR, Rodondi N. Subclinical hypothyroidism: summary of evidence in 2014. *Swiss Med Wkly* 2014;144:w14058. doi:10.4414/smw.2014.14058

12 Canaris GJ, Manowitz NR, Mayor G, *et al.* The Colorado thyroid disease prevalence study. *Arch Intern Med* 2000;160:526–34. doi:10.1001/archinte.160.4.526

13 Cooper DS, Biondi B. Subclinical thyroid disease. *Lancet* 2012;379:1142–54. doi:10.1016/S0140-6736(11)60276-6

14 Garber JR, Cobin RH, Gharib H, *et al.* Clinical practice guidelines for hypothyroidism in adults: cosponsored by the American Association of Clinical Endocrinologists and the American Thyroid Association. *Endocr Pract* 2012;18:988–1028. doi:10.4158/EP12280.GL

15 Allport J, McCahon D, Hobbs FDR, *et al.* Why are GPs treating subclinical hypothyroidism? Case note review and GP survey. *Prim Health Care Res Dev* 2013;14:175–84. doi:10.1017/S1463423612000230

16 Cooper R, Pinkney J, Ayling RM. Appropriateness of prescribing thyroxine in primary care. *Ann Clin Biochem* 2015;52:497–501. doi:10.1177/0004563214568686

17 Villar HCCE, Saconato H, Valente O, *et al.* Thyroid hormone replacement for subclinical hypothyroidism. *Cochrane Database Syst Rev* 2007;:CD003419. doi:10.1002/14651858.CD003419.pub2

18 Somwaru LL, Arnold AM, Joshi N, *et al.* High frequency of and factors associated with thyroid hormone over-replacement and under-replacement in men and women aged 65 and over. *J Clin Endocrinol Metab* 2009;94:1342–5. doi:10.1210/jc.2008-1696

19 Hollowell JG, Staehling NW, Flanders WD, *et al.* Serum TSH, T(4), and thyroid antibodies in the United States population (1988 to 1994): National Health and Nutrition Examination Survey (NHANES III). *J Clin Endocrinol Metab* 2002;87:489–99. doi:10.1210/jcem.87.2.8182

20 Cooper DS, Halpern R, Wood LC, *et al.* L-Thyroxine therapy in subclinical hypothyroidism. A double-blind, placebo-controlled trial. *Ann Intern Med* 1984;101:18–24. doi:10.7326/0003-4819-101-1-18

21 Nyström E, Caidahl K, Fager G, *et al.* A double-blind cross-over 12-month study of L-thyroxine treatment of women with 'subclinical' hypothyroidism. *Clin Endocrinol (Oxf)* 1988;29:63–75. doi:10.1111/j.1365-2265.1988.tb00250.x

22 Staub JJ, Althaus BU, Engler H, *et al.* Spectrum of subclinical and overt hypothyroidism: effect on thyrotropin, prolactin, and thyroid reserve, and metabolic impact on peripheral target tissues. *Am J Med* 1992;92:631–42. doi:10.1016/0002-9343(92)90782-7

23 Monzanil F, Guerra P, Caracciol N, *et al.* Subclinical hypothyroidism: neurobehavioral features and beneficial effect of l-thyroxine treatment. *Clin Investig* 1993;71. doi:10.1007/BF00186625

24 Biondi B. Cardiovascular effects of mild hypothyroidism. *Thyroid* 2007;17:625–30. doi:10.1089/thy.2007.0158

25 Brenta G, Mutti LA, Schnitman M, *et al.* Assessment of left ventricular diastolic function by radionuclide ventriculography at rest and exercise in subclinical hypothyroidism, and its response to L-thyroxine therapy. *Am J Cardiol* 2003;91:1327–30. doi:10.1016/s0002-9149(03)00322-9

26 Owen PJD, Sabit R, Lazarus JH. Thyroid disease and vascular function. *Thyroid* 2007;17:519–24. doi:10.1089/thy.2007.0051

27 Razvi S, Shakoor A, Vanderpump M, *et al.* The influence of age on the relationship between subclinical hypothyroidism and ischemic heart disease: a metaanalysis. *J Clin Endocrinol Metab* 2008;93:2998–3007. doi:10.1210/jc.2008-0167

28 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

29 McDermott MT. Hypothyroidism. *Annals of Internal Medicine* 2020;173:ITC1–16. doi:10.7326/AITC202007070

30 Spezialitätenliste Bundesamt für Gesundheit BAG. https://www.bag.admin.ch/bag/de/home/begriffe-a-z/spezialitaetenliste.html

31 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

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

33 Feller M, Snel M, Moutzouri E, *et al.* Association of thyroid hormone therapy with quality of life and thyroid-related symptoms in patients with subclinical hypothyroidism: a systematic review and meta-analysis. *Jama* 2018;320:1349–59.

34 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

35 Evers S, Goossens M, de Vet H, *et al.* Criteria list for assessment of methodological quality of economic evaluations: Consensus on Health Economic Criteria. *Int J Technol Assess Health Care* 2005;21:240–5.

36 Watts RD, Li IW. Use of Checklists in Reviews of Health Economic Evaluations, 2010 to 2018. *Value Health* 2019;22:377–82. doi:10.1016/j.jval.2018.10.006

37 Roberts M, Russell LB, Paltiel AD, *et al.* Conceptualizing a Model: A Report of the ISPOR-SMDM Modeling Good Research Practices Task Force–2. *Med Decis Making* 2012;32:678–89. doi:10.1177/0272989X12454941

38 Jalal H, Pechlivanoglou P, Krijkamp E, *et al.* An Overview of R in Health Decision Sciences. *Med Decis Making* 2017;37:735–46. doi:10.1177/0272989X16686559

39 Krijkamp EM, Alarid-Escudero F, Enns EA, *et al.* Microsimulation Modeling for Health Decision Sciences Using R: A Tutorial. *Med Decis Making* 2018;38:400–22. doi:10.1177/0272989X18754513

40 Krijkamp EM, Alarid-Escudero F, Enns EA, *et al.* A Multidimensional Array Representation of State-Transition Model Dynamics. *Med Decis Making* 2020;40:242–8. doi:10.1177/0272989X19893973

9 Appendices

Appendix 9.1. Search strategy efficacy, effectiveness, and safety

Table I. Search strategy for the efficacy, effectiveness, and safety systematic literature search for RCTs: PubMed (MEDLINE)

Population: diagnosed with SCH	("hypothyroidism"[mesh] OR hypothyroidism*[tiab] OR thyroid disease*[tiab] OR thyroid deficien*[tiab] OR thyroid insufficien*[tiab] OR thyroid stimulating hormone deficienc*[tiab] OR TSH deficienc*[tiab]) AND (mild[tiab] OR subclinic*[tiab] OR sub-clinic*[tiab] OR moderate[tiab])	
Intervention: levothyroxine	"thyroxine"[mesh] OR thyroxine[tiab] OR levothyroxine[tiab] OR L-thyroxin*[tiab] OR hormone replacement therap*[tiab]	
Comparison	No search string	
Outcomes	No search string	
Limits	*Study design*	
	RCTs and SRs: "randomized controlled trial"[pt] OR "controlled clinical trial"[pt] OR RCT[tiab] OR RCTs[tiab] OR random*[tiab] OR controlled[tiab] OR control-treated[tiab] OR placebo[tiab] OR “cross-over studies“[Mesh] OR "single-blind method"[Mesh] OR single-blind*[tiab] OR singleblind*[tiab] OR single-masked[tiab] OR “double-blind method“[Mesh] OR double-blind*[tiab] OR doubleblind*[tiab] OR double-masked[tiab] OR triple-blind*[tiab] OR tripleblind*[tiab] OR triple-masked[tiab] OR (((systematic*[tiab]	*Optional; Comparative non-randomised studies:* (nonrandomized[tiab] OR non-randomized[tiab] OR nonrandomised[tiab] OR non-randomised[tiab] OR quasiexperimental[tiab] OR quasi-experimental[tiab] OR non-equivalent control*[tiab] OR non-equivalent control*[tiab] OR "cohort studies"[Mesh] OR prospective*[tiab] OR retrospective*[tiab] OR follow-up stud*[tiab] OR followup stud*[tiab] OR longitudinal stud*[tiab] OR cohort[tiab] OR "comparative effectiveness research"[Mesh] OR comparative effectiveness[tiab] OR real-world[tiab]

	OR comprehensive*[tiab]) AND (biblio-graphic*[tiab] OR literature[tiab] OR re-view*[tiab])) OR literature review*[tiab] OR meta-analysis[pt] OR meta-analys*[tiab] OR meta-analyz*[tiab] OR meta-analyt*[tiab] OR metaanalys*[tiab] OR metaanalyz*[tiab] OR metaanalyt*[tiab])	OR real-life[tiab] OR “case-control stud-ies”[Mesh] OR case-control[tiab] OR case-control[tiab] OR case-comparison[tiab] OR case-referent[tiab])
	Publication period: 2000 – current	
	Language: English, German, French, Dutch	

Table II. Search strategy for the efficacy, effectiveness, and safety systematic literature search for RCTs: Embase.com

Population: diagnosed with SCH	('hypothyroidism'/exp OR hypothyroidism*:ti,ab OR "thyroid disease*":ti,ab OR "thyroid deficien*":ti,ab OR "thyroid insufficien*":ti,ab OR "thyroid stimulating hormone deficienc*":ti,ab OR "TSH deficienc*":ti,ab) AND (mild:ti,ab OR subclinic*:ti,ab OR sub-clinic*:ti,ab OR moderate:ti,ab)	
Intervention: levothyroxine	'thyroxine'/exp OR thyroxine:ti,ab OR levothyroxine:ti,ab OR L-thyroxin*:ti,ab OR "hormone replacement therap*":ti,ab	
Comparison	No search string	
Outcomes	No search string	
Limits	*Study design*	
	RCTs and SRs: 'randomized controlled trial'/exp OR 'controlled clinical trial'/exp OR RCT:ti,ab OR RCTs:ti,ab OR random*:ti,ab OR controlled:ti,ab OR control-treated:ti,ab OR placebo:ti,ab OR 'crossover procedure'/exp OR 'single blind procedure'/exp OR single-blind*:ti,ab OR singleblind*:ti,ab OR single-masked:ti,ab OR 'double blind procedure'/exp OR double-blind*:ti,ab OR double-blind*:ti,ab OR double-masked:ti,ab OR 'triple blind procedure'/exp OR triple-blind*:ti,ab OR tripleblind*:ti,ab OR triple-masked:ti,ab OR (((systematic*:ti,ab OR comprehensive*:ti,ab) AND (bibliographic*:ti,ab OR literature:ti,ab OR review*:ti,ab)) OR "literature review*":ti,ab OR 'meta analysis'/exp OR meta-analys*:ti,ab OR meta-analyz*:ti,ab OR meta-analyt*:ti,ab OR metaanalys*:ti,ab OR metaanalyz*:ti,ab OR metaanalyt*:ti,ab)	*Optional; Comparative non-randomised studies:* (nonrandomized:ti,ab OR non-randomized:ti,ab OR nonrandomised:ti,ab OR non-randomised:ti,ab OR quasiexperimental:ti,ab OR quasi-experimental:ti,ab OR "non-equivalent control*":ti,ab OR "non-equivalent control*":ti,ab OR 'cohort analysis'/exp OR prospective*:ti,ab OR retrospective*:ti,ab OR "follow-up stud*":ti,ab OR "followup stud*":ti,ab OR "longitudinal stud*":ti,ab OR cohort:ti,ab OR 'comparative effectiveness'/exp OR "comparative effectiveness":ti,ab OR real-world:ti,ab OR real-life:ti,ab OR 'case control study'/exp OR case-control:ti,ab OR casecontrol:ti,ab OR case-comparison:ti,ab OR case-referent:ti,ab)
	Publication period: 2000 – current	

	Language: English, German, French, Dutch

Appendix 9.2. Search strategy cost-effectiveness

Table I. Search strategy for cost-effectiveness systematic literature search: PubMed (MEDLINE)

Population: diagnosed with SCH	("hypothyroidism"[mesh] OR hypothyroidism*[tiab] OR thyroid disease*[tiab] OR thyroid deficien*[tiab] OR thyroid insufficien*[tiab] OR thyroid stimulating hormone deficienc*[tiab] OR TSH deficienc*[tiab]) AND (mild[tiab] OR subclinic*[tiab] OR sub-clinic*[tiab] OR moderate[tiab])
Intervention: Levothyroxine	"thyroxine"[mesh] OR thyroxine[tiab] OR levothyroxine[tiab] OR L-thyroxin*[tiab] OR hormone replacement therap*[tiab]
Comparison	No search string
Outcomes	No search string
Limits	*Study design economic evaluations:* ("Technology Assessment, Biomedical"[Mesh] OR "Cost-Benefit Analysis"[Mesh] OR "Quality-Adjusted Life Years"[Mesh] OR "technology assessment" [tiab] OR "economic evaluation" [tiab] OR "economic value" [tiab] OR "cost-benefit" [tiab] OR "cost-effective" [tiab] OR "cost-effectiveness" [tiab] OR "cost-utility" [tiab] OR "cost-consequence" [tiab] OR "quality-adjusted life year" [tiab] OR "QALY" [tiab])
	Publication period: No restrictions
	Language: No restrictions

Table II. Search strategy for cost-effectiveness systematic literature search: Embase.com

Population: diagnosed with SCH	('hypothyroidism'/exp OR hypothyroidism*:ti,ab OR "thyroid disease*":ti,ab OR "thyroid deficien*":ti,ab OR "thyroid insufficien*":ti,ab OR "thyroid stimulating hormone deficienc*":ti,ab OR "TSH deficienc*":ti,ab) AND (mild:ti,ab OR subclinic*:ti,ab OR subclinic*:ti,ab OR moderate:ti,ab)
Intervention: levothyroxine	'thyroxine'/exp OR thyroxine:ti,ab OR levothyroxine:ti,ab OR L-thyroxin*:ti,ab OR "hormone replacement therap*":ti,ab
Comparison	No search string
Outcomes	No search string
Limits	*Study design economic evaluations:* ('biomedical technology assessment'/exp OR 'economic evaluation'/exp OR 'quality adjusted life year'/exp OR 'program cost effectiveness'/de OR ((tech-nology NEAR/3 assessment*) OR (economic* NEAR/3 (evaluat* OR value)) OR ((cost OR costs) NEAR/3 (benefit* OR effectiv* OR efficien* OR efficac* OR minim* OR utilit* OR consequen*)) OR (qualit* NEAR/3 adjust* NEAR/3 (life-year* OR lifeyear*)) OR qaly*):ab,ti)
	Publication period: No restrictions
	Language: No restrictions

Table III. Search strategy for cost-effectiveness systematic literature search NHS EED & CEA Registry

Database	NHS EED	The CEA Registry
Population: diagnosed with SCH	(hypothyroidism)	(hypothyroidism)
Intervention: levothyroxine	(l-thyroxine) OR (levothyroxine)	(l-thyroxine) OR (levothyroxine)
Comparison	No search string	No search string
Outcomes	No search string	No search string
Limits	No limits	No limits