Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Federal Department of Home Affairs
Federal Office of Public Health FOPH
Health and Accident Insurance Directorate
Section Health Technology Assessment

Health Technology Assessment (HTA)

HTA Report

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** Pallas Health Research and Consultancy, Rotterdam, the Netherlands
Technology	Levothyroxine
Type of Technology	Pharmaceuticals
Date	29 August 2022

Bundesamt für Gesundheit
Sektion Health Technology Assessment
Schwarzenburgstrasse 157
CH-3003 Bern
Schweiz
Tel.: +41 58 462 92 30
E-mail: hta@bag.admin.ch

Executive Summary

BACKGROUND: 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. 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.

OBJECTIVE: The aim of this health technology assessment (HTA) report is to evaluate the clinical and economic consequences of levothyroxine treatment for patients diagnosed with SCH in Switzerland.

METHODS: For the clinical review, a systematic literature search of the Pubmed (MEDLINE) and Embase databases was conducted adhering to international methodological standards. Randomized controlled trials (RCTs) in patients diagnosed with SCH who are treated with levothyroxine compared with no treatment or placebo were included. Articles including only women with specific sex hormonal state were out of scope. The search was worldwide and restricted to RCTs published from 2000 onwards. The included studies were critically appraised, and the extracted data was summarised in summary tables and narrative text.

For the economic review, cost-effectiveness searches followed the principles of the clinical systematic literature search. A *de novo* model was developed to calculate the budget impact of levothyroxine for adult SCH patients, between 18 and 99 years of age, in Switzerland over a time horizon of five years (years 2020-2024).

Legal, social, ethical, and organisational issues addressed in the studies included in the clinical, cost-effectiveness systematic literature searches were extracted and narratively summarised. In addition, grey literature searches were conducted on this HTA domain.

RESULTS: Evidence from 28 RCTs based on 22 unique studies was evaluated to inform efficacy and safety outcomes of levothyroxine compared to placebo (n=24 articles) or no treatment (n=4 articles) for the treatment of SCH. Twenty-five RCTs reported TSH level changes and 12 studies reported additional clinical efficacy and safety outcomes. The body of evidence was substantial, and the risk of bias ranged from low to high.

Levothyroxine successfully reduced increased levels of TSH in SCH patients. However, levothyroxine treatment did not affect any of the other clinically and patient-relevant efficacy

outcomes that were assessed. The evidence base did not show consistent statistically significant changes between the placebo and levothyroxine treated patients in hypothyroid associated symptoms like fatigue, depression, health-related quality of life, neurological, musculoskeletal or cardiovascular outcomes. Some positive trends were reported in individual studies, but these findings were typically inconsistent between studies and not statistically significant. In none of the studies major safety issues were observed.

No published study was identified from the cost-effectiveness systematic review. In the budget impact analysis, we calculated the total healthcare cost of levothyroxine for adult SCH patients between ages 18 and 99 years, in Switzerland over a time horizon of five years (years 2020-2024). In the absence of Swiss epidemiological data and on the percentage of SCH patients receiving levothyroxine, we estimated the number of patients receiving levothyroxine for SCH based on general population estimates from the Swiss FSO, and European SCH prevalence rates from the literature. Subsequently we used a percentage range (10%-90%) of the prevalent population that would be eligible for levothyroxine in Switzerland. Costs for the treatment with levothyroxine were calculated as a function of drug unit prices and a standard dosing schedule.

The total costs of treatment ranged between 9 Mio CHF and 87 Mio CHF over a time horizon of five years, depending on the percentage of patients that would be treated with levothyroxine.

The literature search did not provide evidence indicating that levothyroxine treatment in SCH patients was associated with any important ethical, legal, social or organisation issues.

CONCLUSION:

Levothyroxine treatment successfully lowers raised levels of serum TSH in SCH patients but evidence of a consistent (positive) treatment effect on other clinical or patient-relevant outcomes like hypothyroid associated symptoms or cardiovascular outcomes is lacking. There are no indications of major safety issues related to levothyroxine treatment. The budget impact analysis showed that depending on the prescription behaviour of physicians the costs of levothyroxine for SCH patients can be substantial (up to 87 Mio CHF over 5 years).

Table of Contents

1 Policy question and context 9

2 Research question 10

3 Medical background 11

4 Technology 15

5 Population, Intervention, Comparator, Outcome (PICO) 15

6 HTA key questions 16

6.1 Additional question(s) 16

7 Efficacy, effectiveness, and safety 17

7.1 Methodology efficacy, effectiveness, and safety 18

7.2 Results efficacy, effectiveness, and safety 24

8 Costs, cost-effectiveness and budget impact 49

8.1 Methodology costs, cost-effectiveness and budget impact 49

8.2 Results costs, cost-effectiveness and budget impact 53

9 Ethical, legal, social and organisational issues 59

9.1 Methodology ethical, legal, social and organisational issues 59

9.2 Results ethical, legal, social and organisational issues 60

10 Additional issues 62

10.1 Guideline Recommendations 62

10.2 Ongoing Trials 63

11 Discussion 64

12 Conclusions 67

13 References 68

14 Appendices 74

14.1 Search strategy efficacy, effectiveness, and safety 74

14.2 Excluded studies during full text selection 76

14.3 Efficacy results 81

14.4 Search strategy cost-effectiveness analysis 112

Abbreviations and acronyms

ATA	American Thyroid Association
BDI	Beck Depression Inventory
CalCAP	California Computerised Assessment Package
CEA	Cost Effectiveness Analysis
CESD	Center for Epidemiologic Studies Depression
CHEC	Consensus Health Economic Criteria
CHF	Swiss Franc
CHD	Coronary Heart Disease
CI	Confidence interval
COGE	Controlling Gesundheitswesen
CV	Cardiovascular
DARTH	Decision Analysis in R for Technologies in Health
DDD	Defined daily dose
EAE	Effectiveness, appropriateness and economic efficiency
ETA	European Thyroid Association
EUnetHTA	European Network for Health Technology Assessment
EQ-5D	EuroQol-5 Dimension
FOPH	Federal Office of Public Health
FSO	Federal Statistical Office
GDS	Geriatric depression scale
GHQ	General health questionnaire
GRADE	Grading of Recommendations, Assessment, Development and Evaluations
HADS	Hospital Anxiety and Depression Scale
HR	Hazard ratio
HRQoL	Health-related quality of life
HTA	Health Technology Assessment
ICER	Incremental Cost Effectiveness Ratio
IEMO80+	Institute for Evidence-Based Medicine in Old Age 80-plus thyroid trial
ISPOR	International Society for Pharmacoeconomics and Outcomes Research
MEAMS	Middlesex Elderly Assessment of Mental State
MMSE	Mini-Mental State Examination
N.A.	Not Applicable

NHS EED	National Health Service Economic Evaluation Database
OECD	Organisation for Economic Cooperation and Development
PFS	Pittsburgh Fatigability Scale
PICO	Population, Intervention, Comparator, Outcome
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
QALY	Quality Adjusted Life Year
RCT	Randomised Controlled Trial
SCH	Subclinical hypothyroidism
SCOLP	Speed and Capacity of Language Processing
SF-36	36-item short form survey
SMDM	Society for Medical Decision Making
SR	Systematic Review
T3	Thyroxine
T4	Triiodothyronine
ThyPRO	Thyroid-specific patient reported outcome
TRUST	Thyroid Hormone Replacement for Subclinical Hypothyroidism Trial
TSH	Thyroid-Stimulating Hormone

Objective of the HTA report

The objective of a health technology assessment (HTA) is to generate a focused assessment of various aspects of a health technology. The analytic methods applied to assess the value of using a health technology, their execution, and the results are described. The analytical process is comparative, systematic, transparent, and involves multiple stakeholders. The domains covered in an HTA report include clinical efficacy, effectiveness, safety, cost-effectiveness, budget impact, and ethical, legal, social, and organisational issues. The purpose is to inform health policy and decision-making to promote an efficient, sustainable, equitable, and high-quality health system.

1 Policy question and context

Each HTA topic entails a policy and a research question. In healthcare, a ***policy question*** is a request to regulate a reimbursement policy and is aimed at securing financing of health technologies. Such a request, related to a particular health technology, typically addresses an existing controversy around a technology. This HTA report addresses the following policy question brought forward by the applicant:

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 between levothyroxine treatment in SCH patients 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] A valid alternative to prescribing levothyroxine is refraining from medical treatment, especially given the general higher age of the patients and additional comorbidities.

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.

2 Research question

To answer a policy question, the research question has to be defined and answered first. The ***research question*** is an answerable inquiry into the HTA topic, which requires data collection and analysis. Research questions are specific and narrow. This HTA report addresses the following research question:

What are the efficacy[1], effectiveness[2], safety[3], cost-effectiveness, and budget impact implications of levothyroxine treatment in patients 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).

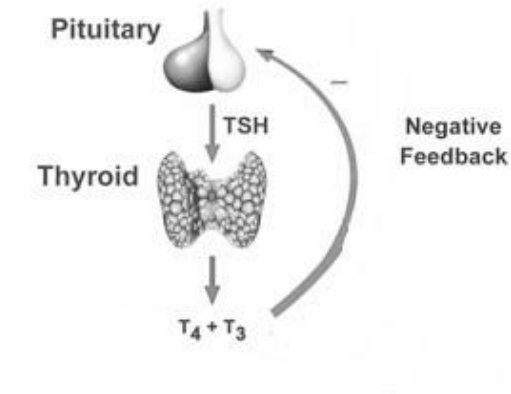
[3]Safety is a judgement of the harmful effects and their severity using the health technology. Relevant adverse events are those that result in death, are life-threatening, require inpatient hospitalisation or cause prolongation of existing hospitalisation (i.e. serious adverse events) and those that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

3 Medical background

Thyroid system

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 converse pathway; when T3/T4 concentrations are low, the TSH production is increased (absence of negative feedback loop), and when T3/T4 concentrations are high, the TSH production is decreased (negative feedback loop). The negative feedback loop is visualised in ***Figure 1***.

Figure 1. The Pituitary-Thyroid axis, including the roles thyroid stimulating hormone (TSH), thyroxine (T4) and triiodothyronine (T3). Minus indicates a negative feedback loop. Reproduced with modifications from Freitas et al., 2012. [5]



Factors influencing the thyroid system and the secretion of thyroid hormones are subject for research. The effect of aging on the pituitary-thyroid function is well investigated but results are still controversial. [6] Aging is reported as associated with a reduction in TSH secretion, but increased TSH levels in relatively healthy elders are also reported. [7,8] A large longitudinal study examined changes in thyroid function and the study authors suggested that the TSH increase arises from age-related alteration in the TSH set point or reduced TSH bioactivity rather than occult thyroid disease.[9]

When considering the effect of gender on thyroid hormone levels, TSH concentrations typically tend to be higher in women than in men. [10–12] Gender-specific alterations in TSH and free thyroid hormone levels have been reported in association with aging. Some studies showed that in men, negative relationships were observed between aging and serum thyroid hormone concentrations, while TSH

levels was not altered with increasing age. Aging did not affect the serum concentrations of free thyroid hormones in women, but TSH levels increased in an aging-dependent manner. [10,11,13]

Thyroid disorders

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. [14] Hyperthyroidism is caused by an overactive thyroid gland producing too much thyroid hormones, leading to decreased TSH levels. [15]. Primary thyroid disorders originate within the thyroid gland (e.g. caused by autoimmune thyroid disease), whereas secondary thyroid disorders refer to a change in the thyroid gland as a result of external factors such as a disease in another organ (e.g. due to a TSH-secreting pituitary adenoma) or age. [16]

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. [17] These biochemical measurements have been used to classify thyroid disorders into overt and subclinical.[18] When TSH levels are above the 95% confidence interval (CI) of a reference population (usually >4.5mlU/L),[18] and free T4-levels (fT4) are below (usually 0.7 to 1.9ng/dL),[19] this is classified as *overt* hypothyroidism. When TSH levels are above the 95% CI of a reference population (usually >4.5mlU/L), [18] and concentrations of fT4 are within the reference range (usually 0.7 to 1.9ng/dL), [19] the condition is called *subclinical* hypothyroidism (SCH). There is no strict consensus on what TSH level should be used as the cut off to distinguish subclinical from overt. Most studies define an abnormal TSH test result as the upper and lower limits of the assay's 95 percent reference range, approximately 0.1 to 4.5 mIU/L.[18]

The prevalence of overt hypothyroidism lies around 0.4% in the general population. [20,21] Symptoms of overt hypothyroidism are subtle and nonspecific and may include hoarse or deep voice, fatigue, feeling cold, weight gain, hair loss, constipation, dry skin, muscle weakness, puffy eyes, poor concentration, slow thinking, and poor memory.[18,21] If overt hypothyroidism is allowed to progress due to lack of treatment or under-treatment, then myxedema coma, a life-threatening condition, can occur. Myxedema coma is generally seen in the elderly and may be precipitated by factors that impair respiration; it is marked by hypothermia, hypoventilation, decreased level of consciousness, and sometimes seizures and death.[18]

SCH is a more commonly diagnosed condition and is found in about 4% of the general population with higher percentage (up to 20%) in women and elderly. [20,21] Most patients with SCH do not have symptoms, have symptoms that are non-specific, or have symptoms similar to those observed in overt hypothyroidism. [22] Among these symptoms, global fatigue is the symptoms that triggers most of the thyroid hormone testing in general practice. [22–24] Over time, the condition of SCH may progress to overt hypothyroidism. The rate of progression can occur in 2% to 3% of SCH patients per year. [25] However, TSH levels can also spontaneously normalise over time. [26,27] 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. [21,28,29]

Treatment of SCH

SCH can be treated with hormone replacement therapy using levothyroxine. Levothyroxine treatment is generally considered 1) to normalise TSH levels and to prevent progression to overt hypothyroidism and, and 2) to reduce symptoms of thyroid hormone deficiency (which appears to be reversible in at least 25% of patients suffering from SCH). [22,30–32]

Several studies have revealed an increased risk for cardiovascular (CV) events in people with SCH, especially in younger SCH patients. [33–35] 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 CV mortality and an increased incidence of CHD was found also in those with serum TSH level <10 µU/ml. [36]

Two guidelines recommend to initiate treatment in SCH patients with TSH levels >10 µU/ml, irrespective of age. [23,37] Guidelines recommend treating younger SCH patients with TSH levels >10 µU/ml, even in the absence of symptoms, [3,17,37] but are less consistent in their recommendations on initiating treatment in older SCH patients (>65/70 years of age) with high TSH levels: levothyroxine is only prescribed when symptoms of hypothyroidism, and/or CV risk factors are present [3,17] while others recommend to watch and wait. [17] In elderly with mild elevated TSH levels, the common recommended approach is to observe. [3,17,37] In young SCH patients with mildly raised TSH levels (4.0-10.0 Um/L) and with thyroid-related symptoms treatment is considered.[3,17,37] Irrespective of age, for SCH patients with a TSH <10 mIU/L the American Thyroid Association (ATA) recommends to consider treatment if symptoms suggestive of hypothyroidism, positive antibodies to thyroid peroxidase, or evidence of atherosclerotic cardiovascular disease, heart failure, or risk factors for these diseases.[23] The usual starting dose is 25 to 75 µg daily,

with upward dose titration based on serum TSH or persistence of symptoms, or both. [3,17,23,38] The average levothyroxine dose requirement in SCH is 0.5 µg/kg/day.[39] 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. [39] Levothyroxine should be taken orally in the morning, ideally 30 minutes before breakfast. [40](see also ***Chapter 10*** for more guideline recommendations on when to treat and how to treat SCH patients).

4 Technology

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

In Switzerland, levothyroxine is approved by the Swiss Agency for Therapeutic Products (Swissmedic) for the treatment of hypothyroidism in adults and children. Except for no treatment, 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" [41] and are reimbursed by the mandatory health insurance.

5 Population, Intervention, Comparator, Outcome (PICO)

The Population Intervention Comparator Outcome (PICO) method is used to specify the research question as defined in ***Chapter 2***. The PICO is described in more detail in ***Table 1***.

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

P:	Patients diagnosed with SCH*
I:	Treatment with levothyroxine
C:	• Placebo • No treatment
O	**Efficacy/effectiveness outcomes:** • Serum TSH levels • Hypothyroid symptoms (i.e. change in occurrence of symptoms, a validated symptom score) • Neuromuscular disorders • Musculoskeletal disorders • Cardiovascular disorders • Health-related quality of life • Withdrawal of treatment due to lack of efficacy of levothyroxine **Safety outcomes:** • Mortality • Serious and clinically important adverse events† **Economic outcomes**: • Medical costs • Cost-effectiveness • Budget impact

* We chose to use the term patient throughout the report, despite the existing controversy about whether (asymptomatic) subjects with SCH should be considered patients or not; †As defined by the included articles

6 HTA key questions

For the evaluation of the technology the following key questions covering central HTA domains, as designated by the European Network for Health Technology Assessment (EUnetHTA) Core Model (clinical efficacy, effectiveness, safety, budget impact, ethical, legal, social, and organisational aspects), are 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. What is the cost-effectiveness of levothyroxine 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?

6.1 Additional question(s)

N.A.

7 Efficacy, effectiveness, and safety

Summary statement efficacy, effectiveness, and safety

A systematic literature search retrieved 28 RCTs, based on 22 unique studies, to inform the efficacy and safety outcomes of levothyroxine treatment in patients with SCH. Levothyroxine was compared to placebo in 24 RCTs and to no treatment in four RCTs.

Articles assessing TSH levels before and after levothyroxine treatment showed almost unanimously, a statistically significant reduction in serum TSH levels and in favour of levothyroxine compared to placebo or no treatment.

However, levothyroxine treatment did not affect any of the clinically and patient-relevant efficacy outcomes that were assessed. The evidence base did not show consistent statistically significant changes between the placebo and levothyroxine treated patients in hypothyroid associated symptoms like fatigue, depression, health-related quality of life, neurological, musculoskeletal or cardiovascular outcomes. Some positive trends were reported in individual studies, but these findings were typically inconsistent between studies and not statistically significant. No indication of major safety issues of levothyroxine has been reported.

In conclusion, there is no robust evidence of clinically meaningful and patient-relevant effects of levothyroxine treatment on clinical outcomes of SCH patients. The treatment is not associated with major safety issues.

7.1 Methodology efficacy, effectiveness, and safety

A systematic review (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). [42–44]

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. Data extraction
5. Critical appraisal (risk of bias assessment)
6. Data synthesis
7. Quality control

Within these fundamental steps, a stepwise approach was implemented:

I. To conduct a systematic literature search for original RCTs and for published systematic reviews including RCTs;

II. In case no RCT or only one RCT are found in Step I. as described above, an additional systematic literature search was 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 were searched for peer-reviewed scientific literature. Grey literature sources were not considered. Since there is considerable overlap in articles included in other literature databases (such as Cochrane Library), the decision was made to search in these two main databases. The searches were 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) were applied in combination with a search string for the study design RCTs and supplemented with a SR search string to find possible missed relevant individual references. Furthermore, animal studies were excluded with an additional search string. The search was conducted in four languages: English, German, French, and

Dutch. The search was restricted to the publication period 2000 to September 13, 2021. The details of the search strategies are included in ***Appendix 14.1***. The literature database output, including all indexed fields per record (e.g. title, authors, and abstract) was exported to Endnote version 20. Duplicates in Endnote were automatically identified and manually deleted.

7.1.1.2 Selection procedure

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

1. Screening of title and abstract: this step yielded the articles that were assessed in full text. The major topics of the articles were assessed on relevancy for the objectives by the title and abstract. In this step, articles that seemed to contain relevant data for the objectives were selected for full text screening, while articles that did not seem to contain relevant data were not selected for full text assessment. In case of doubt, the article was assessed in full text.
2. Screening of full article: the articles selected during the first phase were assessed in full text. Articles were included if the reported information was relevant for the objectives and the methodological description and results section were of sufficient quality, based on the 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 articles with unexplained errors in their patient flow or articles based on duplicate data.

The process of selection and inclusion and exclusion of articles was registered in Microsoft Excel and an Endnote library. The overall exclusion criteria applied during the full text screening phase are reported in a PRISMA flow chart (***Section 7.2.1***) and an overview table with the reasons for exclusion per excluded article (***Appendix 14.2***). The implemented quality control during the selection process is described in ***Section 7.1.4***.

7.1.1.3 Inclusion and exclusion criteria

The inclusion and exclusion criteria which were applied during the selection processes for articles on levothyroxine treatment in patients diagnosed with SCH are presented in ***Table 2***.

Table 2. Inclusion and exclusion criteria for RCTs and published SRs including RCTs

	Inclusion	Exclusion
Period of publication	2000 - September 13, 2021	-
Language of Publication	• English • German • French • Dutch	All other languages
Country of study	All countries	-
Study design/type	• RCTs • Open-label extension studies (i.e. of an original RCT included in this SR)	• SR* • Narrative reviews • (Irrelevant) post-hoc/subgroup analysis of an original RCT already included in our SR • Comparative non-randomised studies • Non-comparative studies • Case reports • Abstract only (e.g. conference abstract) • Non-pertinent publication types (e.g. expert opinion, letter to editor, editorial, comment)
Study quality		• Studies with incomplete reporting • Cross-over trial without washout period
Study population	Patients diagnosed with SCH	• Overt (or clinical) hypothyroidism • Congenital hypothyroidism • Studies including only women with specific sex hormonal states, e.g. pregnant women, non-pregnant women on fertility-related treatment, or menopausal women • Studies including only patients diagnosed with SCH in combination with another specific condition (e.g. patients with SCH and diabetic nephropathy)
Study intervention	Levothyroxine treatment	Any other intervention than levothyroxine
Study comparison	• Placebo • No treatment	• Within-group comparison† • Comparison of different brands or doses of levothyroxine • Comparison with other drug-treatment • No comparison
Study outcomes	See PICO table‡	• Other outcomes • Duplicate data

Keys: PICO = Population-Intervention-Comparator-Outcome, RCT = randomised controlled trial, SCH = subclinical hypothyroidism, SR = systematic review, TSH = thyroid-stimulating hormone. * Relevant SRs were selected during the screening of title and abstract phase. During the full text screening phase, reference lists of good quality SRs were checked for possibly missed relevant individual articles. No data extraction was performed for SRs, only for relevant individual studies. † Outcomes were not extracted when within-group comparison was applied (i.e. baseline versus end of trial values per treatment arm) and no comparison between groups was performed (i.e. levothyroxine versus placebo or versus no treatment). An exception was made for the analytical outcome TSH; this outcome was also extracted when only a within-group comparison was conducted and the baseline values in both treatment arms were comparable to show the efficacy of levothyroxine. ‡ See PICO in Chapter 5.

7.1.2 *Risk of bias assessment*

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 included RCTs was assessed. [45] 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 loss 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 was rated as low risk of bias, moderate risk of bias, high risk of bias or unclear risk of bias (i.e. not reported in the article). Based on the crucial limitations for one or more of these criteria, the risk of bias within the whole study was rated in one of the three categories: low risk of bias, moderate risk of bias, or high risk of bias.

7.1.3 *Methodology data extraction, analysis and synthesis of the domains efficacy, effectiveness, and safety*

Relevant data from the included RCTs found in the peer-reviewed literature is summarised using a standardised data-extraction spreadsheet in Excel. To be selected for full analysis, a between-group statistical analysis (i.e. levothyroxine versus placebo group or levothyroxine versus no treatment group) had to be performed for the clinical outcomes of interest. An exception was made for the analytical outcome TSH to show the efficacy of levothyroxine; TSH within-group comparison data (i.e. baseline compared to end-of-trial values in levothyroxine group or in comparator group) were also considered for

full analysis when baseline values between the treatment arms were comparable. A p-value less than 0.05 is considered statistically significant.

Based on this data-extraction sheet, the data was further summarised and presented in this HTA report in study characteristics tables, risk of bias tables, summary tables, and accompanying text. Separate summary tables were made for the different outcomes of interest. The summary tables present the key features and main outcomes of the included articles. The key features consist of the risk of bias, study name, gender and age criteria at inclusion, the intervention and the comparator including the study duration, and sample size. Additional information on how to interpret the baseline values and the direction of the treatment effects is added in footnotes below the summary tables. Safety data was extracted from the RCTs that were included for the results on efficacy. Part of the neurological, musculoskeletal, and CV outcomes reported in the efficacy section were classified as safety outcomes by the authors of the included RCTs. These results will not be repeated in the safety section.

Clinically relevant strata across articles could not be applied due to the heterogeneity of the data and the wide variety of outcomes. However, if outcomes were stratified by the authors, these stratified data were reported in separate tables as presented in the original article. Pooling of efficacy outcomes was conducted when at least two studies reported on the same clinical outcome and were defined in the same way, and/or sufficient data is reported in the studies. By sufficient data we mean that the articles should address enough information to make an estimation whether pooling is justified, e.g. research questions, enough information on population, setting, follow-up, intervention.

For outcomes for which it was possible to calculate pooled estimates, the Grading of Recommendations, Assessment, Development, and Evaluations (GRADE) assessment for the level of the quality or certainty of the evidence on outcome level was implemented. [45]

7.1.4 *Quality control*

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

- Search strategy
    - An information specialist was consulted during the development of the search strategy. Quality checks were implemented, for example the search strategy was checked by a second researcher and run multiple times at separate days.
    - The proposed approach for SRs was to select relevant SRs during the screening of titles and abstracts and to check reference lists of good quality SRs for possibly missed relevant

individual articles during the full text screening phase. Twelve additional articles were included in the screening process via SRs.

- The supplementary search technique citation chasing (i.e. backward by finding other articles cited within the selected articles) was applied in addition to the database searches. No additional articles were enclosed in the selection process.

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

- Data extraction and synthesis
  - The critical appraisal of included articles was done in duplicate. In case of discrepancy a third researcher was consulted to reach consensus.
  - The data extraction spreadsheet was fully checked with the original articles by a second researcher.
  - The summary tables were fully reviewed by a second researcher.

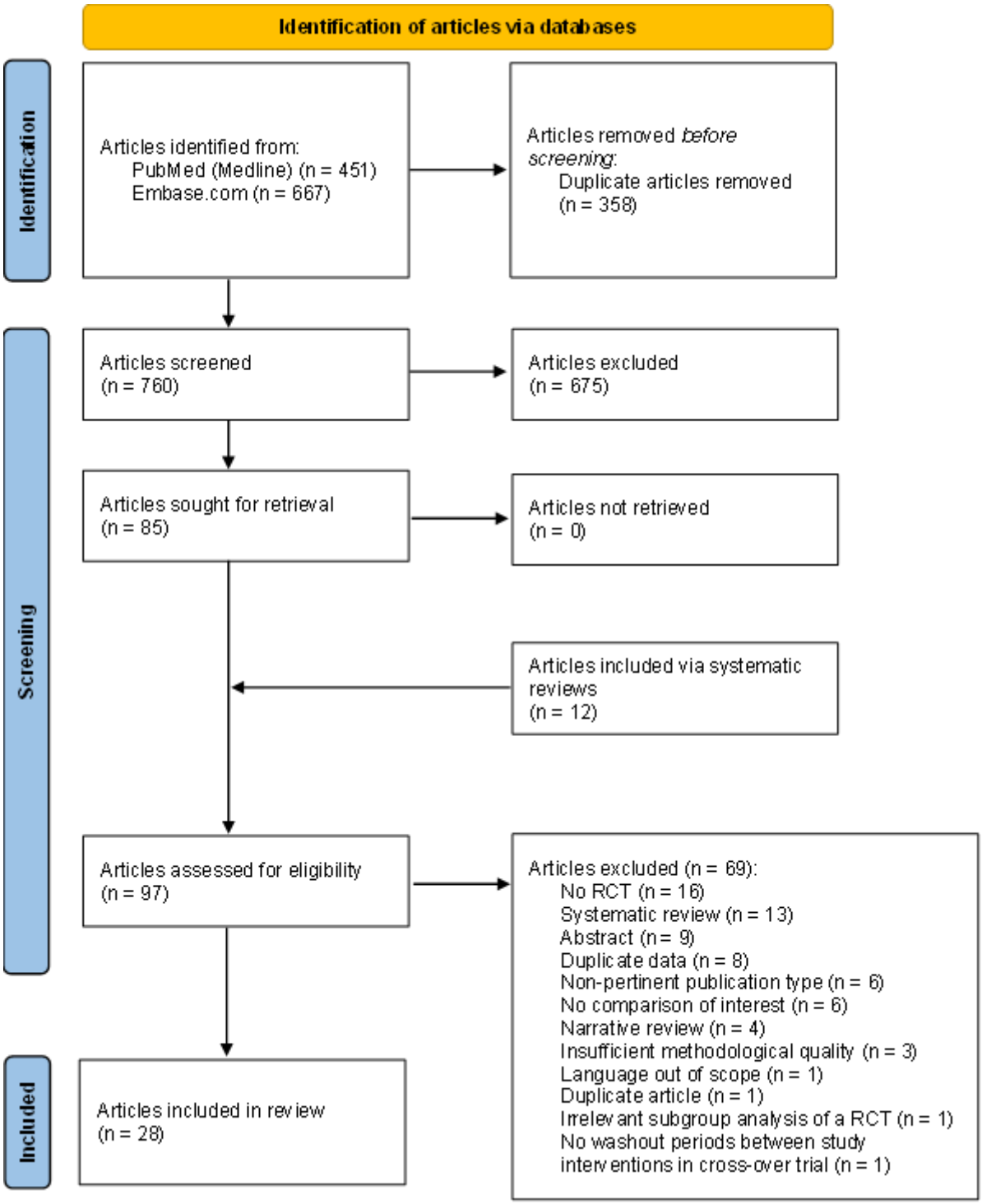
7.2 Results efficacy, effectiveness, and safety

7.2.1 *PRISMA flow diagram*

In total, 760 unique records were identified in PubMed (MEDLINE) and Embase.com with our search strategy. Of those, 675 records were excluded based on their title and abstract, resulting in 85 articles selected to be screened in full text. Twelve additional articles were included in the screening process via SRs. All 97 articles were retrieved in full text. After applying the inclusion and exclusion criteria, 28 articles based on 22 unique studies were included for full analysis. None of the additional articles included in the screening process via SRs were included in the full analysis.

The main reasons for exclusion were no RCT (n=16 articles), systematic review (n=13 articles), abstract only (n=9 articles), duplicate data (n=8 articles), non-pertinent publication type (n=6 articles), no comparison of interest (n=6 articles), and narrative review (n=4 articles). An overview of the reasons for exclusion is enclosed in the PRISMA flow chart (***Figure 2***) and a complete overview of the reasons for exclusion by each excluded article is enclosed in ***Appendix 14.2***.

Figure 2. PRISMA flowchart for the efficacy, effectiveness, and safety systematic literature search



Keys: RCT = randomised controlled trial

7.2.2 *Study characteristics and Risk of Bias assessment of included studies*

Twenty-eight RCTs were selected for full analysis in this HTA report on the efficacy and safety of levothyroxine treatment in populations diagnosed with SCH. A summary of the study characteristics is included in ***Table 3***. The studies were conducted in Brazil, China, Iran, Ireland, Italy, Japan, Kuwait, the Netherlands, Norway, Switzerland, Turkey, UK, and the USA. Ten of the 28 articles had been assigned a study name: Basel Thyroid Study [46]; Birmingham Elderly Thyroid Study [26]; Tromsø study [47,48]; Thyroid Hormone Replacement for Subclinical Hypothyroidism trial (TRUST).[4,49–53] The original TRUST study was included [4], three nested TRUST studies [49,51,53] and two studies combined data of TRUST and Institute for Evidence-Based Medicine in Old Age 80-plus thyroid trial (IEMO80+) data. [50,52] There was patient overlap for outcomes reported in RCTs based on TRUST data. The two articles based on the Tromsø study reported on different outcomes so there was no patient overlap for the outcomes of interest.

The sample size in the 28 articles ranged from 20 to 842 participants. The study populations were adults, except for one RCT, [54] that included participants from the age of 15 years onwards. The percentage included men ranged from 0% to 56.6%. Most RCTs, 24 in total, compared levothyroxine treatment with placebo. The comparator in the remaining four RCTs was no treatment. The starting dose of levothyroxine ranged from 12.5 to 100 µg and this dose was adjusted during the trails based on the TSH levels; three Iranian RCTs [55–57] did not report whether a dose adjustment was applied. The treatment duration ranged from 3 to 36 months. TSH as outcome of interest was reported in 25 articles. [4,26,46,47,49–52,54–70]

The degree of clinical hypothyroidism was assessed in one article [46] and four articles based on two unique studies reported data on hypothyroid symptoms. [4,48,50,51] Neurological outcomes were reported in seven articles based on four unique studies [4,26,48,50,53,61,71], musculoskeletal outcomes in two articles, both based on the same study [4,50] and CV outcomes in three articles, all based on the same study [4,50,52]. Three articles based on two unique studies reported the effect of levothyroxine compared to placebo for the outcome health-related quality of life (HRQoL). [4,50,71] A summary of the study characteristics is included in ***Table 3*** and the risk of bias of the individual RCTs in ***Table 4***. The overall body of evidence was substantial, and the risk of bias ranged from low to high. Six RCTs had a moderate risk of bias [4,26,50,57,60,61] (including two TRUST studies) and four RCTs had a low risk of bias, all four were (nested) TRUST studies. [49,51–53]

Table 3. Study characteristics of the included RCTs

First author, year	Country Name of study	Study population (definition SCH by TSH, fT4 level)			Sample size at randomisation	Intervention - starting dose - dose adjustment - treatment duration - TSH ref range	Comparator	Outcomes
Aghili et al., 2012 [55]	Iran NR	Persons aged 18-64 y with SCH (TSH: 4.5-10 µU/ml; fT4: 0.8-2 ng/dl)			I: n=30 C: n=30	Levothyroxine (Iran Hormone Product) - 100 µg - NR - 3 months - <4.5 µU/ml	Placebo	**TSH** **Neurological outcomes** Attention and memory tests (total memory quotient)
			Age, mean (SD) in y	*Men, %*				
		I	32.37 (11.35)	13.3				
		C	36.07 (11.35)	16.7				
Blum et al., 2018 [49]	Switzerland TRUST (nested study on carotid atherosclerosis)	Community dwelling persons aged ≥65 y with untreated SCH (TSH: 4.6-19.9 µU/ml; fT4: within reference range)			I: n=109 C: n=108	Levothyroxine - 50 µg (25 µg in certain conditions) - yes - 12 months - NR	Placebo	**TSH** **Safety** All-cause death; CV death; fatal & non-fatal CV events; ≥1 serious adverse event
			Age, mean (SD) in y	*Men, %*				
		I	74.3 (5.3)	54.2				
		C	73.9 (6.0)	51.7				
Cabral et al., 2011 [67]	Brazil NR	Women with SCH (TSH: >4.0 µU/ml; fT4: within reference range)			I: n= 14 C: n=18	LT-4 - 0.75 µg/kg/day - yes - 12 months - TSH: 0.4-4.0 µU/ml	No treatment	**TSH**
			Age, mean (SD) in y	*Men, %*				
		I	43.36 (9.8)	0				
		C	47.59 (8.4)	0				
Caraccio et al., 2002 [58]	Italy NR	Persons with SCH (TSH: >3.6 µU/ml; fT4: NR)			I: n=24 C: n=25	LT-4 (Eutirox) - 25 µg - yes - 6 months - 0.30-3.6 µU/ml	Placebo	**TSH**
			Age, mean (SD) in y	*Men, %*				
		I&C	34.5 (9.1)	14.3				
Caraccio et al., 2005 [59]	NR (authors from Italy) NR	Persons with SCH (TSH: >3.6 µU/ml; fT4: within reference range)			I: n=12 C: n=11	LT-4 (Eutirox) - 25 µg - yes - 6 months - 0.30-3.6 µU/ml	Placebo	**TSH**
			Age, mean (SEM) in y	*Men, %*				
		I	34 (2)	8.3				
		C	30 (2)	9.1				
Christ-Crain et al., 2004 [60]	Switzerland NR	Women aged 18-75 y with SCH (TSH: >5.0 µU/ml; fT4: within reference range)			I: n=31 C: n=32	LT-4 - NR - yes - 48 weeks - 0.30-4.0 µU/ml	Placebo	**TSH** By TSH <12; >12 µU/ml
			Age, mean (SD) in y	*Men, %*				
		I	57.3 (9.6)	0				
		C	57.1 (10.2)	0				

First author, year	Country Name of study	Study population (definition SCH by TSH, fT4 level)	Sample size at randomisation	Intervention - starting dose - dose adjustment - treatment duration - TSH ref range	Comparator	Outcomes
Duman et al., 2007 [68]	NR (authors from Turkey) NR	Women with newly diagnosed SCH (TSH: >4.2 µU/ml; fT4: NR) I analysed (n=20): Age, mean (SD) in y: 36 (11); Men, %: 0 C analysed (n=19): Age, mean (SD) in y: 35 (14); Men, %: 0	I: n=22 C: n=19	LT-4 (Tefor) - 25 µg - yes - 8 months - 0.27-4.2 µU/ml	No treatment	**TSH**
Iqbal et al., 2006 [47]	Norway 5th Tromsø substudy	Community-dwelling persons aged >29 y with SCH (TSH: 3.50-10.0 µU/ml; fT4: within reference range) I: Age, mean (SD) in y: 62.0 (11.9); Men, %: 50.0 C: Age, mean (SD) in y: 62.7 (12.4); Men, %: 53.1	I: n=32 C: n=32	Thyroxine - 50 µg - yes - 12 months - 0.2-4.3 µU/ml	Placebo	**TSH**
Jorde et al., 2006 [48]	Norway Tromsø	Community-dwelling persons aged >29 y with SCH (TSH: 3.5-10.0 µU/ml; fT4: within reference range) I analysed (n=36): Age, mean (SD) in y: 61.6 (11.5); Men, %: 52.8 C analysed (n=33): Age, mean (SD) in y: 63.0 (12.4); Men, %: 54.5	I: n=36 C: n=34	T4 - 50 µg (100 µg after 6 weeks) - yes - 12 months - 0.2-4.20 µU/ml	Placebo	**Hypothyroid symptoms** Change of symptoms **Neurological outcomes** Attention and memory tests (digit span forward/backward; verbal/visual recall; word list test; seashore rhythm test). Cognitive processing (CalCAP; digit symbol test; controlled word association test; vocabulary, WAIS; trail making A, B; Stroop Color-Word test parts 1-2, part 3); Depression (BDI total score); GHQ; GHQ with Likert-scoring method (general, factor A to E)
Kong et al., 2002 [61]	NR (authors from UK) NR	Community-dwelling women aged ≥18 y with SCH (TSH: 5-10 µU/ml; fT4: 0.8-16 ng/dL) I: Age, mean (SD) in y: 53 (3); Men, %: 0 C: Age, mean (SD) in y: 45 (4); Men, %: 0	I: n=23 C: n=17	Thyroxine - 50 µg - yes - 6 months - NR	Placebo	**TSH** **Neurological outcomes** Anxiety (HADS); Depression (HADS); GHQ
Mainenti et al., 2009 [69]	Brazil NR	Women aged 30-60 y with SCH (TSH: >4.0 µU/ml; fT4: 0.8-1.9 ng/dL)	I: n=11 C: n=12	L-T4 - 0.75 µg/kg day - yes	No treatment	**TSH**

First author, year	Country Name of study	Study population (definition SCH by TSH, fT4 level)			Sample size at randomisation	Intervention - starting dose - dose adjustment - treatment duration - TSH ref range	Comparator	Outcomes
			Age, mean (SD) in y	*Men, %*		- 6 months - NR		
		I	46.1 (7.2)	0				
		C	44.1 (8.4)	0				
Meek et al., 2006 [62]	USA NR	Persons ≥18 y with SCH (TSH: 5-10 µU/ml; fT4: within reference range)			I: n=12 C: n=12	Levothyroxine - 25 µg - yes - 6 months - 0.3-5.0 µU/ml	Placebo	**TSH**
			Age, median (range) in y	*Men, %*				
		I	66 (36-78	42.0				
		C	62 (52-77)	17.0				
Meier et al., 2001 [46]	Switzerland Basel Thyroid Study	Women aged 18-75 y with SCH (TSH: >5.0 µU/ml; fT4: within reference range)			I: n=33 C: n=33	L-thyroxine - NR - yes - 48 weeks - 0.10-4.0 µU/ml	Placebo	**TSH** **Diagnostic index for hypothyroidism** Billewicz score; Zulewski score **Safety** ≥1 serious adverse event
			Age, mean (SD) in y	*Men, %*				
		I	57.1 (1.8)	0				
		C	57.1 (1.9)	0				
Mikhail et al., 2008 [54]	Kuwait NR	Persons aged 15-60 y with SCH (TSH: 4.0-10 µU/ml; fT4: within reference range)			I: n=60 C: n=60	Levothyroxine - 25 µg - yes - 52 weeks - 0.3-4.0 µU/ml	Placebo	**TSH**
			Age, mean (SD) in y	*Men, %*				
		I	32.13 (10.2)	1.7				
		C	31.76 (9.89)	1.7				
Monzani et al., 2001 [63]	NR (authors from Italy) NR	Persons with SCH (TSH: >3.6 µU/ml; fT4: within reference range)			I: n=10 C: n=10	L-T4 (Eutirox) - 50 µg - yes - 6 months - 0.3-3.6 µU/ml	Placebo	**TSH**
			Age, mean (SD) in y	*Men, %*				
		I	34.3 (12.3)	10.0				
		C	29.2 (9.4)	10.0				
Monzani et al., 2004 [64]	Italy NR	Persons with SCH (TSH: >3.6 µU/ml; fT4: NR)			I: n=23 C: n=22	L-T4 - 25 µg - yes - 6 months - 0.3-3.6 µU/ml	Placebo	**TSH**
			Age, mean (SD) in y	*Men, %*				
		I&C	37 (11)	17.8				
Mooijaart et al., 2019 [50]	Ireland, the Netherlands, Switzerland, UK	Community-dwelling persons aged ≥80 y with SCH (TSH: 4.6-19.9 µU/ml; fT4: within laboratory reference range)			I: n=112 C: n=139	Levothyroxine sodium tablets - 50 µg (25 µg in certain conditions)	Placebo	**TSH** **Hypothyroid symptoms** ThyPRO hypothyroid symptoms score; ThyPRO tiredness score
			Age,	*Men,*				

First author, year	Country Name of study	Study population (definition SCH by TSH, fT4 level)			Sample size at randomisation	Intervention - starting dose - dose adjustment - treatment duration - TSH ref range	Comparator	Outcomes
	TRUST and IEMO (subgroup of those aged ≥80 y)		*mean (SD) in y*	*%*		- yes - 12-36 months - 0.4-4.6 µU/ml		**Neurological outcomes** Cognitive processing tests (letter digit coding test) **Musculoskeletal outcomes** Fracture **CV outcomes** New onset atrial fibrillation; heart failure **HRQoL** EQ-5D descriptive index; EQ VAS; Barthel index; instrumental activities of daily living **Safety** All-cause death; CV death; fatal & non-fatal CV events; ≥1 serious adverse event; ThyPRO hyperthyroid adverse symptom assessment
		I	84.0 (3.3)	53.6				
		C	85.0 (3.7)	52.3				
Nagasaki et al., 2009 [65]	Japan NR	Women with SCH (TSH: >normal upper limit; fT4: within normal levels)			I: n=48 C: n=47	L-T4 - 12.5 µg - yes - 5 months - 0.4-4.7 µU/ml	Placebo	**TSH**
			Age, mean (SEM) in y	*Men, %*				
		I	64.4 (2.59)	0				
		C	66.0 (3.0)	0				
Najafi et al., 2015 [56]	Iran NR	Persons with SCH (TSH: >4.5 µU/ml; fT4: 0.8-2 ng/dl)			I: n=30 C: n=30	Levothyroxine (Iran Hormone Product) - 100 µg - NR - 12 weeks - <4.5 µU/ml	Placebo	**TSH**
			Age, mean (SD) in y	*Men, %*				
		I	32.47 (11.35)	13.3				
		C	36.07 (11.35)	16.7				
Parle et al., 2010 [26]	UK The Birmingham elderly thyroid study	Community-dwelling persons aged ≥65 y with SCH (TSH: >5.5 µU/ml; fT4: 9-20 pmol/L)			I: n=52 C: n=42	T4 - 25 µg - yes - 12 months - 0.4-5.5 µU/ml	Placebo	**TSH** **Neurological outcomes** Cognitive function test (MEAMS, MMSE); Cognitive processing tests (SCOLP; trail making A, B, B-A); Depression (HADS).
			Age, mean (SD) in y	*Men, %*				
		I	73.5 (6.2)	40.4				
		C	74.2 (5.2)	38.1				
Reuters et al., 2012 [71]	Brazil NR	Persons with SCH (TSH: >4.0 µU/ml; fT4: 0.9-1.8 ng/dL)			I: n=35 C: n=36	L-T4 - Depending on TSH level (25, 50, or 75 µg) - yes - 6 months	Placebo	**Neurological outcomes** Anxiety (Hamilton); Depression (Hamilton; BDI total score: depressed mood) **HRQoL** SF-36 (8 domains)
			Age, mean (SD) in y	*Men, %*				

First author, year	Country Name of study	Study population (definition SCH by TSH, fT4 level)			Sample size at randomisation	Intervention - starting dose - dose adjustment - treatment duration - TSH ref range	Comparator	Outcomes
		I analysed (n=25)	51.7 (9.2)	12.0		- 0.4-4.0 µU/ml		
		C analysed (n=32)	48.3 (11.7)	0				
Rezaee et al., 2021 [57]	Iran NR	Persons with SCH (see footnote*)			I: n=15 C: n=15	Levothyroxine (Iran hormone pharmaceutical) - 50 µg - NR - 2 months - NR	Placebo	**TSH**
			Age, mean (SD) in y	*Men, %*				
		I analysed (n=14)	35.07 (9.94)	0				
		C analysed (n=14)	31.30 (4.30)	28.6				
Stott et al., 2017 [4]	Ireland, the Netherlands, Switzerland, UK TRUST (original study)	Community dwelling persons aged ≥65 y with SCH (TSH: 4.60-19.99 µU/ml; fT4: within reference range)			I: n=368 C: n=369	Levothyroxine - 50 µg (25 µg in certain conditions) - yes - 12-36 months - 0.40-4.6 µU/ml	Placebo	**TSH** By gender; by TSH <7; 7-9.99; >9.99 µU/ml **Hypothyroid symptoms** ThyPRO hypothyroid symptoms score (by gender & TSH); ThyPRO tiredness score (by gender & TSH) **Neurological outcomes** Cognitive processing tests (letter-digit coding test) **Musculoskeletal outcomes** Fracture; new diagnosis of osteoporosis **CV outcomes** New onset atrial fibrillation; heart failure **HRQoL** EQ-5D descriptive index; EQ VAS **Safety** All-cause death; CV death; fatal & non-fatal CV events; ≥1 serious adverse event; ThyPRO hyperthyroid adverse symptom assessment
			Age, mean (SD) in y	*Men, %*				
		I	74.0 (5.8)	46.2				
		C	74.8 (6.8)	46.3				
Stuber et al., 2020 [51]	Ireland, Switzerland TRUST (nested study on fatigability)	Community dwelling persons aged ≥65 y with SCH (TSH: 4.60-19.99 µU/ml; fT4: within reference range)			I: n=142 C: n=134	LT-4 - 50 µg (25 µg in certain conditions) - yes - 12 months - NR	Placebo	**TSH** **Hypothyroid symptoms** Fatigue (PFS physical score; PFS mental score)
			Age, mean (SD) in y	*Men, %*				
		I	73.9 (5.1)	56.6				

First author, year	Country Name of study	Study population (definition SCH by TSH, fT4 level)			Sample size at randomisation	Intervention - starting dose - dose adjustment - treatment duration - TSH ref range	Comparator	Outcomes
		C	73.5 (6.3)	55.2				
Wildisen et al., 2021 [53]	the Netherlands, Switzerland (Ireland, included in some analyses) TRUST (nested study on depressive symptoms)	Community-dwelling persons aged ≥65 y with SCH (TSH: 4.6-19.9 µU/ml; fT4: within reference range)			I: n=236 C: n=236	Levothyroxine sodium tablets - 50 µg (25 µg in certain conditions) - yes - 12 months - 0.4-4.6 µU/ml	Placebo	**Neurological outcomes** Depression (GDS-15 score, GDS-15 and CESD-20: incidence of mild depression, GDS-15 and CESD-20: recovery from mild depression; By gender, age and TSH level)
			Age, mean (SD) in y	*Men, %*				
		I analysed (n=211)	73.99 (NR)	44.1				
		C analysed (n=216)	75.04 (NR)	44.0				
Yazici et al., 2004 [66]	Turkey NR	Persons with SCH (TSH: >4.0 µU/ml; fT4: 0.9-1.9 ng/dL)			I: n=22 C: n=23	L-T4 - 100 µg - yes - 1 year - 0.4-4.0 µU/ml	Placebo	**TSH**
			Age, mean (SD) in y	*Men, %*				
		I	39.7 (8.7)	13.6				
		C	40.2 (9.3)	17.4				
Zhao et al., 2016 [70]	China NR	Community-dwelling persons aged ≥40 y with SCH (TSH: 4.2-10 µU/ml; fT4: within reference range)			I: n=215 C: n=163	L-thyroxine - 25 µg - yes - 15 months - 0.27-4.2 µU/ml	No treatment	**TSH**
			Age, mean (SD) in y	*Men, %*				
		I analysed (n=210)	54.98 (7.74)	27.1				
		C analysed (n=159)	55.44 (7.40)	26.4				
Zijlstra et al., 2021 [52]	Ireland, the Netherlands, Switzerland, UK TRUST and IEMO	Community-dwelling persons aged ≥65 y (TRUST) or ≥80 years (IEMO80+) with SCH (TSH: 4.6-19.9 µU/ml; fT4: within reference range)			I: n=420 C: n=422	Levothyroxine - 50 µg (25 µg in certain conditions) - yes - 12 months - NR	Placebo	**TSH** **CV outcomes** New-onset atrial fibrillation; heart failure **Safety** All-cause death; fatal & non-fatal CV events; ≥1 serious adverse event
			Age, median (IQR) in y	*Men, %*				
		I&C	75.0 (69.7-81.1)	46.8				

Keys: BDI = Beck Depression Inventory, C = comparator, CalCAP = California Computerised Assessment Package, CV = cardiovascular; EQ = EuroQol, fT4 = free thyroxine, GDS = Geriatric Depression Scale, GHQ = General Health Questionnaire, HADS = Hospital Anxiety and Depression Scale, HRQoL = Health-related quality of life, I = intervention, IEMO = Institute for Evidence-Based Medicine in Old Age, IQR = interquartile range, MMSE = Mini-Mental State Examination, MEAMS = Middlesex Elderly Assessment of Mental State, NR = not reported, PFS = Piper Fatigue Scale, RCT = randomised controlled trial, SCH = subclinical hypothyroidism, SCOLP = Speed and Capacity of Language Processing Test, SD = standard deviation, SEM = standard

error of the mean, SF-36 = 36-item short form survey, ThyPRO = thyroid-specific patient reported outcome, TRUST = Thyroid Hormone Replacement for Subclinical Hypothyroidism, TSH = thyroid-stimulating hormone, UK = United Kingdom, VAS = visual-analogue scale, WAIS = Wechsler Adult Intelligence Scale, y = years. * TSH level more than 10 µU/ml and positive Anti-thyroperoxidase antibody level; symptomatic patients (menstrual dysfunction, musculoskeletal complaints, hair and skin complaints, cardiac disease); desire for pregnancy or/and infertility and TSH level more than 15 µU/ml.

Table 4. Risk of bias of the included articles

First author, year (study name)	**Allocation concealment**	**Blinding**	**Loss to follow-up** - treatment duration - loss to FU per treatment arm (%) - reasons for loss to FU reported per treatment arm (y/n) - clinically relevant difference in loss to FU per treatment arm (y/n/?)	**Intention to treat**	**Selective outcome reporting**	**Other limitations**	**RISK OF BIAS**
Aghili et al., 2012 [55] (NR)	NR	Double blind	- 3 months - NR - NR - NR	NR	No	Funded by non-industry	High
Blum et al., 2018 [49] (nested study in TRUST in Switzerland)	Yes	Double blind	- 12 months - I: 11.9%; C: 17.6% - yes - no	Yes	No	Funded by non-industry	Low
Cabral et al., 2011[67] (NR)	NR	No	- 12 months - NR - NR - NR	No	No	Funding NR	High
Caraccio et al., 2002 [58] (NR)	NR	NR; blinding possible	- 6 months - NR - NR - NR	NR	No	Funded by non-industry	High
Caraccio et al., 2005 [59] (NR)	NR	Double blind	- 6 months - NR - NR - NR	NR	No	Funded by non-industry	High
Christ-Crain et al., 2004 [60] (NR)	NR	Double blind	- 48 weeks - I: 6.1%; C: 6.1% - yes - no	NR	No	Funded by industry & non-industry	Moderate
Duman et al., 2007 [68] (NR)	NR	No	- 8 months - I: 9.1%; C: 0% - yes - no	NR	No	Funding NR	High
Iqbal et al., 2006 [47] (Tromsø)	NR	Double blind	- 12 months - NR - NR - NR	NR	No	Funded by industry & non-industry	High
Jorde et al., 2006 [48] (Tromsø)	NR	Double blind	- 12 months - I: 0%; C: 2.9% - yes - no	NR	Yes (no p-values reported for TSH)	Funded by industry & non-industry	High

First author, year (study name)	Allocation concealment	Blinding	Loss to follow-up - treatment duration - loss to FU per treatment arm (%) - reasons for loss to FU reported per treatment arm (y/n) - clinically relevant difference in loss to FU per treatment arm (y/n/?)	Intention to treat	Selective outcome reporting	Other limitations	RISK OF BIAS
Kong et al., 2002 [61] (NR)	NR	Double blind	- 6 months - I: 13.0%; C: 11.8% - yes - no	Yes	No	Funding NR	Moderate
Mainenti et al., 2009 [69] (NR)	NR	No	- 6 months - NR - NR - NR	NR	No	Funding NR	High
Meek et al., 2006 [62] (NR)	NR	Double blind	- 6 months - NR - NR - NR	NR	No	Funded by industry	High
Meier et al., 2001 [46] (Basel Thyroid Study)	NR	Double blind	- 48 weeks - I: 6.1%; C: 3.1% - yes - no	No	No	Funded by industry & non-industry	High
Mikhail et al., 2008 [54] (NR)	NR	Double blind	- 52 weeks - NR - NR - NR	NR	No	Funding NR	High
Monzani et al., 2001[63] (NR)	NR	Double blind	- 6 months - NR - NR - NR	NR	No	Funding NR	High
Monzani et al., 2004 [64](NR)	NR	Double blind	- 6 months - NR - NR - NR	NR	No	Funding NR	High
Mooijaart et al., 2019 [50](TRUST & IEMO)	Yes	Double blind	- mean: 17 months (12-36) - I: 17.0%; C: 10.8% - yes - unclear	NR	No	Funded by industry & non-industry	Moderate
Nagasaki et al., 2009 [65] (NR)	NR	Double blind	- 5 months - NR - NR - NR	NR	No	Not funded	High
Najafi et al., 2015 [56](NR)	NR	Double blind	- 12 weeks - NR - NR - NR	NR	No	Funded by non-industry	High
Parle et al., 2010 [26] (The Birmingham elderly thyroid study)	Yes	Double blind	- 12 months - I: 13.5%; C: 35.7% - yes - unclear	Yes	No	Funded by industry & non-industry	Moderate
Reuters et al.,	NR	Double	- 6 months	No	Yes (no	Funded by	High

First author, year (study name)	Allocation concealment	Blinding	Loss to follow-up - treatment duration - loss to FU per treatment arm (%) - reasons for loss to FU reported per treatment arm (y/n) - clinically relevant difference in loss to FU per treatment arm (y/n/?)	Intention to treat	Selective outcome reporting	Other limitations	RISK OF BIAS
2012 [71] (NR)		blind	- I: 28.6%; C: 11.1% - yes - unclear		baseline & end-of-study values reported for outcomes, only variations; Billewicz scale introduced but no results reported)	industry	
Rezaee et al., 2021 [57] (NR)	Yes	Double blind	- 2 months - I: 6.7%; C: 6.7% - NR - NR	NR	No	Funded by non-industry	Moderate
Stott et al., 2017 [4] (original TRUST study)	Yes	Double blind	- 12 months - I: 9.8%; C: 8.7% - NR - NR	Yes	No	Funded by industry & non-industry	Moderate
Stuber et al., 2020 [51] (nested study in TRUST in Ireland & Switzerland)	NR (same as Stott et al., 2017)	Double blind	- 12 months - I: 16.2%; C: 17.2% - yes - no	Yes	No	Funded by industry & non-industry	Low
Wildisen et al., 2021 [53] (nested study in TRUST in the Nether-lands, Switzerland, Ireland)	NR (same as Stott et al., 2017)	Double blind	- 12 months - I: 10.6%; C: 8.5% - yes - no	Yes	No	Funded by non-industry	Low
Yazici et al., 2004 [66] (NR)	NR	Double blind	- 12 months - NR - NR - NR	NR	No	Funding NR	High
Zhao et al., 2016 [70] (NR)	NR	No	- 15 months - I: 2.3%; C: 2.5% - yes - no	NR	No	Funded by non-industry	High
Zijlstra et al., 2021 [52] (TRUST & IEMO)	NR (same as Mooijaart et al., 2019)	Double blind	- 12 months - I: 15.7%; C: 14.7% - yes - no	Yes	No	Funded by non-industry	Low

Keys: C = comparator, FU = follow-up, I = intervention, IEMO = Institute for Evidence-Based Medicine in Old Age, NR = not reported TRUST = Thyroid Hormone Replacement for Subclinical Hypothyroidism. Low risk of bias; Moderate risk of bias; High risk of bias; Unclear risk of bias.

7.2.3 *Findings efficacy*

A statement on GRADE is given in the first paragraph, followed by results of the efficacy outcomes in subparagraphs. In ***paragraph 7.2.3.2***, articles with data on TSH are discussed and when possible, the outcomes are visualised in figures. In ***paragraph 7.2.3.3***, the diagnostic index for hypothyroidism is described, followed by ***paragraph 7.2.3.4*** on hypothyroid symptoms with data on change in occurrence of individual hypothyroid symptoms and validated symptom scores. Subsequently, all articles reporting on neurological, musculoskeletal, and CV outcomes are discussed in ***paragraphs 7.2.3.5, 7.2.3.6***, and ***7.2.3.7***, respectively. At last, the combined outcome HRQoL and activities of daily living is described in ***paragraph 7.2.3.8***.

7.2.3.1 GRADE Summary of Findings

Given the high degree of heterogeneity of the outcome assessment techniques or tools, varying statistical expression parameters or different follow-up periods it was not possible to construct a GRADE summary of findings table. However, all findings were narratively summarised in the concise subparagraph below and presented in outcome tables in detail (***Appendices 14.3.1 to 14.3.7***).

7.2.3.2 TSH

In all included RCTs, thyroid function tests were used for the inclusion of the study population (i.e. diagnosis of SCH based on TSH levels and fT3/fT4). During study treatment, serum levels of TSH were assessed for dose adjustments in the levothyroxine group. The applied TSH reference values to diagnose SCH in the included RCTs ranged between 0.1 µU/ml as lowest cut-off level and 5.5 µU/ml as highest cut-off level. The TSH reference values of the majority of studies ranged between 0.3 and 4.6 µU/ml (see first column, ***Table 15 in Appendix 14.3.1***). TSH data was reported in the RCTs, however not always as a primary outcome with detailed statistical comparisons.

In total, 25 RCTs based on 21 unique studies reported TSH level changes and all but one are shown in ***Table 15 in Appendix 14.3.1***: Christ-Crain et al., 2004 [60] only performed statistical tests for TSH values stratified by TSH level and was not included in the table. Either one or more of the following statistical comparisons were presented in the 25 articles:

1) Within group comparison: baseline TSH value compared to end of trial TSH value within the treatment arm.

2) Comparison of end of trial TSH level between treatment arms: End of trial TSH value of the levothyroxine group compared to the end of trial TSH value of the placebo or no treatment group.
3) Comparison of difference in change in TSH level between treatment arms: the difference of the within-treatment arm TSH value change between treatment arms.

Fourteen RCTs, all based on unique studies compared the baseline TSH value to the end of trial TSH value within each treatment arm, of which thirteen RCTs reported a (statistically significant) reduction in TSH level within the levothyroxine group but not in the comparator group (n=10 placebo, n=3 no treatment; see ***Table 15 in Appendix 14.3.1***). [46,47,55,56,58,59,63–66] [67–69] In one RCT, the TSH level was statistically significantly decreased in the levothyroxine group and in the no treatment group. [70]

Fifteen RCT performed a statistical analysis between treatment arms of which eleven RCTs based on ten unique studies compared the TSH level at the end of the trial between the levothyroxine group and the placebo group. All but one showed a statistically significantly lower TSH level at the end of trial in the levothyroxine group compared to the placebo group (see ***Table 15 in Appendix 14.3.1***). Rezaee et al. [57] reported non-significant difference in end of trial TSH level between treatment arms (p=0.66), but it should be considered that at baseline, the TSH level differed significantly between the levothyroxine group and the placebo group (p=0.03).

Four RCTs [4,49,50,61] based on two unique studies analysed the difference in change of mean TSH level between the levothyroxine group and placebo group. The (statistically significant) differences in change of mean TSH level between treatment arms ranging from -1.64 to -3.0 µU/ml in favour of levothyroxine (see ***Table 15 in Appendix 14.3.1***).

One RCT [26] reported the percentage of participants that achieved euthyroidism. After 12 months of treatment, 84.4% of the participants in the levothyroxine group were within the euthyroid TSH range (i.e. 0.4 to 5.5 µU/ml, as applied in this study) versus 50.0% of the participants in the placebo group.

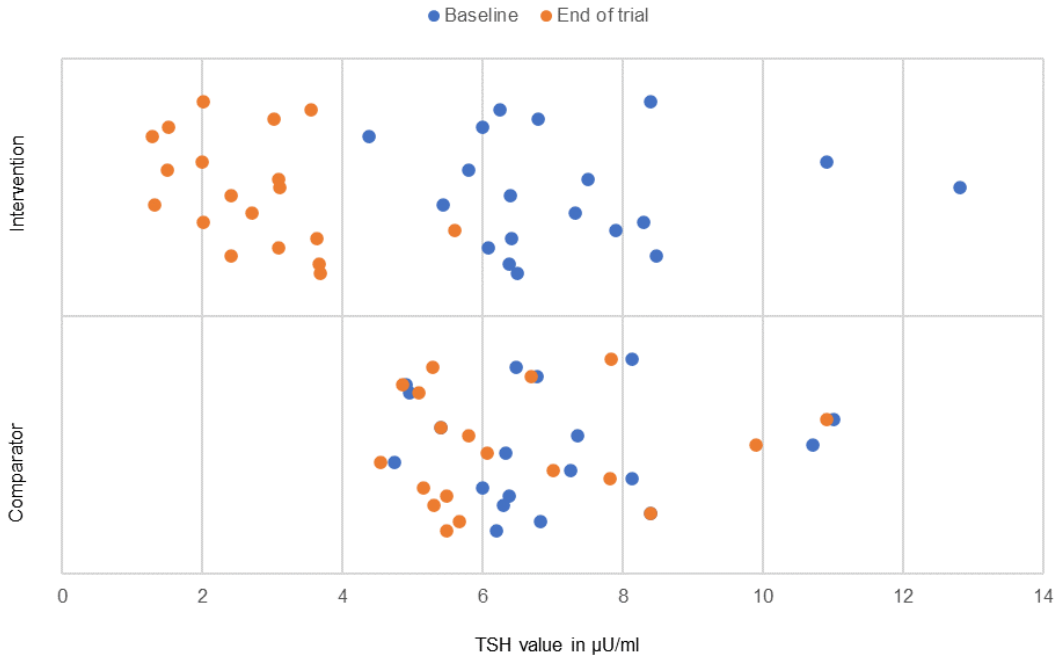
One RCT [4] reported TSH data stratified by gender (***Table 16 in Appendix 14.3.1***), and two RCTs [4,60] stratified data by TSH levels (***Table 17 in Appendix 14.3.1***). Within these stratified groups similar results were reported, with (statistically significant) reductions in mean TSH levels after levothyroxine treatment compared to placebo.

Nineteen RCTs [4,46,47,49–52,54–59,63,65–69], of which five were based on the TRUST study, reported mean TSH levels at baseline and end of trial in the intervention and comparator groups and were included in ***Figure 3*** to visualise the data. When the mean TSH levels at baseline and end of trial of the two study arms are outlined in a figure, a clear pattern is seen (***Figure 3***). For the levothyroxine group, the TSH levels show a distinct reduction from above 4.0 µU/ml at baseline to below 4.0 µU/ml at the end of trial

in 18 of the 19 RCTs. The mean TSH levels of the comparator groups at the end of trial are scattered around the mean baseline TSH levels.

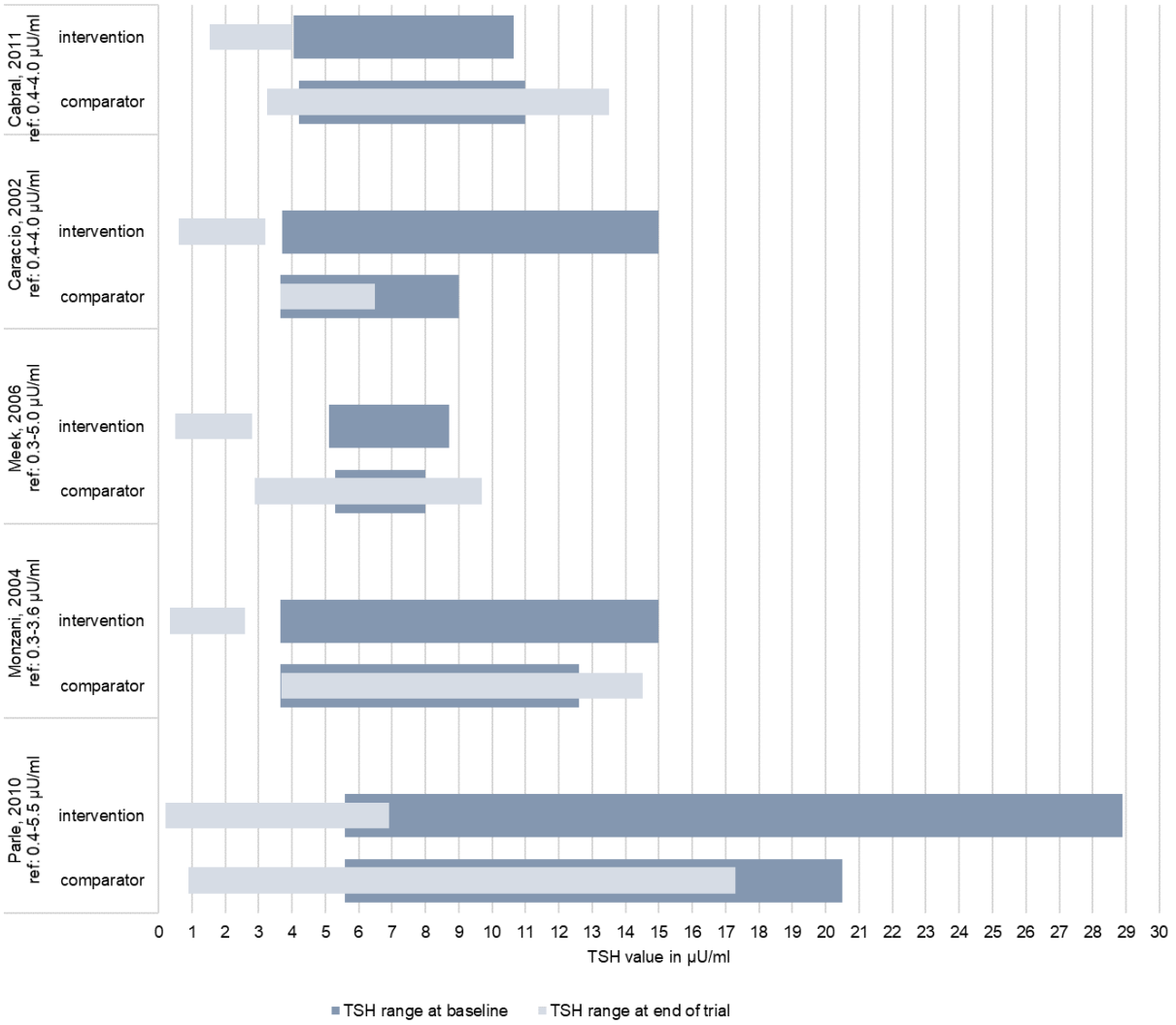
Five RCTs, all based on unique studies [26,58,62,64,67] reported the minimum and maximum TSH levels at baseline (dark blue) and at the end of trial (light blue) per treatment arm, which provides more insight in the variation within the study population and is visualised in ***Figure 4***. The ranges are wide in both the intervention and the comparator group. In all studies, the range at the end of trial is smaller than the range at baseline in the intervention group although this is not seen in all comparator groups. Within the comparator group some participants reach mean TSH levels at the end of trial which are below the TSH reference range as applied in the study.

Figure 3. Mean TSH levels at baseline and end of trial in the intervention and comparator groups of individual RCTs*



* For a complete overview of the data, see Table 15 in Appendix 14.3.1.

Figure 4. Ranges of TSH levels at baseline and end of trial in the intervention and comparator groups of individual RCTs



7.2.3.3 Diagnostic index for hypothyroidism

At baseline and end of study, one RCT [46] assessed the degree of clinical hypothyroidism with two different questionnaires: the Billewicz and Zulewski scores (***Table 18 in Appendix 14.3.2***). After 48 weeks of treatment, Meier et al., 2001 did not find a statistically significant treatment effect between the levothyroxine and placebo group for the Billewicz score or the Zulewski score. In both treatment arms, the Billewicz score increased and the Zulewski score decreased.

7.2.3.4 Hypothyroid symptoms

Two RCTs, both based on the TRUST study [4,50] measured the change in hypothyroid symptoms and tiredness scores with the thyroid-specific patient reported outcome (ThyPRO) questionnaire. Compared to placebo, there was no statistically significant association of levothyroxine with change in the hypothyroid symptoms score or tiredness score after 12 months of treatment (***Table 19 in Appendix 14.3.3***). In both articles, the ThyPRO Hypothyroid Symptoms score declined in both treatment arms and the ThyPRO Tiredness score increased in both treatment arms. Less consistency in the direction of the effect was observed in the stratified results.

Jorde et al., 2006 [48] assessed the change in occurrence of individual hypothyroid symptoms at the end of study with ten specific questions. Only on one of these questions showed a statistically significant difference between the levothyroxine and the placebo group; those in the levothyroxine group scored statistically significantly more in the positive direction (clothes looser) than the patients in the placebo group, one year after treatment start (***Table 20 in Appendix 14.3.3***). Both the levothyroxine treatment arm and the placebo arm had a change in symptoms score in the hypothyroid direction but none were statistically significant.

A nested study within the TRUST trial was conducted to gain more insight in the symptom fatigability in older persons with SCH. Stuber et al., 2020 [51] assessed the physical and mental fatigability with two independent subscores of the Pittsburgh Fatigability Scale (PFS) and did not find statistically significant between treatment arms differences between the levothyroxine and placebo group after 12 months (***Table 21 in Appendix 14.3.3***). Looking to the direction of the effect, the number of SCH patients with higher physical and mental fatigability was decreased in the levothyroxine group and the percentage was increased in the placebo group.

7.2.3.5 Neurological outcomes

Neurological outcomes were extracted from the included RCTs and categorised in six groups: cognitive function tests, attention and memory tests, cognitive processing tests, anxiety, depression, and general health questionnaire (GHQ).

Cognitive function tests

Parle et al., 2010 [26] administered the Middlesex Elderly Assessment of Mental State (MEAMS) and Mini-Mental State Examination (MMSE) to measure the cognitive functioning in elderly with SCH receiving levothyroxine or placebo and did not find statistically significant group difference between the two groups after 12 months (***Table 22 in Appendix 14.3.4***). For both MEAMs and MMSE, negligible effects were found in both the levothyroxine group and the placebo group.

Attention and memory tests

Two RCTs, based on unique studies [48,55], applied attention and memory tests and evaluated the change in scores. After three months, the mean total memory quotient score was statistically significantly more increased in the levothyroxine group compared to the placebo group (p=0.002). [55] No statistically significant difference was found for any of the other attention and memory test-scores between the levothyroxine and placebo group after 12 months in Jorde et al., 2006 [48] (***Table 23 in Appendix 14.3.4***). Negligible effects were observed for the Digit span Forward and Digit span backward in both treatment arms. An increase in verbal recall score, visual recall score and word list test score was observed in both treatment arm. The Seashore Rhythm test score decreased in the levothyroxine group after 12 months and increased in the placebo group.

Cognitive processing tests

Four RCTs based on three unique studies **[4,26,48,50]** reported the outcomes of eleven different cognitive processing tests. After 12 months of treatment, none of the articles reported a statistically significant treatment effect between the levothyroxine and placebo group for any of the tests (***Table 24 and Table 25 in Appendix 14.3.4***). Similar direction of effect was found for the levothyroxine group and the placebo group for the outcomes CalCAP (decrease in score), digit symbol test, controlled word association test and the vocabulary test (increase in score). Two articles based on the same study reported a decrease in score for the letter-digit coding test in the levothyroxine group indicating worse executive cognitive functioning while only one article reported opposite direction of effect in the placebo group. A lower score was reported for the levothyroxine group after 12 months for the Speed and Capacity of Language Processing (SCOLP) test and a slightly higher score in the placebo group at the end of the trial.

Anxiety

Anxiety was investigated in two articles based on two different studies [61,71] by the Hospital Anxiety and Depression Scale (HADS) and the Hamilton scale. Compared to placebo, there was no statistically significant association of levothyroxine with anxiety after 6 months of treatment in the articles (***Table 26 in Appendix 14.3.4***). In both treatment arms, most SCH patients did not change in case status, 75% and 72% in the levothyroxine group and placebo group, respectively. In Reuters et al., 2012 both treatment arms reported a negative variation in the Hamilton scale at the end of trial.

Depression

Five unique studies [26,48,53,61,71] assessed the effect of levothyroxine provided for six to twelve months on depression. None of the studies found a statistically significant difference in depression (in scores or case status) between the levothyroxine and placebo group. No treatment effect of levothyroxine was found on the incidence and recovery of mild depression [53] (***Table 27 in Appendix 14.3.4***) .One RCT [53] reported depression score stratified by gender, age, and TSH levels (***Table 28 in Appendix 14.3.4***). Within these stratified groups, no statistically significant between group differences were found between the levothyroxine and placebo group after 12 months. An increase in score was observed in both treatment arms in HADS and GDS-15 score, and a decrease in score in both treatment arms in BDI total score and at the Hamilton scale. The case status was in 86% of the patients unchanged in the placebo group compared to 80% in the levothyroxine group after 6 months of treatment. The incidence of mild depression based on scores of GDS-15 and Center for Epidemiologic Studies Depression scale (CESD) was slightly higher in the levothyroxine group (5.1%) compared to the placebo group (4.9%) although not statistically significant. The recovery of mild depression based on scores of GDS-15 and CESD was higher in the placebo group (60.0%) compared to the levothyroxine group (39.1%). The GDS-15 scores stratified by age and gender was above >0 and therefore in benefit of placebo. The direction of effect varied among the GDS-5 scores stratified by TSH: for SCH patients with TSH level 4.5 to <7.0 µU/ml, the GDS-15 score above >0 and therefore in benefit of placebo while for the categories TSH 7.0 to <10 µU/ml and TSH 10.0 to 10.0 µU/ml, the GDS-15 score above <0 and therefore in benefit of levothyroxine.

GHQ

Two unique studies reported in two articles used the GHQ to assess the effect of levothyroxine on mental health status, using the GHQ-scoring method (***Table 29 in Appendix 14.3.4***) and/or the Likert scoring method (***Table 30 in Appendix 14.3.4***). Compared to placebo, there was no statistically significant difference with levothyroxine regarding change in GHQ scores (irrespective of measure method) after 6 months [61] or 12 months. [48]. The mean GHQ-30 score increased in both the levothyroxine and placebo group. The case status as indicator for general psychological disorder did not change for most SCH patients in the levothyroxine group (60%), while in the placebo group most SCH patients improved from

case to non-case (58%). The GSH with Likert-scoring method for the total score and all factors was increased in both the levothyroxine group and placebo group.

7.2.3.6 Musculoskeletal outcomes

Two RCTs, both based on the TRUST study [4,50], recorded the number of fractures and number of new diagnoses of osteoporosis after at least twelve months in the levothyroxine and placebo group. Both outcomes were similar in the two groups at the end of the trial (***Table 31 in Appendix 14.3.5***).

7.2.3.7 Cardiovascular outcomes

Three RCTs, all based on the TRUST study [4,50,52], reported the number of new onset atrial fibrillation and heart failure in patients receiving levothyroxine and placebo. The occurrence of cardiovascular events was low in both treatment groups: it ranged from 3 cases with heart failure (0.8%) in the levothyroxine group to six cases (4.3%) for both new onset atrial fibrillation and heart failure in the placebo group. There was no statistically significant difference in cardiovascular outcome between the two treatment arms after at least 12 months (***Table 32 in Appendix 14.3.6***). The estimated risk difference for new onset atrial fibrillation was around zero, but the adjusted hazard ratio (HR) was below one in the two articles based on the TRUST study. The same direction of effect was found for heart failure (i.e. estimated risk difference around zero, adjusted HR below 1).

7.2.3.8 Activities of daily living and HRQoL

Mooijaart et al., 2019 [50] assessed the change in activities of daily living by two different instruments, the Barthel index and the Instrument activities of daily living, with a higher score favourable in both instruments (***Table 33 in Appendix 14.3.7***). For none of the outcomes, there was statistically significant difference in score between the levothyroxine group and the placebo group. The scores of both the Barthel index and the instrumental activities of daily living decreased in both treatment arms.

Three RCTs based on two unique studies [4,50,71] assessed the change in HRQoL by two different instruments (***Table 34 and Table 35 in Appendix 14.3.7***); higher score is favourable in both instruments. In one article, the levothyroxine group scored statistically significantly higher (indicating better satisfaction with QoL) compared to the placebo group on two outcomes of the 36-item short form survey (SF-36) (i.e. role physical and role pain) after 6 months of treatment. The study authors stated that these results suggest that levothyroxine tends to improve some physical aspects of HRQoL in patients with SCH but did not provide an explanation for this finding.

The EuroQol-5 Dimension (EQ-5D) descriptive index score was slightly decreased in the levothyroxine group in two articles based on the same study, but slightly increased in the placebo group in only one article (borderline statistically significant). The EQ-5D visual-analogue scale score was slightly decreased in the levothyroxine group and slightly increased in the placebo group in two articles based on the TRUST study. The mean variation of the SF-36 score was positive in both treatment arms for five SF-36 elements (general health, physical function, social function, role emotional and mental health). The SF-36 vitality score decreased in the levothyroxine group and increased in the placebo group. The direction of effect of SF-36 Role physical and SF-36 Role pain are discussed above.

7.2.3.9 Withdrawal of treatment due to lack of efficacy of levothyroxine

No articles were included that reported on withdrawal of treatment due to lack of efficacy of levothyroxine.

7.2.4 *Findings effectiveness*

Since the systematic literature search for RCTs included 28 RCTs, it was decided not to proceed with the systematic literature search for comparative non-randomised studies.

7.2.5 *Findings safety*

Part of the neurological, musculoskeletal, and CV outcomes reported in the efficacy section were classified as safety outcomes by the authors of the included RCTs. These results will not be repeated in this safety section.

Five of the twenty-eight included RCTs reported safety data on all-cause death, CV death, fatal or non-fatal CV events, at least one serious adverse event, and the adverse symptom assessment with the ThyPRO Hyperthyroid questionnaire.

Four RCTs with data on death and fatal or non-fatal CV events, all based on the TRUST study, did not find statistically significant differences for the number of deaths in the levothyroxine and placebo groups (***Table 5***). In the articles, the percentage of all-cause mortality ranged from 2.7% to 4.5% in participants treated with levothyroxine and from 0.9% to 2.9% in participants receiving a placebo. The percentage of fatal or non-fatal CV events occurring in the study arms was comparable, ranging from 4.5% to 6.4% in the levothyroxine group and from 5.4% to 10.1% in the placebo group.

Table 5. Safety results on levothyroxine treatment: Death and fatal or non-fatal CV events

First author, year (study name; gender; age) Risk of bias RCT	Intervention (duration) Comparator (duration)	Sample size	All-cause death		CV death		Fatal or non-fatal CV events	
			n (%)	Adjusted HR (95% CI); p-value	n (%)	Adjusted HR (95% CI); p-value	n (%)	Adjusted HR (95% CI); p-value
Blum et al., 2018 [49] (nested study in TRUST in Switzerland; mix; ≥65y) Low risk of bias	Levothyroxine (12 months)	109	4 (3.7%)	NR	2 (1.8%)	NR	7 (6.4%)	NR
	Placebo (12 months)	108	1 (0.9%)		0 (0%)		6 (5.6%)	
Mooijaart et al., 2019 [50] (TRUST and IEMO; mix; ≥80y) Moderate risk of bias	Levothyroxine (12 months)	112	5 (4.5%)	Unadjusted HR: 1.39 (0.37 to 5.19); NR	0 (0%)	NR	7 (6.3%)	Unadjusted HR: 0.61 (0.24 to 1.50); NR
	Placebo (12 months)	139	4 (2.9%)		1 (0.7%)		14 (10.1%)	
Stott et al., 2017 [4] (original TRUST study; mix; ≥65y) Moderate risk of bias	Levothyroxine (12 months)	368	10 (2.7%)	1.91 (0.65 to 5.60); 0.238	2 (0.5%)	NR	18 (4.9%)	0.89 (0.47 to 1.69); 0.728
	Placebo (12 months)	369	5 (1.4%)		1 (0.3%)		20 (5.4%)	
Zijlstra et al., 2021 [52] (TRUST & IEMO; mix; ≥65y or ≥80y) Low risk of bias	Levothyroxine (12 months)	420	12 (2.9%)	1.28 (0.54 to 3.03); NR	-	-	19 (4.5%)	0.74 (0.41 to 1.35); NR
	Placebo (12 months)	422	9 (2.1%)		-		25 (5.9%)	

Keys: CI = confidence interval, CV = cardiovascular, HR = hazard ratio, IEMO = Institute for Evidence-Based Medicine in Old Age, n = number, NR = not reported, RCT = randomised controlled trial, TRUST = Thyroid Hormone Replacement for Sub-clinical Hypothyroidism, y = years.

Five RCTs, of which four were based on the TRUST study, reported the number of participants with at least one serious adverse event. Two RCTs [49,50] provided a definition of a serious adverse event (see footnote below ***Table***) and one RCT [50] specified the most common serious adverse events: stroke (n=3 participants), anaemia (n=2), and pneumonia (n=2) in the levothyroxine group, and pneumonia (n=4 participants), cardiac failure (n=2), and respiratory failure (n=2) in the placebo group. Two articles reporting on data of the TRUST study found a statistically significant difference for the number of patients with at least one serious adverse event. No details were reported on the type of serious adverse events in these articles. In both articles, the number was slightly higher in the placebo group than in the levothyroxine group. In Stott et al., the authors stated that the adverse events were spread among a range of body systems and no particular pattern of event type was observed that contributed to this difference and report this as a chance finding. The data on serious adverse events of Zijlstra et al. was partly based on the same study (i.e. combination of TRUST and IEMO) and no statements were made

by the study authors on this outcome. In addition, two articles based on the TRUST study [4,50] reported data of the adverse symptom assessment with the ThyPRO Hyperthyroid questionnaire. No statistically significant differences were found for levothyroxine compared with placebo.

Table 6. Safety results on levothyroxine treatment: Serious adverse events

First author, year (study name; gender; age) Risk of bias RCT	Intervention (duration) Comparator (duration)	Sample size	≥1 serious adverse event* n (%)	Estimated risk difference (95% CI)‡; p-value	Adjusted HR (95% CI) ;p-value	ThyPRO Hyperthyroid; adverse symptom assessment† Mean score (SD)	Adjusted HR (95% CI); p-value
Blum et al., 2018 [49] (nested study in TRUST in Switzerland; mix; ≥65y) Low risk of bias	Levothyroxine (12 months)	109	30 (27.5%)	NR	NR; 0.43	-	-
	Placebo (12 months)	108	35 (32.4%)			-	
Meier et al., 2001 [46] (Basel Thyroid Study; women; 18-75y) High risk of bias	Levothyroxine (48 weeks)	31	0 (0%)	NR	NR	-	-
	Placebo (48 weeks)	32	0 (0%)			-	
Mooijaart et al., 2019 [50] (TRUST and IEMO; mix; ≥80y) Moderate risk of bias	Levothyroxine (12 months)	112	33 (29.5%)	-0.01 (-0.04 to 0.01); NR	NR	10.9 (11.3)	-0.50 (-2.62 to 1.63); NR
	Placebo (12 months)	139	40 (28.8%)			9.1 (10.8)	
Stott et al., 2017 [4] (original TRUST study; mix; ≥65y) Moderate risk of bias	Levothyroxine (12 months)	368	78 (21.2%)	NR	0.94 (0.88 to 1.00); 0.049	10.5 (10.8)	0.6 (-0.7 to 1.9); 0.35
	Placebo (12 months)	369	103 (27.9%)			10.3 (11.3)	
Zijlstra et al., 2021 [52] (TRUST & IEMO; mix; ≥65y or ≥80y) Low risk of bias	Levothyroxine (12 months)	420	90 (21.4%)	NR	0.73 (0.55 to 0.96); NR	-	-
	Placebo (12 months)	422	116 (27.5%)			-	

Keys: CI = confidence interval, HR = hazard ratio, IEMO = Institute for Evidence-Based Medicine in Old Age, n = number, NR = not reported, RCT = randomised controlled trial, SD = standard deviation, ThyPRO = thyroid-specific patient reported outcome, TRUST = Thyroid Hormone Replacement for Subclinical Hypothyroidism, y = years. * Serious adverse events were defined as: 1) Any untoward medical occurrence that resulted in death, was life-threatening, requires inpatient hospitalisation or prolongation of existing hospitalisation, or resulted in persistent or significant disability/incapacity [49] ; 2) All undesired medical events involving a participant, which are not necessarily associated with the treatment, that are fatal, threaten the life of the participant, make hospital admission or an extension of the admission necessary, cause persistent or significant invalidity or work disability, manifest themselves in a congenital abnormality or malformation, or could, according to the researchers, have developed to a serious undesired medical event but were prevented because of premature interference [50]; or 3) Not defined [4,46,52]. † Hyperthyroid symptom score from the

ThyPRO questionnaire to assess for overtreatment or the development of hyperthyroidism; 0-100 scale, higher scores indicating more symptoms; minimum clinically important difference estimated as 9 points; ‡ Estimated risk difference were estimated using Cox proportional hazard regression models, presented as risk differences and 95%CIs, obtained through bootstrap resampling in 1000 iterations. Statistically significant results

8 Costs, cost-effectiveness and budget impact

Summary statement costs, cost-effectiveness and budget impact

---

From the systematic literature search no published cost-effectiveness analyses were identified. Costs of the treatment with levothyroxine were calculated as a function of drug unit prices and a standard dosing schedule. In the budget impact analysis, we calculated the total healthcare cost of levothyroxine for adult SCH patients between ages 18 and 99 years, in Switzerland over a time horizon of five years (years 2020-2024). In the absence of reliable epidemiological data on the number of SCH patients in Switzerland, nor on the percentage of SCH patients receiving levothyroxine, we estimated the number of SCH patients applying European prevalence rates from the literature to population estimates from the Swiss FSO, and we used a percentage range (10% - 90%) of the estimated prevalent population that would be eligible for levothyroxine in Switzerland.

The total costs of treatment ranged between 9 Mio CHF and 87 Mio CHF over a time horizon of five years.

8.1 Methodology costs, cost-effectiveness and budget impact

8.1.1 *Databases and search strategy*

8.1.1.1 Search strategy

The cost-effectiveness systematic literature search was conducted according to the principles of the systematic literature search outlined in ***Chapter 7.1***. PubMed (MEDLINE) and Embase.com databases were searched for peer-reviewed scientific literature. In addition, economic databases, using the Cost Effectiveness Analysis (CEA) Registry and NHS EED, were searched. The searches were 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 search terms for cost-effectiveness were developed together with a medical information specialist. The details of the search strategy are presented in ***Appendix 14.4***.

8.1.1.2 Selection procedure

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

1. Screening of title and abstract:

The major topics of the articles were 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 were selected for full-text screening, while articles that did not seem to contain relevant data were not selected for full-text assessment.

2. Screening of full-text article:

The full-text articles were assessed based on the inclusion and exclusion criteria as defined in this HTA report (***Table 6***). Articles were included if they fulfil the eligibility criteria. The process of selection and inclusion and exclusion of articles were recorded in both Microsoft Excel and Rayyan (www.rayyan.ai). This method provides transparency regarding all selection steps and assures reproducibility. The selection procedure applied during the full-text screening phase was presented in a PRISMA flow chart (see ***Figure 5***).

8.1.2 *Other sources*

N.A.

8.1.3 *Inclusion and exclusion criteria*

The inclusion and exclusion criteria applied during the selection processes are listed in ***Table 6***.

Table 6. Inclusion and exclusion criteria for cost-effectiveness analyses studies

	Inclusion	Exclusion
Language of Publication	No restriction	-
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 • Budget impact analysis • Costing studies	Resource use measurement
Study population	Patients diagnosed with SCH	• Overt (or clinical) hypothyroidism • 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	• Medical costs • Cost-effectiveness • Budget impact See PICO table‡	Other outcomes

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

8.1.4 *Assessment of quality of evidence*

- 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 were assessed for relevancy in duplicate by two independent researchers. If there were differences between the findings of the two researchers, these differences were identified and discussed. In case of discrepancy or disagreements a third researcher were consulted.

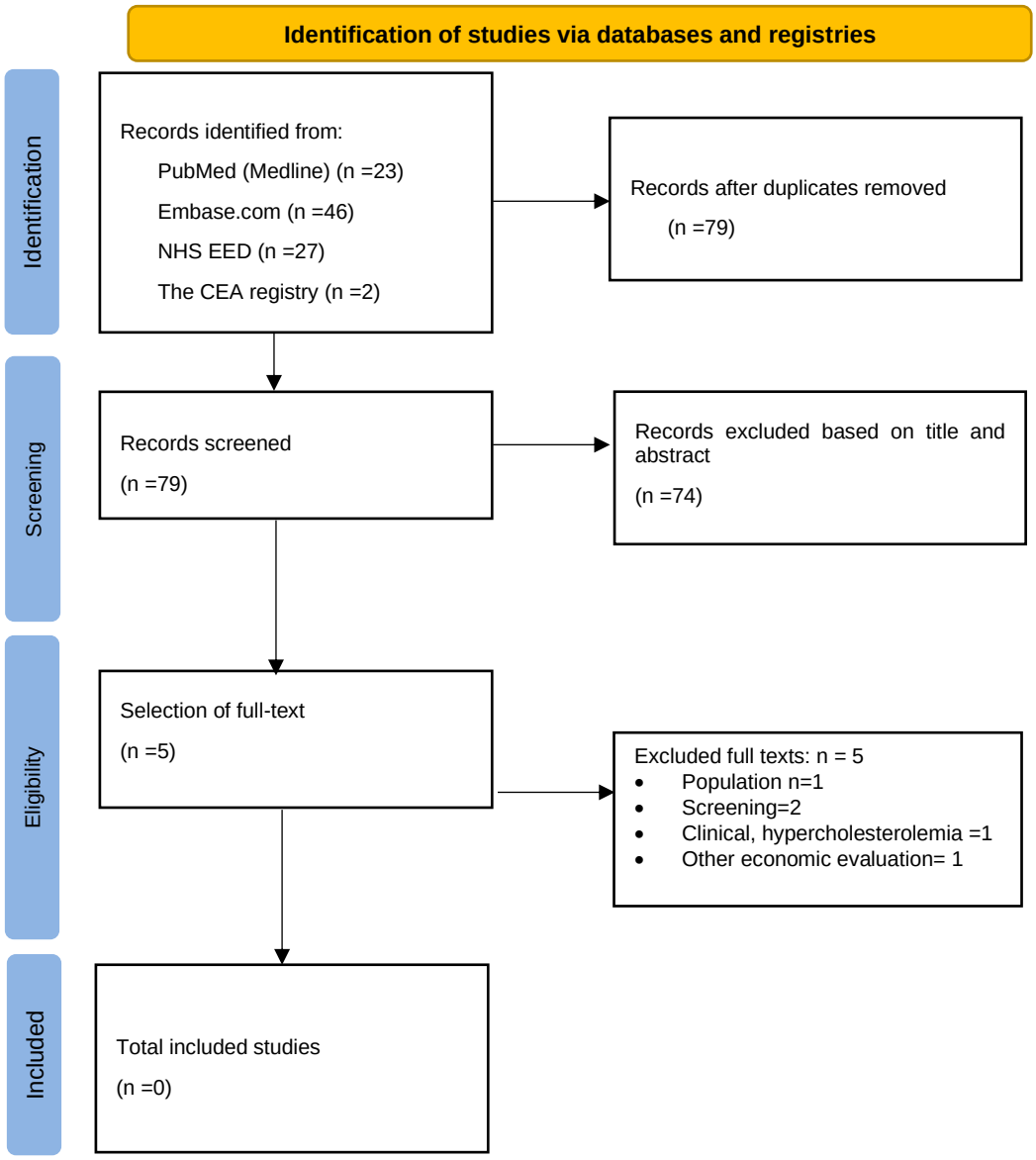
8.1.5 *Methodology data extraction, analysis and synthesis of health economic data*

N.A. Data extraction, analysis and synthesis of health economic data were not applicable for this HTA report.

8.2 Results costs, cost-effectiveness and budget impact

8.2.1 *PRISMA flow diagram*

Figure 5. PRISMA flowchart of the cost-effectiveness systematic literature search



Abbreviations: CEA: Cost-Effectiveness Analysis; NHS EED: National Health Service Economic Evaluation Database

8.2.2 *Study characteristics and quality assessment of included studies*

N.A. No published cost-effectiveness study was identified for levothyroxine treatment in patients diagnosed with SCH.

8.2.3 *Evidence table*

N.A.

8.2.4 *Findings costs*

8.2.4.1 Drug acquisition costs

Drug acquisition costs for levothyroxine were calculated as a function of drug unit prices, and a standard dosing schedule. Drug unit prices were sourced from Spezialitätenliste (www.spezialitätenliste.ch), published by the FOPH on June 13, 2022. To calculate drug acquisition costs, we used the public price (i.e. Publikumspreis). The dosage of Novothyral tablets were not provided in the Spezialitätenliste. The information on the dosage was sourced from the swissmedicinfo.ch website. All considered preparations for Levothyroxine, and their prices, are summarised in ***Table 39*** of ***Appendix 14.4***Table 39.

For this analysis, we assumed an average defined daily dose (DDD) of 50µg for the treatment of SCH. This assumption was based on the average dosages included in the clinical trials [4,48,50] and individual studies [57,61,63,71] included in the clinical systematic review (see ***section 7.2.2, Table 3***). Based on this assumed DDD of 50µg, we calculated the average price per DDD, and the average yearly costs of taking 50µg daily (mean price per DDD: 0.19 CHF, price per year: 69.51 CHF).

8.2.5 *Findings cost-effectiveness*

N.A.

8.2.6 *Findings budget impact*

8.2.6.1 Patient population

The exact numbers of individuals with SCH in Switzerland, and those eligible for therapy with levothyroxine are not known. The prevalent population of SCH was estimated based on population data from the Swiss Federal Statistical Office (FSO), [72] and prevalence rates reported in the literature. [20] From this population, we estimated the number of eligible patients for levothyroxine. These steps are detailed in ***subsections 8.2.6.2 - 8.2.6.4***.

8.2.6.2 Swiss population estimates

To estimate the treatment costs for the total Swiss population, we used the "Population trend" dataset from the Swiss FSO, which presents data scenarios for the time period 2020-2050. [72]The dataset was accessed through the R package BFS: Search and Download Data from the Swiss FSO (version 0.4.1).

In this projection, the FSO described the development in Switzerland's resident population over the next few decades. The FSO notes that the scenarios do not represent forecasts and that the accuracy of these scenarios depend on whether the underlying hypothesis are actually correct. Furthermore, the FSO states that the currently published scenarios were calculated before the COVID-19 pandemic. As it remains unclear to what extend this pandemic will influence the development of the Swiss population, it is possible that the FSO could adjust these scenarios in the months to come. This dataset holds information on the estimated number of individuals for the years 2020 through 2050, and categorises age data in five-year intervals, until the age of 120 years. For these analyses, we linearly interpolated the number of individuals in each age group (i.e. assuming an equal number of individuals for each five years within one age group) to allow more flexibility in re-categorising age, and applying age-specific prevalence rates. The dataset used for the "Population trend" analysis is available in the Italian, French, and German language. For this analysis, the German dataset was used, and data were translated into English.

We selected the Swiss adult population between ages 18 and 99 years. ***Table 7*** summarises the estimated number of individuals by year from 2020 to 2024.

Table 7. Number of individuals between ages 18 and 99 years in Switzerland in the FSO Population trend dataset

Year					Total
2020	**2021**	**2022**	**2023**	**2024**	
7'227'894	7'285'575	7'344'499	7'404'359	7'465'323	36'727'650

8.2.6.3 Prevalence of patients with SCH

In 2014, Garmendia Madariaga et al. conducted a meta-analysis, and estimated the total prevalence of thyroid dysfunction for SCH in Europe. [20] Prevalence rates were not specifically reported for age and gender, but pooled across ten included studies. The total prevalence of thyroid dysfunction for SCH from the pooled meta-analysis was 3.8% (95% credibility interval 3.48% - 4.15%). The prevalence rates from this meta-analysis were applied to the Swiss population estimate detailed in ***section 8.2.6.2***. The estimated number of Swiss adults with SCH by year from 2020 to 2024 is presented in ***Table 8***.

Table 8. Estimated number of individuals with SCH between ages 18 and 99 years in Switzerland

Year					Total
2020	**2021**	**2022**	**2023**	**2024**	
274’660	276’852	279’091	281’366	283’682	1’395’651

8.2.6.4 Patients treated for SCH in Switzerland

There are no reliable figures available on the percentage of patients that are treated with levothyroxine. Therefore, we used a percentage range (10% - 90%) of the prevalent population calculated in ***section 8.2.6.3***, that would receive levothyroxine. The estimated number of individuals with SCH treated with levothyroxine, assuming the 10% - 90% range is summarised by year from 2020 to 2024 in ***Table 9***.

Table 9. Individuals with SCH treated with levothyroxine, assuming the 10% - 90% range in Switzerland

Year					Total	Range (%)
2020	**2021**	**2022**	**2023**	**2024**		
27’466	27’685	27’909	28’137	28’368	139’565	**10%**
54’932	55’370	55’818	56’273	56’736	279’130	**20%**
82’398	83’056	83’727	84’410	85’105	418’695	**30%**
109’864	110’741	111’637	112’546	113’473	558,260	**40%**
137’330	138’426	139’545	140’683	141’841	697’825	**50%**
164’796	166’111	167’455	168’819	170’209	837’390	**60%**
192’262	193’796	195’364	196’956	198’578	976’956	**70%**
219’728	221’481	223’273	225’093	226’946	1’116’521	**80%**
247’194	249’167	251’182	253’230	255’314	1’256’086	**90%**

8.2.6.5 Budget-impact results

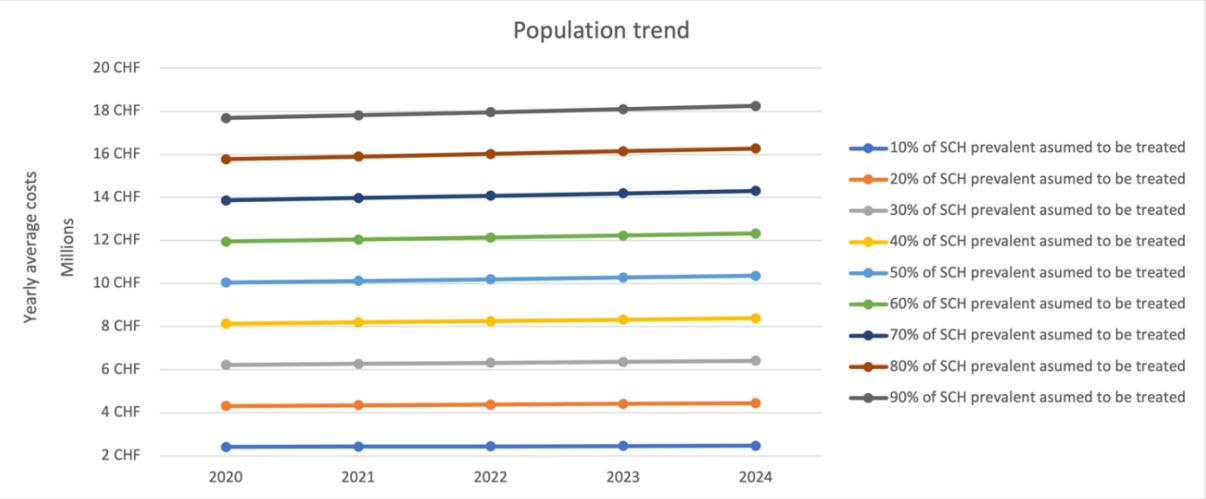
After multiplying the population eligible for levothyroxine with the yearly cost of levothyroxine treatment (based on an assumed DDD of 50µg) and applying a 10% - 90% range of treated patients, the total costs of treatment ranged between 9’701’140 CHF and 87’310’263 CHF over a time horizon of five years.

The total costs of levothyroxine per year in CHF are summarised in ***Table 10*** and depicted in ***Figure 6***.

Table 10. Total costs of levothyroxine per year in CHF, assuming the 10% - 90% range of treated SCH individuals in Switzerland

Year					Total	Range (%)
2020	**2021**	**2022**	**2023**	**2024**		
1’909’156	1’924’392	1’939’956	1’955’767	1’971’870	9’701’140	**10%**
3’818’312	3’848’783	3’879’911	3’911’534	3’943’740	19’402’281	**20%**
5’727’468	5’773’175	5’819’867	5’867’301	5’915’610	29’103’421	**30%**
7’636’624	7’697’567	7’759’823	7’823’068	7’887’479	38’804’561	**40%**
9’545’780	9’621’959	9’699’779	9’778’835	9’859’349	48’505’702	**50%**
11’454’936	11’546’350	11’639’734	11’734’602	11’831’219	58’206’842	**60%**
13’364’092	13’470’742	13’579’690	13’690’369	13’803’089	67’907’982	**70%**
15’273’248	15’395’134	15’519’646	15’646’136	15’774’959	77’609’123	**80%**
17’182’404	17’319’525	17’459’602	17,601’903	17’746’829	87’310’263	**90%**

Figure 6. Total costs of levothyroxine per year in CHF, assuming the 10% - 90% range of treated SCH individuals in Switzerland



The budget impact results only cover acquisition costs associated with the treatment of levothyroxine for patients with SCH. The real-world budget impact might be higher.

The total estimated revenue for levothyroxine between January and December 2020, based on the COGE ASL database, was 17’023’076 CHF (data provided by FOPH). This estimate includes revenues for all listed levothyroxine products, regardless of their indication. Hence, this estimate extends beyond the use of levothyroxine for SCH only. Consequently, any estimated average yearly cost in this analysis that are close, equal to, or above the 17 million CHF revenue need to be regarded as an overestimation of the calculated budget impact. Since SCH constitutes only a proportion of all indications for which levothyroxine is prescribed, the average yearly costs for SCH are assumed to be lower than the total estimated revenue in 2020. However, it is yet unknown, what are the proportions of patients treated for SCH to all other indications for which the COGE data were available.

9 Ethical, legal, social and organisational issues

Summary statement ethical, legal, social and organisational issues

---

The literature search retrieved three studies addressing ethical, legal, social and organisational issues directly or indirectly related to levothyroxine treatment for patients with SCH. Identified issues included: the general physician dilemma when confronted with patients that either lack the ability to provide patient consent or participate in decision-making regarding treatment, or when patients disagree with clinical diagnostic results; a substantial risk of misdiagnosis, given the fact that thyroid disorders tend to develop slowly. The chapter further addresses the importance of bioequivalence of replacement or generic drugs and international differences in prescription strategies for SCH patients.

9.1 Methodology ethical, legal, social and organisational issues

A search strategy was developed to address potential issues pertaining to the use of levothyroxine in patients with SCH. This search strategy focused on potential ethical, legal, social and/or organisational issues, and is described in further detail below.

9.1.1 *Databases and search strategy*

While conducting the systematic literature searches for efficacy, effectiveness, safety and cost-effectiveness, articles discussing issues pertaining to other HTA domains were included from the literature. Beyond this initial selection, there were no specific in- or exclusion criteria used.

9.1.2 *Other sources*

Grey literature searches were also performed on the ethical, legal, social and organisational HTA domains using the following publicly available sources:

In addition to the systematic search results the following publicly available sources were used:

- European Thyroid Association (https://www.eurothyroid.com/)
- Thyroid Federation International (https://thyroid-fed.org/)
- American Thyroid association (https://www.thyroid.org/)

9.1.3 *Assessment of quality of evidence*

N.A.

9.1.4 *Methodology data extraction, analysis and synthesis of the domains ethical, legal, social and organisational issues*

The synthesis of issues related to the ethical, legal, social, and organisational domains are presented in the sections of findings for each respective domain. Findings are narratively summarised.

9.2 Results ethical, legal, social and organisational issues

9.2.1 *PRISMA flow diagram*

N.A.

9.2.2 *Study characteristics of included studies*

The included studies consisted of one peer-reviewed research papers, one conference abstract, and one news article.

9.2.3 *Evidence table*

N.A.

9.2.4 *Findings ethical issues*

Rosenthal et al. [73] reported several ethical issues which can occur in patients with hypothyroidism. Some of these issues may also be applicable to subclinical hypothyroid patients.

The first issue occurs due to (limited) mental impairment on behalf of the patient which can hamper their ability to provide informed consent and shared decision-making. The authors wrote that this impairment can be present in patients with mild forms of (clinical) hypothyroidism, which has small clinical differences from SCH.

Another ethical issue that might arise stems from diagnostic challenges of (slowly) progressing diseases such as (hypo)thyroidism. For example, under diagnosis of patients showing signs of thyroiditis or Graves' disease, misdiagnosis, or diagnosis of a mild case can cause difficulty in decisions regarding best time to initiate therapy. In practice, some patients are diagnosed with SCH as a result of misdiagnosis, when in fact they may benefit from a mild (clinical) hypothyroidism diagnosis and medical intervention already.

The last issue arises when patients themselves claim to have hypothyroidism, despite clinical evidence that present them as euthyroid. As is generally the case, a healthcare practitioner should engage in truth telling and trust building with the patient, and should not prescribe unnecessary medications and/or therapies, even if the patients are persistent. This ethical issue may also apply to SCH patients, if they claim to have progressed to (clinical) hypothyroidism, despite the lack of clinical evidence.

9.2.5 *Findings legal issues*

Casassus et al. [74] provides insight into an ongoing legal issue. In France, over 300'000 patients reported adverse events after the pharmaceutical manufacturer, Merck, replaced the excipients in the Levoythyrox brand of levothyroxine. The active ingredient of Levoythyrox remained unchanged. However, adverse events including chronic fatigue, hair and/or memory loss, heart palpitations, weight gain, insomnia, dizziness, and nausea were reported after the excipient change.

Approximately 4'000 patients filed suit against the manufacturer in 2017. The manufacturer lost this case, and the local appeal court ordered the manufacturer to pay €1'000 (£853; $1'180) compensation to each patient for the anxiety caused by the adverse events due to this new formulation. In the first half of 2022, a decision is expected by the French supreme court.

The study indicated that Levothyrox generics could not replace the original version, and concluded that the manufacturer's average quality bioequivalence studies "did not show or justify scientifically that the

new formula was interchangeable". Especially for a drug like Levoythyrox, which exhibits a narrow therapeutic area, it is important that generics are interchangeable with the branded drug.

9.2.6 *Findings social issues*

Social issues applicable within the context of levothyroxine treatment of SCH patients were not identified in the literature search.

9.2.7 *Findings organisational issues*

Den Elzen et al. [75] investigated the treatment variation of general practitioners (GPs) concerning older patients with SCH, based on patient's characteristics and country. 524 GPs participated in the survey, of which 262 GPs were from Switzerland.

In general, GPs were less inclined to start treatment in 85-year-old than in 70-year-old women. Women with a TSH of 15 mU/L were more likely to get treated than women with TSH 6 mU/L. Differences in treatment strategy were observed between countries. The Swiss GPs, as well as German and Irish GPs, were more inclined to initiate treatment than GPs in England, New Zealand and the Netherlands. These differences between countries were most pronounced for the TSH levels of 6 mU/L.

10 Additional issues

10.1 Guideline Recommendations

Treatment recommendations on when to treat and how to treat SCH from four guidelines are summarised in the ***Table 11***.

Table 11. Overview guidelines on treatment of (subclinical) hypothyroidism*

	When to treat	How to treat	Dose adjustments
ATA/AACE, 2012[23] **Clinical practice guidelines for hypothyroidism in adults**	**Subclinical hypothyroidism** • TSH >10 mIU/L • TSH between upper limit of a given laboratory's reference range and 10 mIU/L and additional criteria[4]	**Overt hypothyroidism** • Treat with L-thyroxine monotherapy • Young healthy adults: full replacement dose • 50-60 years without evidence of coronary heart disease: L-thyroxine dose of 50 µg daily **Subclinical hypothyroidism** • 25 to 75 µg daily (depending on the degree of TSH elevation).	**Subclinical hypothyroidism** Further adjustments should be guided by clinical response and follow up laboratory determinations including TSH values

[4] Symptoms suggestive of hypothyroidism, positive TPOAb or evidence of atherosclerotic cardiovascular disease, heart failure, or associated risk factors for these diseases

ATA task force, 2014[38] **Guidelines for the Treatment of Hypothyroidism**	Not reported	**Hypothyroidism** • Treat with levothyroxine • Dose depends on patient's weight, lean body mass, pregnancy status, etiology of hypothyroidism, degree of thyrotropin elevation, age, general clinical context and serum TSH goal appropriate for the clinical situation. • In general: start with low doses • Start with initial full replacement or as partial replacement, with gradual increments in the dose titrated upward.	**Hypothyroidism** After low dose start: dose titrated slowly based on serum TSH. Dose adjustments should be made when there are large changes in body weight, with aging, and with pregnancy, with thyrotropin assessment 4–6 weeks after any dosage change.

ETA guideline, 2013[3]
 	
Management of Subclinical Hypothyroidism
	Subclinical hypothyroidism
	• Age >65y and serum TSH ≥10 mU/L and clear symptoms hypothyroidism/ high vascular risk
	• Age <65y and serum TSH ≥10 mU/L
	• Age <65y and serum TSH < 10 mU/L and hypothyroid symptoms (3 months trial, re-evaluate thereafter)
	• Patients with persistent SCH and diffuse or nodular goitre.
		Subclinical hypothyroidism
		• Treat with oral L -thyroxine monotherapy daily
		• Weight-related dose for patients without cardiac disease: ~1.5 µg/kg/day
		  ○ 75 or 100 µg/day for a woman
		  ○ 100 or 125 µg for a man
		• For patients with cardiac disease and elderly: 25 or 50 µg daily
			Subclinical hypothyroidism
			For patients with cardiac disease and elderly: increased by 25 µg/day every 14–21 days until full replacement dose is reached.
			For patients aged <65y and serum TSH <10 mU/L and hypothyroid symptoms: If symptoms does not improve and serum TSH levels are within reference range: stop treatment.

NICE guideline, 2019[17] **NG145: Thyroid disease: assessment and management**	**Subclinical hypothyroidism** • Adults with persistent elevated TSH ≥10 mIU/L.	**Overt hypothyroidism** • Treat with levothyroxine monotherapy • Weight-related dose for patients aged <65y without history of cardiovascular disease: 1.6 µg/kg/day • For patients aged ≥65y with history of cardiovascular disease: starting dose 25 to 50 µg/day with titration. **Subclinical hypothyroidism** • Treat with levothyroxine monotherapy • For patients <65y with persistent TSH above reference range but < 10 mIU/L, and with symptoms of hypothyroidism: 6-month trial of levothyroxine.	**Hypothyroidism** Adjust dose if symptoms persist **Subclinical hypothyroidism** If symptoms does not improve and serum TSH levels remain high: adjust dose If symptoms does not improve and serum TSH levels are within reference range: stop treatment.

Keys: AACE = American Association Of Clinical Endocrinologists; ATA = America Thyroid Association; ETA = European thyroid association; TSH = thyroid-stimulating hormone; y = years. *In RCTs evaluating the effects of levothyroxine treatment, the age and TSH level thresholds applied as eligibility criteria vary and do not necessarily correspond with those in the guidelines.

10.2 Ongoing Trials

A search on clinicaltrial.gov (access date 31 May 2022) did not reveal any ongoing randomized clinical trials that investigated our research question.

11 Discussion

This HTA evaluated the efficacy, safety, and budget impact of the use of levothyroxine in patients diagnosed with SCH. In this section, the main findings, strengths, and limitations are discussed.

The systematic literature search for studies on the efficacy, effectiveness, and safety of the use of levothyroxine in patients diagnosed with SCH in multiple peer-reviewed scientific literature databases retrieved 28 articles based on 22 unique studies. Articles including only women with specific sex hormonal states, e.g. pregnant women, non-pregnant women on fertility-related treatment, or menopausal women were out of scope. The body of evidence was substantial, and the risk of bias ranged from low to high. A rigorous methodology, adhering to international methodological standards such as Cochrane and PRISMA, was applied to identify, critically appraise, analyse, and summarise the available evidence.

Regarding the TSH assessment, available data showed a statistically significant reduction in serum TSH levels in response to levothyroxine treatment in patients with SCH and favoured levothyroxine treatment compared to placebo or no treatment. There is a large body of evidence for no clinically meaningful patient-relevant effects of levothyroxine treatment. The majority of the articles did not find a statistically significant effect of levothyroxine treatment. Three articles based on three different studies reported a statistically significant result in favour of levothyroxine compared to placebo for four efficacy outcomes related to symptoms of hypothyroidism, neurological features and physical aspects of HRQoL. The risk of bias of the studies was high. The available articles did not show major safety issues of levothyroxine treatment. Most studies did not stratify subjects by age, gender and/or TSH level. The few articles that did stratify, found no effect of levothyroxine either. It was not possible to calculate pooled of estimates, since almost all outcomes reported by two or more articles were based on the same study (TRUST study) and different assessment techniques or tools were used to assess the same outcome with its own scoring system. Moreover, outcomes were not reported in the same way (i.e. mean SD and mean SE) or different follow-up periods were used (i.e.; 6 months and 12 months).

One challenge of this HTA report was the variation in assessment techniques or tools used to assess the efficacy outcomes of levothyroxine treatment. It is possible that more overlap between study results were detected when similar assessment techniques and tools had been applied. For example, only one of the seven applied attention and memory tests reported a favourable effect of levothyroxine treatment. Other factors that contributed to the heterogeneity between studies were the use of different durations of follow-up (three months to three years), differences in age and sex distribution (inclusion of women only; excluding elderly), and different inclusion criteria (definition of SCH; duration of thyroid dysfunction). The definition of SCH varied between studies: some studies selected patients with TSH

levels between 4.0- 10 µU/ml, or between 4.6-19.9 µU/ml, while others included all patients with a TSH level above 3.6 µU/ml. In patients with relatively mild degree of SCH (i.e. lower TSH levels), the effect of levothyroxine might be less evident compared to the effect in patients with more pronounced SCH. In most articles, the participants with TSH >10 µU/ml accounted for a small percentage of the study population and the results may thus not be generalizable to this subgroup with markedly elevated TSH levels. The RCTs included patients based on TSH and fT4 tests, but one might argue that more evident effect of levothyroxine was observed if patients were recruited based on symptoms or a combination thereof.

The extracted data was not suitable to apply common clinically relevant strata across articles due to the heterogeneity of the data and the wide variety of outcomes. In total, 19 RCTs applied inclusion criteria on gender and/or age. Eight RCTs included women only. The articles varied in choice of efficacy or safety outcomes. Fourteen RCTs applied criteria on age. Seven RCTs included SCH patients aged ≥65 year; six of them used data from the TRUST, inducing overlap in patients. Of the remaining seven RCTs with restrictions on age, four RCTs excluded elderly, using a maximum age range from 60 to 75 years. The only similar outcome in the 11 RCTs that applied criteria on age was the safety outcome 'number of patients with at least one adverse event'.

Most RCTs included patients with persistent elevated TSH levels measured at several occasions, which has been emphasized as an important inclusion criterion, since TSH levels can fluctuate and can even normalise over time in some SCH patients. [4,26,27,33] Sometimes however, the recruitment into the study was based on a single thyroid function test. There is therefore a possibility that ineligible euthyroid patients may have been recruited and hence reducing the change of detecting an existing clinical effect of levothyroxine treatment. [26] Because of the small sample size (i.e. almost 40% studies have ≤25 patients per study arm), many studies were underpowered and did not have sufficient ability to detect the differences between the levothyroxine and placebo groups [4,49,52,55] However, seven studies were included with a sample size above 100 patients per study arm.

The objectives of the included clinical studies were diverse and did often not correspond with our PICO outcomes. For example, the effect of levothyroxine on surrogate efficacy markers such as carotid intima media thickness, homocysteine metabolism, lipoprotein profile and osteocalcin levels were investigated. In less than 50% of the articles identified in the systematic literature search, efficacy outcomes could be extracted. TSH itself was reported in almost all articles and treated as an analytical outcome rather than a clinical efficacy outcome. In the included articles, TSH levels were assessed to monitor the dosage of levothyroxine, and the majority of the studies adjusted the treatment dose based on intermitted TSH levels in the levothyroxine group to result in TSH levels within the specific reference ranges.

Several studies had revealed an increased risk for CV and neuromuscular events in people with SCH. [63,76,77] Although not confirmed by the RCTs included in this HTA, there is some evidence suggesting that levothyroxine treatment has beneficial effect on these outcomes in SCH patients [1,32,33,78]. For this reason, part of the results classified as safety outcomes in the included RCTs were reported in the efficacy section of the HTA report. Being underpowered and too short follow-up period might be the reason no included articles reported a beneficial effect of levothyroxine on CV outcomes or other (safety) outcomes. Side effects from levothyroxine will likely result from overtreatment and could lead to symptoms associated with hyperthyroidism.

The cost-effectiveness systematic review searches did not identify any published cost-effectiveness analyses for levothyroxine in patients diagnosed with SCH. A *de novo* model was developed to calculate the budget impact of levothyroxine for adult SCH patients, between ages 18 and 99 years, in Switzerland over a time horizon of five years (years 2020-2024).

There are no reliable epidemiological data on the number of SCH patients in Switzerland, nor on the percentage of SCH patients receiving levothyroxine. We based our estimates on general population estimates from the Swiss FSO, [72] and European prevalence rates from the literature. . [20] For the budget impact analysis, we used a percentage range (10% - 90%) of the estimated prevalent population eligible for levothyroxine treatment in Switzerland. Costs for the treatment with levothyroxine were calculated as a function of drug unit prices and an average defined daily dose of 50µg per patient with SCH. No additional healthcare costs were assumed. Drug unit prices were sourced from the Spezialitätenliste (www.spezialitätenliste.ch), published by the FOPH on June 13, 2022. To calculate drug acquisition costs, we used the public price (i.e. Publikumspreis). Since the dosage of Novothyral tablets were not provided in the Spezialitätenliste, information on this dosage was sourced from the website swissmedicinfo.ch.

After multiplying the population assumed to be eligible for levothyroxine with the yearly cost of levothyroxine treatment (based on an assumed DDD of 50ug), the total costs of treatment ranged between 9’701’140 CHF and 87’310’263 CHF (10% - 90% range estimation) over a time horizon of five years.

Other HTA domains (ethical, legal, social and organisational) were searched to identify issues related to the treatment of levothyroxine in SCH patients. From the three studies included, identified issues were: the dilemma of physicians when confronted with patients that either lack the ability to provide patient consent or participate in decision-making regarding treatment, or when patients disagree with clinical diagnostic results; a substantial risk of misdiagnosis given the fact that thyroid disorders tend to

develop slowly; the importance of bioequivalence of replacement or generic drugs; and international differences in prescription strategies for SCH patients.

Overall, the limitations of this study are mostly related to available data. Analyses on effectiveness in specific group of patients with associated risk factors and comorbid conditions were not feasible given lack of appropriate data to perform a subpopulation scenario analysis.

12 Conclusions

Levothyroxine treatment successfully lowers raised levels of serum TSH in SCH patients but evidence of a consistent (positive) treatment effect on other clinical or patient-relevant outcomes like hypothyroid associated symptoms or cardiovascular outcomes is lacking. There are no indications of major safety issues related to levothyroxine treatment.

The budget impact analysis showed that total costs of levothyroxine for SCH patients ranged from approximately 9 Mio CHF to 87 Mio CHF over 5 years, depending mainly on the number of patients having levothyroxine being prescribed for their subclinical status of hypothyroidism.

13 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. *N Engl J Med* 2017;**376**:2534–44. doi:10.1056/NEJMoa1603825

5 de Freitas J. Development and validation of in vitro bioassays for thyroid hormone receptor mediated endocrine disruption. 2012.https://edepot.wur.nl/240578 (accessed 10 Jan 2022).

6 Gesing A. The thyroid gland and the process of aging. *Thyroid Res* 2015;**8**:A8. doi:10.1186/1756-6614-8-S1-A8

7 da Costa VMC, Rosenthal D. Effects of aging on thyroidal function and proliferation. *Curr Aging Sci* 2008;**1**:101–4. doi:10.2174/1874609810801020101

8 Mariotti S, Franceschi C, Cossarizza A, *et al.* The aging thyroid. *Endocr Rev* 1995;**16**:686–715. doi:10.1210/edrv-16-6-686

9 Bremner AP, Feddema P, Leedman PJ, *et al.* Age-Related Changes in Thyroid Function: A Longitudinal Study of a Community-Based Cohort. *The Journal of Clinical Endocrinology & Metabolism* 2012;**97**:1554–62. doi:10.1210/jc.2011-3020

10 Vanderpump MPJ, Tunbridge WMG. Epidemiology and prevention of clinical and subclinical hypothyroidism. *Thyroid* 2002;**12**:839–47. doi:10.1089/105072502761016458

11 Vanderpump MP, Tunbridge WM, French JM, *et al.* The incidence of thyroid disorders in the community: a twenty-year follow-up of the Whickham Survey. *Clin Endocrinol (Oxf)* 1995;**43**:55–68. doi:10.1111/j.1365-2265.1995.tb01894.x

12 Sawin CT, Castelli WP, Hershman JM, *et al.* The aging thyroid. Thyroid deficiency in the Framingham Study. *Arch Intern Med* 1985;**145**:1386–8.

13 Roelfsema F, Pereira AM, Veldhuis JD, *et al.* Thyrotropin Secretion Profiles Are Not Different in Men and Women. *The Journal of Clinical Endocrinology & Metabolism* 2009;**94**:3964–7. doi:10.1210/jc.2009-1155

14 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

15 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

16 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

17 NICE Guideline. Thyroid Disease: assessment and management. 2019.https://www.nice.org.uk/guidance/ng145/resources/thyroid-disease-assessment-and-management-pdf-66141781496773 (accessed 12 May 2021).

18 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.

19 UCLA H. UCLA Endocrine Center Thyroid Hormone Levels. https://www.uclahealth.org/endocrine-center/ (accessed 12 May 2021).

20 Garmendia Madariaga A, Santos Palacios S, Guillén-Grima F, *et al.* The Incidence and Prevalence of Thyroid Dysfunction in Europe: A Meta-Analysis. *The Journal of Clinical Endocrinology & Metabolism* 2014;**99**:923–31. doi:10.1210/jc.2013-2409

21 Canaris GJ, Manowitz NR, Mayor G, *et al.* The Colorado Thyroid Disease Prevalence Study. *Archives of Internal Medicine* 2000;**160**:526–34. doi:10.1001/archinte.160.4.526

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

23 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

24 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

25 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

26 Parle J, Roberts L, Wilson S, *et al.* A randomized controlled trial of the effect of thyroxine replacement on cognitive function in community-living elderly subjects with subclinical hypothyroidism: The Birmingham elderly thyroid study. *Journal of Clinical Endocrinology and Metabolism* 2010;**95**:3623–32. doi:10.1210/jc.2009-2571

27 Díez JJ, Iglesias P. Spontaneous subclinical hypothyroidism in patients older than 55 years: an analysis of natural course and risk factors for the development of overt thyroid failure. *J Clin Endocrinol Metab* 2004;**89**:4890–7. doi:10.1210/jc.2003-032061

28 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

29 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

30 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

31 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

32 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

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

34 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

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

36 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

37 Ross D. Subclinical hypothyroidism in nonpregnant adults. https://www.uptodate.com/contents/subclinical-hypothyroidism-in-nonpregnant-adults (accessed 5 Mar 2022).

38 Jonklaas J, Bianco AC, Bauer AJ, *et al.* Guidelines for the Treatment of Hypothyroidism: Prepared by the American Thyroid Association Task Force on Thyroid Hormone Replacement. *Thyroid* 2014;**24**:1670–751. doi:10.1089/thy.2014.0028

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

40 Perez CLS, Araki FS, Graf H, *et al.* Serum Thyrotropin Levels Following Levothyroxine Administration at Breakfast. *Thyroid* 2013;**23**:779–84. doi:10.1089/thy.2012.0435

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

42 Higgins J, Green S. *Cochrane Handbook for Systematic Reviews of Intervnetions Version 5.1.0 [updated March 2011]*. The Cochrane Collaboration, 2011. www.training.cochrane.org/handbook

43 Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. http://www.prisma-statement.org (accessed 9 Mar 2021).

44 Page MJ, McKenzie JE, Bossuyt PM, *et al.* The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ* 2021;:n71. doi:10.1136/bmj.n71

45 Schunemann 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

46 Meier C, Staub JJ, Roth CB, *et al.* TSH-controlled L-thyroxine therapy reduces cholesterol levels and clinical symptoms in subclinical hypothyroidism: A double blind, placebo-controlled trial (basel thyroid study). *Journal of Clinical Endocrinology and Metabolism* 2001;**86**:4860–6. doi:10.1210/jc.86.10.4860

47 Iqbal A, Jorde R, Figenschau Y. Serum lipid levels in relation to serum thyroid-stimulating hormone and the effect of thyroxine treatment on serum lipid levels in subjects with subclinical hypothyroidism: The Tromsø Study. *Journal of Internal Medicine* 2006;**260**:53–61. doi:10.1111/j.1365-2796.2006.01652.x

48 Jorde R, Waterloo K, Storhaug H, *et al.* Neuropsychological function and symptoms in subjects with subclinical hypothyroidism and the effect of thyroxine treatment. *Journal of Clinical Endocrinology and Metabolism* 2006;**91**:145–53. doi:10.1210/jc.2005-1775

49 Blum MR, Gencer B, Adam L, *et al.* Impact of thyroid hormone therapy on atherosclerosis in the elderly with subclinical hypothyroidism: A randomized trial. *Journal of Clinical Endocrinology and Metabolism* 2018;**103**:2988–97. doi:10.1210/jc.2018-00279

50 Mooijaart SP, Du Puy RS, Stott DJ, *et al.* Association between Levothyroxine Treatment and Thyroid-Related Symptoms among Adults Aged 80 Years and Older with Subclinical Hypothyroidism. *JAMA - Journal of the American Medical Association* 2019;**322**:1977–86. doi:10.1001/jama.2019.17274

51 Stuber MJ, Moutzouri E, Feller M, *et al.* Effect of Thyroid Hormone Therapy on Fatigability in Older Adults With Subclinical Hypothyroidism: A Nested Study Within a Randomized Placebo-Controlled Trial. *The journals of gerontology Series A, Biological sciences and medical sciences* 2020;**75**:e89–94. doi:10.1093/gerona/glaa123

52 Zijlstra LE, Jukema JW, Westendorp RGJ, *et al.* Levothyroxine Treatment and Cardiovascular Outcomes in Older People With Subclinical Hypothyroidism: Pooled Individual Results of Two Randomised Controlled Trials. *Frontiers in Endocrinology* 2021;**12**. doi:10.3389/fendo.2021.674841

53 Wildisen L, Feller M, Del Giovane C, *et al.* Effect of Levothyroxine Therapy on the Development of Depressive Symptoms in Older Adults with Subclinical Hypothyroidism: An Ancillary Study of a Randomized Clinical Trial. *JAMA Network Open* 2021;**4**. doi:10.1001/jamanetworkopen.2020.36645

54 Mikhail GS, Alshammari SM, Alenezi MY, *et al.* Increased atherogenic low-density lipoprotein cholesterol in untreated subclinical hypothyroidism. *Endocrine Practice* 2008;**14**:570–5. doi:10.4158/EP.14.5.570

55 Aghili R, Khamseh ME, Malek M, *et al.* Changes of subtests of Wechsler Memory Scale and cognitive function in subjects with subclinical hypothyroidism following treatment with levothyroxine. *Archives of Medical Science* 2012;**8**:1096–101. doi:10.5114/aoms.2012.32423

56 Najafi L, Malek M, Hadian A, *et al.* Depressive symptoms in patients with subclinical hypothyroidism-the effect of treatment with levothyroxine: A double-blind randomized clinical trial. *Endocrine Research* 2015;**40**:121–6. doi:10.3109/07435800.2014.896924

57 Rezaee R, Mohebbi M, Afkhamizadeh M, *et al.* The effect of levothyroxine in comparison with placebo on serum osteocalcin levels in patients with subclinical hypothyroidism. *Current drug safety* Published Online First: 2021. doi:10.2174/1574886316666210525102230

58 Caraccio N, Ferrannini E, Monzani F. Lipoprotein profile in subclinical hypothyroidism: Response to levothyroxine replacement, a randomized placebo-controlled study. *Journal of Clinical Endocrinology and Metabolism* 2002;**87**:1533–8. doi:10.1210/jc.87.4.1533

59 Caraccio N, Natali A, Sironi A, *et al.* Muscle metabolism and exercise tolerance in subclinical hypothyroidism: A controlled trial of levothyroxine. *Journal of Clinical Endocrinology and Metabolism* 2005;**90**:4057–62. doi:10.1210/jc.2004-2344

60 Christ-Crain M, Meier C, Huber PR, *et al.* Effect of L-thyroxine replacement therapy on surrogate markers of skeletal and cardiac function in subclinical hypothyroidism. *Endocrinologist* 2004;**14**:161–6.

61 Kong WM, Sheikh MH, Lumb PJ, *et al.* A 6-month randomized trial of thyroxine treatment in women with mild subclinical hypothyroidism. *American Journal of Medicine* 2002;**112**:348–54. doi:10.1016/S0002-9343(02)01022-7

62 Meek S, Smallridge RC. Effect of thyroid hormone replacement on methionine-stimulated homocysteine levels in patients with subclinical hypothyroidism: A randomized, double-blind, placebo-controlled study. *Endocrine Practice* 2006;**12**:529–34. doi:10.4158/EP.12.5.529

63 Monzani F, Di Bello V, Caraccio N, *et al.* Effect of levothyroxine on cardiac function and structure in subclinical hypothyroidism: A double blind, placebo-controlled study. *Journal of Clinical Endocrinology and Metabolism* 2001;**86**:1110–5. doi:10.1210/jc.86.3.1110

64 Monzani F, Caraccio N, Kozàkowà M, *et al.* Effect of Levothyroxine Replacement on Lipid Profile and Intima-Media Thickness in Subclinical Hypothyroidism: A Double-Blind, Placebo-Controlled Study. *Journal of Clinical Endocrinology and Metabolism* 2004;**89**:2099–106. doi:10.1210/jc.2003-031669

65 Nagasaki T, Inaba M, Yamada S, *et al.* Decrease of brachial-ankle pulse wave velocity in female subclinical hypothyroid patients during normalization of thyroid function: A double-blind, placebo-controlled study. *European Journal of Endocrinology* 2009;**160**:409–15. doi:10.1530/EJE-08-0742

66 Yazici M, Gorgulu S, Sertbas Y, *et al.* Effects of thyroxin therapy on cardiac function in patients with subclinical hypothyroidism: Index of myocardial performance in the evaluation of left ventricular function. *International Journal of Cardiology* 2004;**95**:135–43. doi:10.1016/j.ijcard.2003.05.015

67 Cabral MD, Teixeira P, Soares D, *et al.* Effects of thyroxine replacement on endothelial function and carotid artery intima-media thickness in female patients with mild subclinical hypothyroidism. *Clinics* 2011;**66**:1321–7. doi:10.1590/S1807-59322011000800003

68 Duman D, Sahin S, Esertas K, *et al.* Simvastatin improves endothelial function in patents with subclinical hypothyroidism. *Heart and Vessels* 2007;**22**:88–93. doi:10.1007/s00380-006-0950-0

69 Mainenti MRM, Vigário PS, Teixeira PFS, *et al.* Effect of levothyroxine replacement on exercise performance in subclinical hypothyroidism. *Journal of Endocrinological Investigation* 2009;**32**:470–3. doi:10.3275/6106

70 Zhao M, Liu L, Wang F, *et al.* A Worthy Finding: Decrease in Total Cholesterol and Low-Density Lipoprotein Cholesterol in Treated Mild Subclinical Hypothyroidism. *Thyroid* 2016;**26**:1019–29. doi:10.1089/thy.2016.0010

71 Reuters VS, Almeida CP, Teixeira PFS, *et al.* Effects of subclinical hypothyroidism treatment on psychiatric symptoms, muscular complaints, and quality of life. *Arquivos Brasileiros de Endocrinologia e Metabologia* 2012;**56**:128–36. doi:10.1590/S0004-27302012000200006

72 Swiss Federal Statistical Office. Population scenarios for Switzerland and cantons 2020-2050. https://www.bfs.admin.ch/bfs/en/home/statistics/catalogues-databases/press-releases.gnpdetail.2020-0194.html

73 Rosenthal MS. Ethical issues in the management of thyroid disease. *Endocrinol Metab Clin North Am* 2014;**43**:545–64. doi:10.1016/j.ecl.2014.02.013

74 Casassus B. French to take legal action against drug regulator over levothyroxine. *BMJ* 2021;:n2243. doi:10.1136/bmj.n2243

75 den Elzen WPJ, Lefèbre–van de Fliert AA, Virgini V, *et al.* International variation in GP treatment strategies for subclinical hypothyroidism in older adults. *European Geriatric Medicine* 2013;**4**:S147. doi:10.1016/j.eurger.2013.07.482

76 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

77 Monzani F, Caraccio N, Del Guerra P, *et al.* Neuromuscular symptoms and dysfunction in subclinical hypothyroid patients: beneficial effect of L-T4 replacement therapy. *Clin Endocrinol (Oxf)* 1999;**51**:237–42. doi:10.1046/j.1365-2265.1999.00790.x

78 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

14 Appendices

14.1 Search strategy efficacy, effectiveness, and safety

Table 12. Search strategy for the efficacy, effectiveness, and safety systematic literature search: PubMed (MEDLINE)

Population: diagnosed with subclinical hypothyroidism	("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] OR comprehensive*[tiab]) AND (bibliographic*[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])	*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 real-life[tiab] OR “case-control studies”[Mesh] OR case-control[tiab] OR casecontrol[tiab] OR case-comparison[tiab] OR case-referent[tiab])
	Publication period: 2000 – September 13, 2021	
	Language: English, German, French, Dutch	

Table 13. Search strategy for the efficacy, effectiveness, and safety systematic literature search: Embase.com

Population: diagnosed with subclinical hypothyroidism	('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 doubleblind*: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 re-view*: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 – September 13, 2021	
	Language: English, German, French, Dutch	

14.2 Excluded studies during full text selection

Table 14. Excluded studies found with the efficacy, effectiveness, and safety systematic literature search

Reference	Reason for exclusion
Abreu IM, Lau E, Pinto BS, et al. Subclinical hypothyroidism: To treat or not to treat, that is the question! a systematic review with meta-analysis on lipid profile. Endocrine Connections 2017;6(3):188-99. doi: 10.1530/EC-17-0028	Systematic review
Adrees M, Gibney J, El-Saeity N, et al. Effects of 18 months of L-T4 replacement in women with subclinical hypothyroidism. Clinical endocrinology 2009;71(2):298-303.	No RCT
Anagnostis P, Efstathiadou Z, Slavakis A, et al. The effect of L-thyroxine substitution on lipid profile, glucose homeostasis, inflammation and coagulation in patients with subclinical hypothyroidism. International journal of clinical practice 2014;68(7):857-63	No RCT
Bauer D, Rodondi N, Kearney P, et al. Thyroid hormone therapy does not improve hypothyroid symptoms in older adults with subclinical hypothyroidism even among those with more symptoms at baseline: The trust randomized trial. Thyroid 2017;27:A13. doi: 10.1089/thy.2017.29046.abstracts	Abstract
Bekkering GE, Agoritsas T, Lytvyn L, et al. Thyroid hormones treatment for subclinical hypothyroidism: A clinical practice guideline. BMJ (Online) 2019;365 doi: 10.1136/bmj.l2006.	Non-pertinent publication type (guideline)
Beyhan Z, Ertürk K, Üçkaya G, et al. Restoration of euthyroidism does not improve cardiovascular risk factors in patients with subclinical hypothyroidism in the short term. Journal of endocrinological investigation 2006;29(6):505-10.	No RCT
Çakal B, Cakal E, Demirbaş B, et al. Homocysteine and fibrinogen changes with L-thyroxine in subclinical hypothyroid patients. Journal of Korean medical science 2007;22(3):431-35.	No RCT
Canpolat U, Yorgun H, Sunman H, et al. Evaluation of cardiac autonomic functions by heart rate variability, QT dispersion and heart rate recovery index in patients with subclinical hypothyroidism. International Journal of Cardiology 2011;147:S113. doi: 10.1016/S0167-5273(11)70315-1	Abstract
Chen Y, Tai HY. Levothyroxine in the treatment of overt or subclinical hypothyroidism: A systematic review and meta-analysis. Endocrine Journal 2020;67(7):719-32. doi: 10.1507/endocrj.EJ19-0583	Systematic review
Christ-Crain M, Meier C, Guglielmetti M, et al. Elevated C-reactive protein and homocysteine values: Cardiovascular risk factors in hypothyroidism? A cross-sectional and a double-blind, placebo-controlled trial. Atherosclerosis 2003;166(2):379-86. doi: 10.1016/S0021-9150(02)00372-6	Duplicate data
Christ-Crain M, Meier C, Huber P, et al. Effect of restoration of euthyroidism on peripheral blood cells and erythropoietin in women with subclinical hypothyroidism. Hormones (Athens) 2003;2(4):237-42. doi: 10.14310/horm.2002.11105 [published Online First: 2006/09/28]	Duplicate data
Christ-Crain M, Morgenthaler NG, Meier C, et al. Pro-A-type and N-terminal pro-B-type natriuretic peptides in different thyroid function states.	Duplicate data

Swiss Medical Weekly 2005;135(37-38):549-54.	
De Montmollin M, Feller M, Beglinger S, et al. Do patients with subclinical hypothyroidism who suffer from greater symptom burden benefit from levothyroxine therapy? Further analysis of the randomized placebo-controlled trust trial. Journal of General Internal Medicine 2019;34(2):S192-S93. doi: 10.1007/11606.1525-1497	Non-pertinent publication type (presentation)
de Montmollin M, Feller M, Beglinger S, et al. L-Thyroxine Therapy for Older Adults With Subclinical Hypothyroidism and Hypothyroid Symptoms: Secondary Analysis of a Randomized Trial. Annals of internal medicine 2020;172(11):709-16. doi: 10.7326/M19-3193	Irrelevant subgroup analysis of a RCT included in this review
Deicher R, Vierhapper H. Homocysteine: a risk factor for cardiovascular disease in subclinical hypothyroidism? Thyroid 2002;12(8):733-36.	No RCT
Dejanović M, Ivetić V, Nestorović V, et al. The value of P300 event related potentials in the assessment of cognitive function in subclinical hypothyroidism. Minerva Endocrinologica 2017;42(1):15-23. doi: 10.23736/S0391-1977.16.02327-0	No RCT
Deshmukh VC, Varthakavi PK. Role of l-thyroxine therapy in treatment of subclinical hypothyroidism (SCHT): 6-month randomized controlled trial. Thyroid 2012;22:A52-A53. doi: 10.1089/thy.2012.2209.abs	Abstract
Duman D, Demirtunc R, Sahin S, et al. The effects of simvastatin and levothyroxine on intima-media thickness of the carotid artery in female normolipemic patients with subclinical hypothyroidism: A prospective, randomized-controlled study. Journal of Cardiovascular Medicine 2007;8(12):1007-11. doi: 10.2459/JCM.0b013e3282f03bc1	No comparison of interest
Faber J, Petersen L, Wiinberg N, et al. Hemodynamic changes after levothyroxine treatment in subclinical hypothyroidism. Thyroid 2002;12(4):319-24.	No RCT
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(13):1349-59. doi: 10.1001/jama.2018.13770 [published Online First: 2018/10/05]	Systematic review
Franzoni F, Galetta F, Fallahi P, et al. Effect of l-thyroxine treatment on left ventricular function in subclinical hypothyroidism. Biomedicine and Pharmacotherapy 2006;60(8):431-36. doi: 10.1016/j.biopha.2006.07.010	Insufficient methodological quality (unclear/inconsistent flow of patient numbers, not reported whether baseline values were comparable between study arms, no quantitative data reported on outcome of interest)
Galetta F, Franzoni F, Fallahi P, et al. Heart rate variability and QT dispersion in patients with subclinical hypothyroidism. Biomedicine and Pharmacotherapy 2006;60(8):425-30. doi: 10.1016/j.biopha.2006.07.009	No comparison of interest
Gencer B, Moutzouri E, Blum MR, et al. The Impact of Levothyroxine on Cardiac Function in Older Adults With Mild Subclinical Hypothyroidism: A Randomized Clinical Trial. American Journal of Medicine 2020;133(7):848-56.e5. doi: 10.1016/j.amjmed.2020.01.018	Duplicate data
Gencer B, Moutzouri E, Blum MR, et al. The impact of levothyroxine on cardiac function in older adults with subclinical hypothyroidism: A randomized clinical trial. European Heart Journal 2019;40:359. doi: 10.1093/eurheartj/ehz747.0357	Abstract
He W, Li S, Zhang JA, et al. Effect of levothyroxine on blood pressure in patients with subclinical hypothyroidism: A systematic review and meta-	Systematic review

analysis. Frontiers in Endocrinology 2018;9(AUG) doi: 10.3389/fendo.2018.00454	
Hennessey JV, Espaillat R. Reversible morbidity markers in subclinical hypothyroidism. Postgraduate Medicine 2015;127(1):78-91. doi: 10.1080/00325481.2015.998158	Narrative review
Iwen KA, Brabant G. Thyroid hormone therapy in old age. Internist 2020 doi: 10.1007/s00108-020-00790-4	Narrative review
Jensovsky J, Ruzicka E, Spackova N, et al. Changes of event related potential and cognitive processes in patients with subclinical hypothyroidism after thyroxine treatment. Endocrine Regulations 2002;36(3):115-22.	No RCT
Joffe RT, Pearce EN, Hennessey JV, et al. Subclinical hypothyroidism, mood, and cognition in older adults: A review. International Journal of Geriatric Psychiatry 2013;28(2):111-18. doi: 10.1002/gps.3796	Narrative review
Kebapcilar L, Comlekci A, Tuncel P, et al. Effect of levothyroxine replacement therapy on paraoxonase-1 and carotid intima-media thickness in subclinical hypothyroidism. Medical Science Monitor 2009;16(1):CR41-CR47.	No RCT
Kim S-K, Kim S-H, Park K-S, et al. Regression of the increased common carotid artery-intima media thickness in subclinical hypothyroidism after thyroid hormone replacement. Endocrine journal 2009;56(6):753-58.	No RCT
Kong WM, Sheikh MH, Lumb PJ, et al. Erratum: A 6-month randomized trial of thyroxine treatment in women with mild subclinical hypothyroidism (American Journal of Medicine (2002) 112 (348-354) PII: S0002934302010227). American Journal of Medicine 2002;113(3):264. doi: 10.1016/S0002-9343(02)01240-8	Non-pertinent publication type (erratum)
Köroglu B, Bagci Ö, Ersoy I, et al. Effects of levothyroxine treatment on cardiovascular risk profile and carotid intima media thickness in patients with subclinical hypothyroidism. Acta Endocrinologica (1841-0987) 2012;8(3)	Insufficient methodological quality (control group not further specified, i.e. not clear if the comparator was placebo or no treatment)
Leng O, Razvi S. Treatment of subclinical hypothyroidism: assessing when treatment is likely to be beneficial. Expert Rev Endocrinol Metab 2021;16(2):73-86. doi: 10.1080/17446651.2020.1738924 [published Online First: 2020/03/29]	Narrative review
Li X, Meng Z, Jia Q, et al. Effect of L-thyroxine treatment versus a placebo on serum lipid levels in patients with sub-clinical hypothyroidism. Biomedical Reports 2016;5(4):443-49. doi: 10.3892/br.2016.745	Systematic review
Li X, Wang Y, Guan Q, et al. The lipid-lowering effect of levothyroxine in patients with subclinical hypothyroidism: A systematic review and meta-analysis of randomized controlled trials. Clinical Endocrinology 2017;87(1):1-9. doi: 10.1111/cen.13338	Systematic review
Liu L, Yu Y, Zhao M, et al. Benefits of levothyroxine replacement therapy on nonalcoholic fatty liver disease in subclinical hypothyroidism patients. International Journal of Endocrinology 2017;2017 doi: 10.1155/2017/5753039	No comparison of interest (no comparator in the study arm with significant SCH (TSH ≥10 µU/ml))
Mainenti MRM, Teixeira PFS, Oliveira FP, et al. Effect of hormone replacement on exercise cardiopulmonary reserve and recovery performance in subclinical hypothyroidism. Brazilian Journal of Medical and Biological Research 2010;43(11):1095-101. doi: 10.1590/S0100-879X2010007500116	Duplicate data

Martins RM, Fonseca RHA, Duarte MMT, et al. Impact of subclinical hypothyroidism treatment in systolic and diastolic cardiac function. Arquivos Brasileiros de Endocrinologia e Metabologia 2011;55(7):460-67. doi: 10.1590/S0004-27302011000700005	No comparison of interest (no statistical comparison for outcome of interest)
Meier C, Beat M, Guglielmetti M, et al. Restoration of euthyroidism accelerates bone turnover in patients with subclinical hypothyroidism: A randomized controlled trial. Osteoporosis International 2004;15(3):209-16. doi: 10.1007/s00198-003-1527-8	Duplicate data
Meier C, Christ-Crain M, Guglielmetti M, et al. Prolactin Dysregulation in Women with Subclinical Hypothyroidism: Effect of Levothyroxine Replacement Therapy. Thyroid 2003;13(10):979-85. doi: 10.1089/105072503322511391	Duplicate data
Niknam N, Khalili N, Khosravi E, et al. Endothelial dysfunction in patients with subclinical hypothyroidism and the effects of treatment with levothyroxine. Adv Biomed Res 2016;5:38. doi: 10.4103/2277-9175.178783 [published Online First: 2016/04/22]	No RCT
Özben B, Toprak A, Yavuz D, et al. Effects of restoration of the euthyroid state on epicardial adipose tissue and carotid intima media thickness in subclinical hypothyroid patients. Endocrine 2015;48(3):909-15.	No RCT
Ozcan O, Cakir E, Yaman H, et al. The effects of thyroxine replacement on the levels of serum asymmetric dimethylarginine (ADMA) and other biochemical cardiovascular risk markers in patients with subclinical hypothyroidism. Clinical endocrinology 2005;63(2):203-06.	No RCT
Peng CCH, Huang HK, Wu BBC, et al. Association of Thyroid Hormone Therapy with Mortality in Subclinical Hypothyroidism: A Systematic Review and Meta-Analysis. Journal of Clinical Endocrinology and Metabolism 2021;106(1):292-303. doi: 10.1210/clinem/dgaa777	Systematic review
Perez A, Cubero J, Sucunza N, et al. Emerging cardiovascular risk factors in subclinical hypothyroidism: lack of change after restoration of euthyroidism. Metabolism 2004;53(11):1512-15.	No RCT
Razvi S, Ingoe L, Keeka G, et al. The beneficial effect of L-thyroxine on cardiovascular risk factors, endothelial function, and quality of life in subclinical hypothyroidism: Randomized, crossover trial. Journal of Clinical Endocrinology and Metabolism 2007;92(5):1715-23. doi: 10.1210/jc.2006-1869	No washout periods between study interventions in cross-over trial
Rodriguez EG, Löwe AL, Del Giovane C, et al. Impact of thyroid hormone therapy on bone health in older adults with subclinical hypothyroidism: A Randomized clinical trial. Journal of Bone and Mineral Research 2018;33:317-18.	Abstract
Rodriguez EG, Stuber M, Giovane CD, et al. Skeletal effects of levothyroxine for subclinical hypothyroidism in older adults: A TRUST randomized trial nested study. Journal of Clinical Endocrinology and Metabolism 2020;105(1):1-8. doi: 10.1210/clinem/dgz058	Duplicate data
Schneider R, Reiners C. The Effect of Levothyroxine Therapy on Bone Mineral Density: A Systematic Review of the Literature. Experimental and Clinical Endocrinology and Diabetes 2003;111(8):455-70. doi: 10.1055/s-2003-44704	Systematic review
Shahebrahimi K, Naseri R, Kalantarian TS, et al. Effects of levothyroxine treatment on lipid profile in subclinical hypothyroidism: A randomized clinical trial. Galen Medical Journal 2015;4(2):72-77.	Insufficient methodological quality (unclear when TSH measure is performed, i.e. at 3 months at end of treatment, or at 6 months assay)

Shatynska-Mytsyk I, Kyyak Y, Mytsyk Y. Subclinical hypothyroidism as an independent risk factor for coronary heart disease in postmenopausal women. European Journal of Cardiovascular Nursing 2010;9:S35. doi: 10.1016/S1474-5151(10)60124-7	Abstract
Siddhanta S, Sahana PK, Goswami S, et al. A prospective study of effect of levothyroxine replacement on cardiovascular risk factors and quality of life in subjects with subclinical hypothyroidism. Indian Journal of Endocrinology and Metabolism 2017;21(8):S57.	Abstract
Stuber MJ, Moutzouri E, Feller M, et al. Effect of thyroid hormone replacement on fati-gability in older adults with subclinical hypothy-roidism: A randomized placebo controlled trial. Journal of General Internal Medicine 2019;34(2):S201-S02. doi: 10.1007/11606.1525-1497	Non-pertinent publication type (presentation)
Teixeira PFS, Reuters VS, Ferreira MM, et al. Treatment of subclinical hypothyroidism reduces atherogenic lipid levels in a placebo-controlled double-blind clinical trial. Hormone and Metabolic Research 2008;40(1):50-55. doi: 10.1055/s-2007-993216	No comparison of interest (no statistical comparison for outcome of interest)
Teixeira PDFDS, Reuters VS, Ferreira MM, et al. Lipid profile in different degrees of hypothyroidism and effects of levothyroxine replacement in mild thyroid failure. Translational Research 2008;151(4):224-31. doi: 10.1016/j.trsl.2007.12.006	No comparison of interest (no statistical comparison for outcome of interest)
Unal O, Erturk E, Ozkan H, et al. Effect of levothyroxine treatment on QT dispersion in patients with subclinical hypothyroidism. Endocrine practice : official journal of the American College of Endocrinology and the American Association of Clinical Endocrinologists 2007;13(7):711-15.	No RCT
Ünal OK, Erturk E, Sarandöl E, et al. Is there any beneficial effect of L-thyroxine replacement theraphy on cardiovascular risk factors in patients with subclinical hypothyroidism? Endocrine Abstracts 2010;20:P85.	Abstract
Unsal IO, Topaloglu O, Cakir E, et al. Effect of L-thyroxin therapy on thyroid volume and carotid artery intima-media thickness in the patients with subclinical hypothyroidism. Journal of Medical Disorders 2014;2:1.	Duplicate article
Unsal IO, Topaloglu O, Colak NB, et al. Effect of l-thyroxin therapy on thyroid volume and carotid artery intima-media thickness in patients with subclinical hypothyroidism. Endocrine Reviews 2012;33(3)	No RCT
Van Harten AC, Leue C, Verhey FRJ. Should depressive symptoms in patients with subclinical hypothyroidism be treated with thyroid hormone? Tijdschrift voor Psychiatrie 2008;50(8):539-43.	Systematic review
Villar HCCE, Saconato H, Valente O, et al. Thyroid hormone replacement for subclinical hypothyroidism. Cochrane Database of Systematic Reviews 2007(3) doi: 10.1002/14651858.CD003419.pub2	Systematic review
Watt J. In older adults with subclinical hypothyroidism, levothyroxine did not improve symptoms. Annals of Internal Medicine 2020;172(8):JC44. doi: 10.7326/ACPJ202004210-044	Non-pertinent publication type (comment)
Watt J. In older adults with subclinical hypothyroidism, levothyroxine therapy did not reduce depressive symptoms at 12 mo. Annals of internal medicine 2021;174(7):JC81. doi: 10.7326/ACPJ202107200-081	Non-pertinent publication type (comment)
Yetmiş M, Kazancioğlu R, Erkoç R, et al. Changes in lipid profile and body mass index in patients with subclinical hypothyroidism: Evaluation of L-Thyroxine treatment. Haseki Tip Bulteni 2011;49(4):131-35.	Language out of scope (article in Turkish)
Zhao M, Liu L, Guan Q, et al. Effects of l-thyroxine replacement therapy on serum lipid profiles in patients with mild subclinical hypothyroidism: An	Abstract

open-label, randomized, controlled trial. Endocrine Reviews 2016;37(2) doi: 10.1210/endo-meetings.2016.THPTA.13.OR11-5	
Zhao T, Chen B, Zhou Y, et al. Effect of levothyroxine on the progression of carotid intima-media thickness in subclinical hypothyroidism patients: A meta-analysis. BMJ Open 2017;7(10) doi: 10.1136/bmjopen-2017-016053	Systematic review
Zhao T, Chen BM, Zhao XM, et al. Subclinical hypothyroidism and depression: a meta-analysis. Translational Psychiatry 2018;8(1) doi: 10.1038/s41398-018-0283-7	Systematic review
Zhou Y, Chen Y, Cao X, et al. Association between plasma homocysteine status and hypothyroidism: A meta-analysis. International Journal of Clinical and Experimental Medicine 2014;7(11):4544-53.	Systematic review

14.3 Efficacy results

14.3.1 *TSH*

Table 15 Efficacy results on levothyroxine treatment: TSH levels

First author, year (study name; gender; age) **Reference range TSH in µU/ml** **Risk of bias RCT**	**Intervention (duration)** **Comparator (duration)**	**Sample size**	**TSH level** **Mean (SD) at baseline, µU/ml** **Mean (SD) at end of trial, µU/ml**	**p-value for comparison of baseline TSH level between treatment arms**	**p-value for comparison of end of trial TSH level between treatment arms**	**Comparison of difference in adjusted mean change in TSH level between treatment arms in µU/ml (95% CI); p-value**
Aghili et al., 2012 [55] (NR; mix; 18-64y) <4.5 High risk of bias	Levothyroxine (3 months)	30	8.39 (4.09)	*ns***	NR	NR; NR
		30	2.01 (1.34)			
	p-value (within group): <0.001					
	Placebo (3 months)	30	8.12 (3.20)			
		30	7.83 (5.17)			
	p-value (within group): 0.733					
Blum et al., 2018 [49] (nested study in TRUST in Switzerland; mix; ≥65y) NR Low risk of bias	Levothyroxine (12 months)	109	6.24 (1.73)	*ns***	NR	-1.64 (NR); <0.001
		96	3.55 (2.14)			
	p-value (within group): NR					
	Placebo (12 months)	108	6.47 (2.17)			
		89	5.29 (2.21)			
	p-value (within group): NR					
Cabral et al., 2011 [67] (NR; women; NA*) 0.4-4.0 High risk of bias	Levothyroxine (12 months)	14	6.79 (2.0)	*ns***	NR	NR; NR
		14	3.02 (0.89)†			
	p-value (within group): <0.05					
	No treatment (12 months)	18	6.77 (1.96)			
		18	6.69 (3.51)†			
	p-value (within group): *ns***					

First author, year (study name; gender; age) Reference range TSH in µU/ml Risk of bias RCT	Intervention (duration) Comparator (duration)	Sample size	TSH level Mean (SD) at baseline, µU/ml Mean (SD) at end of trial, µU/ml	p-value for comparison of baseline TSH level between treatment arms	p-value for comparison of end of trial TSH level between treatment arms	Comparison of difference in adjusted mean change in TSH level between treatment arms in µU/ml (95% CI); p-value
Caraccio et al., 2002 [58] (NR; mix; NA‡) 0.30-3.6 High risk of bias	Levothyroxine (6 months)	24	6.00 (NR)	*ns***	<0.0001	NR; NR
		24	1.52 (NR)			
	p-value (within group): <0.0001					
	Placebo (6 months)	25	4.90 (NR)			
		25	4.86 (NR)			
	p-value (within group): *ns***					
Caraccio et al., 2005 [59] (NR; mix; NA§) 0.30-3.6 High risk of bias	Levothyroxine (6 months)	12	4.38 (SEM 0.16)	*ns***	<0.0001	NR; NR
		12	1.28 (SEM 0.14)			
	p-value (within group): <0.001					
	Placebo (6 months)	11	4.95 (SEM 0.34)			
		11	5.08 (SEM 0.56)			
	p-value (within group): *ns***					
Duman et al., 2007 [68] (NR; women; NA‖) 0.27-4.2 High risk of bias	Levothyroxine (8 months)	20	10.9 (5.8)	*ns***	NR	NR; NR
		20	2.0 (1.1)			
	p-value (within group): <0.001					
	No treatment (8 months)	19	11.0 (7.5)			
		19	10.9 (6.8)			
	p-value (within group): *ns***					
Iqbal et al., 2006 [47] (Tromsø; mix; >29y) 0.2-4.3 High risk of bias	Levothyroxine (12 months)	32	5.8 (1.8)	*ns***	NR	NR; NR
		32	1.5 (1.4)			
	p-value (within group): <0.001					
	Placebo (12 months)	32	5.4 (1.3)			
		32	5.4 (2.0)			
	p-value (within group): *ns***					
Kong et al., 2002 [61] (NR; women; ≥18y) NR Moderate risk of bias	Levothyroxine (6 months)	23	8.0 (1.5)	0.2	NR	-3.0 (-3.3 to -0.6)¶; NR
		20	NR¶			
	p-value (within group): NR					
	Placebo (6 months)	17	7.3 (1.6)			
		15	NR¶			
	p-value (within group): NR					
Mainenti et al., 2009 [69] main (NR; women; 30-60y) NR High risk of bias	Levothyroxine (6 months)	11	7.50 (2.52)	*ns***	NR	NR; NR
		11	3.08 (1.03)			
	p-value (within group): 0.01					
	No treatment (6 months)	12	7.35 (1.93)			
		12	5.80 (1.38)			
	p-value (within group): *ns***					

First author, year (study name; gender; age) Reference range TSH in µU/ml Risk of bias RCT	Intervention (duration) Comparator (duration)	Sample size	TSH level Mean (SD) at baseline, µU/ml Mean (SD) at end of trial, µU/ml	p-value for comparison of baseline TSH level between treatment arms	p-value for comparison of end of trial TSH level between treatment arms	Comparison of difference in adjusted mean change in TSH level between treatment arms in µU/ml (95% CI); p-value
Meek et al., 2006 [62] (NR; mix; ≥18y) 0.3-5.0 High risk of bias	Levothyroxine (6 months)	12	Median (IQR): 6.7 (NR)	0.52	<0.001	NR; NR
		12	Median (IQR): 1.8 (NR)			
	p-value (within group): NR					
	Placebo (6 months)	12	Median (IQR): 6.6 (NR)			
		12	Median (IQR): 6.0 (NR)			
	p-value (within group): NR					
Meier et al., 2001 [46] (Basel Thyroid study; women; 18-75y) 0.1-4.0 High risk of bias	Levothyroxine (48 weeks)	31	12.8 (SEM 1.4)	*ns***	NR	NR; NR
		31	3.1 (SEM 0.3)			
	p-value (within group): <0.001					
	Placebo (48 weeks)	32	10.7 (SEM 0.9)			
		32	9.9 (SEM 0.6)			
	p-value (within group): 0.108					
Mikhail et al., 2008 (NR; mix; 15-60y) [54] 0.3-4.0 High risk of bias	Levothyroxine (52 weeks)	60	6.40 (1.41)	0.87	0.0001	NR; NR
		60	2.41 (0.86)			
	p-value (within group): NR					
	Placebo (52 weeks)	60	6.32 (1.38)			
		60	6.06 (1.34)			
	p-value (within group): NR					
Monzani et al., 2001 [63] (NR; mix; NA#) 0.3-3.6 High risk of bias	Levothyroxine (6 months)	10	5.44 (2.41)	*ns***	0.0001	NR; NR
		10	1.32 (0.47)			
	p-value (within group): 0.0001					
	Placebo (6 months)	10	4.74 (1.14)			
		10	4.55 (1.20)			
	p-value (within group): *ns***					
Monzani et al., 2004 [64] (NR; mix; NA**) 0.3-3.6 High risk of bias	Levothyroxine (6 months)	23	Median (IQR): 6.03 (NR)	*ns***	<0.0001	NR; NR
		23	Median (IQR): 1.32 (NR)			
	p-value (within group): <0.0001					
	Placebo (6 months)	22	Median (IQR): 5.68 (NR)			
		22	Median (IQR): 6.01 (NR)			
	p-value (within group): *ns***					

First author, year (study name; gender; age) Reference range TSH in µU/ml Risk of bias RCT	Intervention (duration) Comparator (duration)	Sample size	TSH level Mean (SD) at baseline, µU/ml Mean (SD) at end of trial, µU/ml	p-value for comparison of baseline TSH level between treatment arms	p-value for comparison of end of trial TSH level between treatment arms	Comparison of difference in adjusted mean change in TSH level between treatment arms in µU/ml (95% CI); p-value
Mooijaart et al., 2019 [50] (TRUST & IEMO; mix; ≥80y) 0.4-4.6 Moderate risk of bias	Levothyroxine (12 months)	90	6.50 (1.80)	NR	NR	-1.97 (-2.49 to -1.45); <0.001
		90	3.69 (1.81)			
	p-value (within group): NR					
	Placebo (12 months)	122	6.20 (1.48)			
		122	5.49 (2.21)			
	p-value (within group): NR					
Nagasaki et al., 2009 [65] (NR; women; NA[††]) 0.4-4.7 High risk of bias	Levothyroxine (5 months)	48	7.32 (SEM 0.64)	NR	NR	NR; NR
		48	2.7 (SEM 0.21)			
	p-value (within group): <0.0001					
	Placebo (5 months)	47	7.25 (SEM 0.69)			
		47	7.01 (SEM 0.67)			
	p-value (within group): *ns****					
Najafi et al., 2015 [56] (NR; mix; NA[‡‡]) <4.5 High risk of bias	Levothyroxine (12 weeks)	30	8.29 (4.9)	NR	NR	NR; NR
		30	2.01 (1.34)			
	p-value (within group): 0.000					
	Placebo (12 weeks)	30	8.12 (3.12)			
		30	7.82 (5.17)			
	p-value (within group): 0.733					
Parle et al., 2010 [26] (The Birmingham elderly thyroid study; mix; ≥65y) 0.4-5.5 Moderate risk of bias	Levothyroxine (12 months)	52	Median (IQR): 6.6 (6.0-8.5)	*ns****	0.0002	NR; NR
		49	Median (IQR): 3.7 (2.8-4.9)			
	p-value (within group): NR					
	Placebo (12 months)	42	Median (IQR): 6.65 (5.9-8.3)			
		36	Median (IQR): 5.45 (3.9-9.2)			
	p-value (within group): NR					
Rezaee et al., 2021 [57] (NR; mix; NA[§§]) NR Moderate risk of bias	Levothyroxine (2 months)	14	7.89 (2.94)	0.03	0.66	NR; NR
		14	5.60 (3.20)[‖‖]			
	p-value (within group): <0.01					
	Placebo (2 months)	14	5.99 (0.68)			
		14	5.15 (2.04)[‖‖]			
	p-value (within group): 0.09					
Stott et al., 2017 [4] (original TRUST study; mix; ≥65y)	Levothyroxine (12 months)	368	6.41 (2.01)	*ns****	NR	-1.92 (-2.24 to -1.59); <0.001
		318	3.63 (2.11)			
	p-value (within group): NR					

First author, year (study name; gender; age) Reference range TSH in µU/ml Risk of bias RCT	Intervention (duration) Comparator (duration)	Sample size	TSH level Mean (SD) at baseline, µU/ml Mean (SD) at end of trial, µU/ml	p-value for comparison of baseline TSH level between treatment arms	p-value for comparison of end of trial TSH level between treatment arms	Comparison of difference in adjusted mean change in TSH level between treatment arms in µU/ml (95% CI); p-value
0.40-4.59 Moderate risk of bias	Placebo (12 months)	369	6.38 (2.01)			
		320	5.48 (2.48)			
	p-value (within group): NR					
Stuber et al., 2020 [51] (nested study in TRUST in Ireland & Switzerland; mix; ≥65y) NR Low risk of bias	Levothyroxine (12 months)	119	6.08 (1.80)	*ns***	<0.001	NR; NR
		119	3.08 (1.32			
	p-value (within group): NR					
	Placebo (12 months)	111	6.29 (2.01)			
		111	5.30 (2.34)			
	p-value (within group): NR					
Yazici et al., 2004 [66](NR; mix; NA¶¶) 0.4-4.0 High risk of bias	Levothyroxine (12 months)	22	8.47 (1.9)	*ns***	<0.0001	NR; NR
		22	2.41 (1.3)			
	p-value (within group): <0.0001					
	Placebo (12 months)	23	8.39 (2.3)			
		23	8.40 (2.3)			
	p-value (within group): NR					
Zhao et al., 2016 [70] (NR; mix; ≥40y) 0.27-4.2 High risk of bias	Levothyroxine (15 months)	215	Median (IQR): 5.96 (2.24)	0.188	NR	NR; NR
		210	Median (IQR): 2.78 (1.43)			
	p-value (within group): <0.001					
	No treatment (15 months)	163	Median (IQR): 5.50 (1.85)			
		159	Median (IQR): 4.88 (2.26)			
	p-value (within group): <0.001					
Zijlstra et al., 2021 [52] (TRUST & IEMO; mix; ≥65y or ≥80y) NR Low risk of bias	Levothyroxine (12 months)	420	NR##	NR	<0.001	NR; NR
		363	3.66 (2.1)			
	p-value (within group): NR					
	Placebo (12 months)	422	NR##			
		368	5.66 (3.3)			
	p-value (within group): NR					

Keys: CI = confidence interval, IEMO = Institute for Evidence-Based Medicine in Old Age, NA = not applicable, NR = not reported, *ns* = not significant, RCT = randomised controlled trial, SD = standard deviation, TRUST = Thyroid Hormone Replacement for Subclinical Hypothyroidism, TSH = thyroid-stimulating hormone, y = years. * Age was not an inclusion or exclusion criterion; the mean (SD) age at baseline was 43.36 (9.8) years in the intervention group and 47.59 (8.4) years in the no treatment group. † The p-value for the mean changes of THS values during the study period between the levothyroxine and no treatment group is: <0.05, change in TSH values not provided. ‡ Age was not an inclusion or exclusion criterion; for the intervention and placebo group the mean (SD) age at baseline was 34.5 (9.1) years. § Age was not an inclusion or exclusion criterion; the mean (SEM) age at baseline

was 34 (2) years in the intervention group and 30 (2) years in the placebo group. ‖ Age was not an inclusion or exclusion criterion; the mean (SD) age at the analysis was 36 (11) years in the intervention group and 35 (14) years in the no treatment group. ¶ Mean (SD) change from baseline within levothyroxine group: -4.6 (2.3) µU/ml; mean (SD) change from baseline within comparator group (placebo): -1.7 (2.0) µU/ml; the p-value for the changes within the levothyroxine and comparator (placebo) group is: <0.001. # Age was not an inclusion or exclusion criterion; the mean (SD) age at baseline was 34.3 (12.3) years in the intervention group and 29.2 (9.4) years in the placebo group. ** Age was not an inclusion or exclusion criterion; for the intervention and placebo group the mean (SD) age at baseline was 37 (11) years. †† Age was not an inclusion or exclusion criterion; the mean (SEM) age at baseline was 64.4 (2.59) years in the intervention group and 66.0 (3.0) years in the placebo group. ‡‡ Age was not an inclusion or exclusion criterion; the mean (SD) age at baseline was 32.47 (11.35) years in the intervention group and 36.07 (11.35) years in the placebo group. §§ Age was not an inclusion or exclusion criterion; the mean (SD) age at the analysis was 35.07 (9.94) years in the intervention group and 31.30 (4.30) years in the placebo group. ‖‖ The p-value for the changes in THS values of the levothyroxine and placebo group is: 0.05, change in TSH values not provided. Because the levothyroxine group and the placebo group were statistically significant different at baseline, the statistically significant result of the within-levothyroxine group comparison was not counted. ¶¶ Age was not an inclusion or exclusion criterion; the mean (SD) age at baseline was 39.7 (8.7) years in the intervention group and 40.2 (9.3) years in the placebo group. ## Mean (SD) TSH baseline value for both intervention and placebo group were 6.38 (5.7) µU/ml; *** Full p-value not available. Statistically significant results

Table 16. Efficacy results on levothyroxine treatment: TSH levels stratified by gender

First author, year (study name; gender; age in years) Reference range TSH in µU/ml Risk of bias RCT	Intervention (duration) Comparator (duration)	Sample size	TSH level Mean (SD) at baseline, µU/ml Mean (SD) at end of trial, µU/ml	p-value for comparison of baseline TSH level between treatment arms	p-value for comparison of end of trial TSH level between treatment arms	Comparison of difference in adjusted mean change in TSH level between treatment arms in µU/ml (95% CI); p-value
Stott et al., 2017 [4] (original TRUST study; mix; ≥65y) 0.40-4.59 Moderate risk of bias	No stratification					
	Levothyroxine (12 months)	368	6.41 (2.01)	*ns**	NR	-1.92 (-2.24 to -1.59); <0.001
		318	3.63 (2.11)			
	p-value (within group): NR					
	Placebo (12 months)	369	6.38 (2.01)			
		320	5.48 (2.48)			
	p-value (within group): NR					
	Stratified by gender: Men					
	Levothyroxine (12 months)	157	6.37 (1.87)	*ns**	NR	-1.83 (-2.41 to -1.25); <0.001
		157	3.76 (1.97)			
	p-value (within group): NR					
	Placebo (12 months)	155	6.34 (2.02)			
		155	5.58 (2.55)			
	p-value (within group): NR					
	Stratified by gender: Women					
	Levothyroxine (12 months)	174	6.41 (2.16)	*ns**	NR	-2.14 (-2.69 to -1.58); <0.001
		174	3.55 (2.22)			
	p-value (within group): NR					
	Placebo (12 months)	173	6.35 (1.81)			
		173	5.67 (4.06)			
	p-value (within group): NR					

Keys: CI = confidence interval, NR = not reported; *ns* = not significant, RCT = randomised controlled trial, SD = standard deviation, TRUST = Thyroid Hormone Replacement for Subclinical Hypothyroidism, TSH = thyroid-stimulating hormone, y = years. * full p-value not available Statistically significant results

Table 17. Efficacy results on levothyroxine treatment: TSH levels stratified by TSH

First author, year (study name; gender; age) Reference range TSH in µU/ml Risk of bias RCT	Intervention (duration) Comparator (duration)	Sample size	TSH level Mean (SD) at baseline, µU/ml Mean (SD) at end of trial, µU/ml	p-value for comparison of baseline TSH level between treatment arms	p-value for comparison of end of trial TSH level between treatment arms	Comparison of difference in adjusted mean change in TSH level between treatment arms in µU/ml (95% CI); p-value
Christ-Crain et al., 2004 [60] (NR; women; 18-75y) 0.3-4.0 Moderate risk of bias	No stratification					
	Levothyroxine (48 weeks)	31	14.1 (9.6) 3.1 (1.6)	NR	NR	NR; NR
	p-value (within group): NR					
	Placebo (48 weeks)	31	10.6 (3.8) NR*			
	p-value (within group): NR					
	Stratified by TSH level: TSH <12 µU/ml					
	Levothyroxine (48 weeks)	14 14	8.8 (2.2) 3.1 (1.2)	NR	NR	NR; NR
	p-value (within group): <0.001					
	Placebo (48 weeks)	23 23	8.9 (2.2) 8.3 (2.1)			
	p-value (within group): *ns*†					
	Stratified by TSH level: TSH >12 µU/ml					
	Levothyroxine (48 weeks)	17 17	21.5 (11.5) 3.0 (2.1)	NR	NR	NR; NR
	p-value (within group): <0.001					
	Placebo (48 weeks)	8 8	15.4 (2.4) 13.1 (2.6)			
	p-value (within group): *ns*†					
Stott et al., 2017 [4] (original TRUST study; mix; ≥65y) 0.40-4.59 Moderate risk of bias	No stratification					
	Levothyroxine (12 months)	368 318	6.41 (2.01) 3.63 (2.11)	*ns*†	NR	-1.92 (-2.24 to -1.59); <0.001
	p-value (within group): NR					
	Placebo (12 months)	369 320	6.38 (2.01) 5.48 (2.48)			
	p-value (within group): NR					
	Stratified by TSH level: <7 µU/ml					
	Levothyroxine (12 months)	256 256	5.55 (0.66) 3.4 (1.7)	*ns*†	NR	-1.61 (-2.07 to -1.16); <0.001
	p-value (within group): NR					
	Placebo (12 months)	250 250	5.52 (0.64) 5.0 (3.3)			

First author, year (study name; gender; age) Reference range TSH in µU/ml Risk of bias RCT	Intervention (duration) Comparator (duration)	Sample size	TSH level Mean (SD) at baseline, µU/ml Mean (SD) at end of trial, µU/ml	p-value for comparison of baseline TSH level between treatment arms	p-value for comparison of end of trial TSH level between treatment arms	Comparison of difference in adjusted mean change in TSH level between treatment arms in µU/ml (95% CI); p-value
	p-value (within group): NR					
	Stratified by TSH level: 7-9.99 µU/ml					
	Levothyroxine (12 months)	57	8.18 (0.89)	*ns*†	NR	-3.01 (-3.95 to -2.07); <0.001
		57	3.99 (2.26)			
	p-value (within group): NR					
	Placebo (12 months)	62	8.08 (0.84)			
		62	6.95 (2.18)			
	p-value (within group): NR					
	Stratified by TSH level: >9.99 µU/ml					
	Levothyroxine (12 months)	18	12.78 (2.43)	*ns*†	NR	-4.19 (-5.95 to -2.43); <0.001
		18	6.10 (3.94)			
	p-value (within group): NR					
	Placebo (12 months)	16	12.48 (2.35)			
		16	10.06 (5.15)			
	p-value (within group): NR					

Keys: CI = confidence interval, NR = not reported; *ns* = not significant, RCT = randomised controlled trial, SD = standard deviation, TRUST = Thyroid Hormone Replacement for Subclinical Hypothyroidism, TSH = thyroid-stimulating hormone, y = years. *No change in TSH could be seen in patients with placebo according to the study authors; † Full p-value not available. Statistically significant results

14.3.2 *Diagnostic index for hypothyroidism*

Table 18. Efficacy results on levothyroxine treatment: diagnostic index for hypothyroidism

First author, year (study name; gender; age) Risk of bias RCT	Intervention (duration) Comparator (duration)	Sample size	Billewicz score* Mean (SEM) at baseline, score Mean (SEM) at end of trial, score	p-value for comparison of score at end of trial between treatment arms	Zulewski score† Mean (SEM) at baseline, score Mean (SEM) at end of trial, score	p-value for comparison of score at end of trial between treatment arms
Meier et al., 2001 [46] (Basel Thyroid Study; women; 18-75y) High risk of bias	Levothyroxine (48 weeks)	31	-25.7 (2.7)	0.31	2.1 (0.3)	0.53
		31	-32.1 (2.1)		1.5 (0.2)	
	Placebo (48 weeks)	31	-28.3 (2.5)		2.0 (0.2)	
		31	-30.8 (2.5)		1.6 (0.2)	

Keys: RCT = randomised controlled trial, SEM = standard error of the mean, vs = versus, y = years. * Euthyroidism indicated by score of ≤-30 points, borderline hypothyroidism by -29 to +24 points, clinical hypothyroidism by ≥25 points. †Euthyroidism indicated by score of 0-1 point, borderline hypothyroidism by 2-5 points, clinical hypothyroidism by >=5 points, including an age-correcting factor.

14.3.3 *Hypothyroid symptoms*

Table 19. Efficacy results on levothyroxine treatment: ThyPRO scores

First author, year (study name; gender; age) Risk of bias RCT	Intervention (duration) Comparator (duration)	Sample size	ThyPRO Hypothyroid Symptoms score*		ThyPRO Tiredness score†	
			Mean (SD) at baseline, score Mean (SD) at end of trial, score	Difference in change in adjusted mean score (95% CI) between treatment arms; p-value	Mean (SD) at baseline, score Mean (SD) at end of trial, score	Difference in change in adjusted mean score (95% CI) between treatment arms; p-value
Mooijaart et al., 2019 [50] (TRUST and IEMO; mix; ≥80y) Moderate risk of bias	Levothyroxine (12 months)	90	21.7 (19.5) 19.3 (18.2)	1.27 (−2.69 to 5.23); 0.53	25.2 (21.5) 28.2 (20.0)	-0.10 (-4.51 to 4.31); 0.96
	Placebo (12 months)	122	19.8 (19.6) 17.4 (18.1)		25.1 (19.5) 28.7 (19.9)	
Stott et al., 2017 [4] (original TRUST study; mix; ≥65y) Moderate risk of bias	No stratification					
	Levothyroxine (12 months)	368 318	17.5 (18.8) 16.6 (16.9)	0.0 (-2.0 to 2.1); 0.99	25.9 (20.6) 28.7 (20.2)	0.4 (-2.1 to 2.9); 0.77
	Placebo (12 months)	369 320	16.9 (17.9) 16.7 (17.5)		25.5 (20.3) 28.6 (19.5)	
	Stratified by gender: Men					
	Levothyroxine (12 months)	157	12.6 (14.6) 14.1 (15.8)	1.6 (-1.4 to 4.6); *ns*§	20.9 (16.5) 26.8 (19.0)	1.6 (-2.0 to 5.2); *ns*§
	Placebo (12 months)	155	14.3 (15.5) 13.5 (15.8)		21.9 (19.2) 25.8 (18.8)	
	Stratified by gender: Women					
	Levothyroxine (12 months)	174	18.7 (19.5) 19.6 (18.5)	- 2.0 (-4.8 to 0.8); *ns*‡	29.0 (22.3) 31.0 (21.6)	-0.4 (-3.8 to 3.0); *ns*‡
	Placebo (12 months)	173	20.9 (20.5) 20.3 (18.6)		29.2 (20.7) 31.6 (20.2)	
	Stratified by TSH level: <7 µU/ml					
	Levothyroxine (12 months)	256	17.1 (17.7) 17.4 (17.2)	-0.8 (-3.1 to 1.6); *ns*‡	25.2 (20.1) 30.3 (20.9)	0.9 (-1.9 to 3.7); *ns*‡
	Placebo (12 months)	250	17.0 (17.9) 18.1 (18.0)		5.9 (20.4) 29.9 (19.8)	
	Stratified by TSH level: 7-9.99 µU/ml					
	Levothyroxine (12 months)	57	18.1 (21.5) 16.4 (18.8)	1.2 (-3.6 to 6.0); *ns*‡	25.1 (19.3) 24.4 (16.3)	-0.8 (-6.5 to 5.0); *ns*‡
	Placebo (12 months)	62	14.4 (16.2) 12.9 (13.7)		25.5 (21.4) 24.8 (19.3)	
	Stratified by TSH level: >9.99 µU/ml					
	Levothyroxine (12 months)	18	12.2 (18.2) 14.2 (16.4)	0.6 (-8.3 to 9.6); *ns*‡	25.7 (24.8) 28.8 (20.3)	-2.8 (-13.6 to 8.0); *ns*‡

First author, year (study name; gender; age) Risk of bias RCT	Intervention (duration) Comparator (duration)	Sample size	ThyPRO Hypothyroid Symptoms score*		ThyPRO Tiredness score†	
			Mean (SD) at baseline, score Mean (SD) at end of trial, score	Difference in change in adjusted mean score (95% CI) between treatment arms; p-value	Mean (SD) at baseline, score Mean (SD) at end of trial, score	Difference in change in adjusted mean score (95% CI) between treatment arms; p-value
	Placebo (12 months)	16	20.2 (22.5) 19.1 (23.4)		25.0 (15.8) 28.8 (20.3)	

Keys: CI = confidence interval, IEMO = Institute for Evidence-Based Medicine in Old Age, NR = not reported; *ns* = not significant, RCT = randomised controlled trial, SD = standard deviation, ThyPRO = thyroid-specific patient reported outcome, TRUST = Thyroid Hormone Replacement for Subclinical Hypothyroidism, y = years. * 0-100 scale, higher scores indicating more symptoms; † 0-100 scale, higher scores indicating more symptoms; ‡ Full p-value not available.

Table 20a. Efficacy results on levothyroxine treatment: Change of symptoms

First author, year (study name; gender; age) Risk of bias RCT	Intervention (duration) Comparator (duration)	Sample size	If comparing with 1y ago, do you feel warmer/no change/ colder? - Warmer - No change - Colder, in n (%)	p-value	If comparing with 1y ago, is your skin less dry/no change/more dry? - Less dry - No change - More dry, in n (%)	p-value	If comparing with 1y ago, are your clothes looser/no change/ tighter? - Looser - No change - Tighter, in n (%)	p-value	If comparing with 1y ago, is your hair less dry/no change/more dry? - Less dry - No change - More dry, in n (%)	p-value	If comparing with 1y ago, are your muscles stronger/no change/ weaker? - Stronger - No change - Weaker, in n (%)	p-value	If comparing with 1y ago, are your nails less brittle/ no change/more brittle? - Less brittle - No change - More brittle, in n (%)	p-value
Jorde et al., 2006 [48] (Tromsø; mix; >29y) High risk of bias	Levothyroxine (12 months)	36	6 (16.7%) 27 (75.0%) 3 (8.3%)	*ns*§	4 (11.1%) 25 (69.4%) 7 (19.4%)	*ns*§	10 (27.8%) 24 (66.7%) 2 (5.6%)	<0.05	3 (8.3%) 25 (69.4%) 8 (22.2%)	*ns*§	0 (0.0%) 32 (88.9%) 4 (11.1%)	*ns*§	3 (8.3%) 29 (80.6%) 4 (11.1%)	*ns*§
	Placebo (12 months)	33	4 (12.1%) 26 (78.8%) 3 (9.1%)		1 (3.0%) 27 (81.8%) 5 (15.2%)		4 (12.1%) 22 (66.7%) 7 (21.2%)		1 (3.0%) 28 (84.8%) 4 (12.1%)		2 (6.1%) 26 (78.8%) 5 (15.2%)		1 (3.1%)† 26 (81.3%)† 5 (15.6%)†	

Table 20b. Efficacy results on levothyroxine treatment: Change of symptoms - Continued

First author, year (study name; gender; age) Risk of bias RCT	Intervention (duration) Comparator (duration)	Sample size	If comparing with 1y ago, are you less short of breath/no change/more short of breath? - Less short of breath - No change - More short of breath, in n (%)	p-value	If comparing with 1y ago, is your voice less coarse/no change/ more coarse? - Less coarse - No change - More coarse, in n (%)	p-value	If comparing with 1y ago, is your skin warmer/no change/colder? - Warmer - No change - Colder, in n (%)	p-value	If comparing with 1y ago, do you feel less constipated/no change/ more constipated? - Less constipated - No change - More constipated, in n (%)	p-value	If comparing with 1y ago, do you feel better/no change/ worse? - Better - No change - Worse, in n (%)	p-value	Change of symptom score* Change of symptom score, at end of trial	p-value
Jorde et al., 2006 [48] (Tromsø; mix; >29y) High risk of bias	Levothyroxine (12 months)	36	2 (5.5%) 31 (86.1%) 3 (8.3%)	*ns*§	1 (2.8%) 30 (83.3%) 5 (13.9%)	*ns*§	3 (8.3%) 32 (88.9%) 1 (2.8%)	*ns*§	6 (16.7%) 26 (72.2%) 4 (11.1%)	*ns*§	4 (11.1%) 29 (80.6%) 3 (8.3%)	*ns*§	-0.08 (2.23)‡	*ns*§
	Placebo (12 months)	33	2 (6.1%) 25 (75.8%) 6 (18.2%)		1 (3.1%)† 25 (78.1%)† 6 (18.8%)†		3 (9.1%) 28 (84.8%) 2 (6.1%)		2 (6.1%) 30 (90.9%) 1 (3.0%)		3 (9.1%) 25 (75.8%) 5 (15.2%)		-0.65 (2.04)‡	

Keys: NR = not reported; *ns* = not significant, RCT = randomised controlled trial, y = years. * +1 point if the change was in the better (more euthyroid) direction, -1 point if the change was in the hypothyroid direction, and 0 point if there was no change; † Sample size of comparator group for this question is 32 instead of 33; ‡ Sample size of intervention group and comparator for the outcome symptom score is not reported; § Full p-value not available. Statistically significant results.

Table 21. Efficacy results on levothyroxine treatment: Fatigue

First author, year (study name; gender; age) Risk of bias RCT	Intervention (duration) Comparator (duration)	Sample size	PFS physical score* Crude mean (SD) at baseline, score Crude mean (SD) at end of trial, score	Difference in change of adjusted mean score (95% CI) between treatment arms; p-value	PFS mental score† Crude mean (SD) at baseline, score Crude mean (SD) at end of trial, score	Difference in change of adjusted mean score (95% CI) between treatment arms; p-value	PFS - participants with higher physical fatigability‡ n of patients (%) at baseline n of patients (%) at end of trial	Difference in change of adjusted mean score (95% CI) between treatment arms; p-value	PFS - participants with higher mental fatigability§ n of patients (%) at baseline n of patients (%) at end of trial	Difference in change of adjusted mean score (95% CI) between treatment arms; p-value
Stuber et al., 2020 [51] (nested study in TRUST in Ireland & Switzerland; mix; ≥65y) Low risk of bias	No stratification									
	Levothyroxine (12 months)	119	14.7 (9.3)‖	0.2 (-1.8 to 2.1); 0.88	7.4 (8.0)¶	-1.0 (-2.8 to 0.8); 0.26	55 (46.2%)	1.0 (0.5 to 1.8); 0.88	27 (22.7%)	0.6 (0.3 to 1.4); 0.23
			14.8 (9.6)		6.0 (7.8)		52 (43.7%)		19 (16.0%)	
	Placebo (12 months)	111	11.1 (9.1)‖		5.1 (6.9)¶		33 (29.7%)		14 (12.6%)	
			12.4 (9.3)		6.0 (8.0)		41 (36.9%)		20 (18.0%)	
	Stratified by TSH level: TSH ≥6.76 µU/ml (upper quartile)									
	Levothyroxine (12 months)	28	11.5 (8.5)	-0.01 (-4.2 to 4.2); 0.1	5.5 (7.7)	0.62 (-2.8 to 4.0); 0.72	-	-	-	-
			13.0 (8.5)		6.1 (8.7)		-		-	
	Placebo (12 months)	28	6.3 (6.3)		3.4 (4.1)		-		-	
			10.3 (7.5)		4.1 (5.1)		-		-	

Keys: CI = confidence interval, n = number, PFS = Pittsburgh Fatigability Scale, RCT = randomised controlled trial, SD = standard deviation, TRUST = Thyroid Hormone Replacement for Subclinical Hypothyroidism, TSH = thyroid-stimulating hormone, y = years. * 0-50 scale, higher scores indicating greater fatigability. Minimal clinically important difference is estimated at 2 to 3 points. † 0-50 scale, higher scores indicating greater fatigability. Minimal clinically important difference is estimated at 2 to 3 points. ‡ Cut points for higher fatigability is set at ≥15 points. § Cut points for higher fatigability is set at ≥13 points. ‖ p-value for PFS physical score at baseline in intervention and comparator: 0.003. ¶ p-value for PFS mental score at baseline in intervention and comparator: 0.023.

14.3.4 *Neurological outcomes*

Table 22. Efficacy results on levothyroxine treatment: Cognitive function tests

First author, year (study name; gender; age) Risk of bias RCT	Intervention (duration) Comparator (duration)	Sample size	MEAMS* Mean (SE) at baseline, score Adjusted mean (SE) at end of trial, score	Difference in change of adjusted mean score (95% CI) between treatment arms; p-value‡	MMSE† Mean (SE) at baseline, score Adjusted mean (SE) at end of trial, score	Difference in change of adjusted mean score (95% CI) between treatment arms; p-value‡
Parle et al., 2010 [26] (The Birmingham elderly thyroid study; mix; ≥65y) Moderate risk of bias	Levothyroxine (12 months)	NR	11.72 (0.13)	0.07 (-0.21 to 0.36); 0.57	28.26 (0.30)	0.03 (-1.12 to 1.17); 0.18
		46	11.67 (0.09)		28.24 (0.38)	
	Placebo (12 months)	NR	11.21 (0.16)		28.17 (0.36)	
		36	11.60 (0.11)		28.22 (0.43)	

Keys: CI = confidence interval, MEAMS = Middlesex Elderly Assessment of Mental State, MMSE = Mini-Mental State Examination, NR = not reported, RCT = randomised controlled trial, SE = standard error, y = years. * Scores range 0-30. Higher score indicates less dysfunction; † Score range 0-12. Higher score indicates less dysfunction; ‡ p-value group x time, calculated for baseline score, intermittent score (not presented in table) and end of trial score.

Table 23a. Efficacy results on levothyroxine treatment: Attention and memory tests

First author, year (study name; gender; age) Risk of bias RCT	Intervention (duration) Comparator (duration)	Sample size	Total memory quotient (Wechsler's memory test)* Mean (SD) at baseline, score Mean (SD) at end of trial, score	p-value‡	Digit span Forward (subtest from Wechsler Memory Scale)† Mean (SD) at baseline, score Mean (SD) at end of trial, score	p-value§	Digit span backward (subtest from Wechsler Memory Scale)† Mean (SD) at baseline, score Mean (SD) at end of trial, score	p-value§‡
Aghili et al., 2012 [55] (NR; mix; 18-64y) High risk of bias	Levothyroxine (3 months)	30	101.45 (7.34)	0.002	-	-	-	-
			107.03 (7.44)		-		-	
	Placebo (3 months)	30	106.4 (12.1)		-		-	
			105.4 (12.06)		-		-	
Jorde et al., 2006 [48] (Tromsø; mix; >29y) High risk of bias	Levothyroxine (12 months)	36	-	-	5.8 (1.0)	*ns***	4.1 (0.9)	*ns***
			-		5.8 (1.0)		4.2 (0.9)	
	Placebo (12 months)	33	-		5.5 (1.2)		4.2 (1.4)	
			-		5.8 (1.3)		4.1 (1.1)	

Table 23b continued. Efficacy results on levothyroxine treatment: Attention and memory tests

First author, year (study name; gender; age) Risk of bias RCT	Intervention (duration)	Sample size	Verbal recall (immediate and delayed recall; subtest from Wechsler Memory Scale)†		Visual recall (immediate and delayed recall; subtest from Wechsler Memory Scale)†		Word list test (subtest from California Verbal Learning Test)†		Seashore Rhythm test†	
	Comparator (duration)		Mean (SD) at baseline, score	p-value‡	Mean (SD) at baseline, score	p-value‡	Mean (SD) at baseline, score	p-value‡	Mean (SD) at baseline, score	p-value‡
			Mean (SD) at end of trial, score		Mean (SD) at end of trial, score		Mean (SD) at end of trial, score		Mean (SD) at end of trial, score	
Aghili et al., 2012 [55] (NR; mix; 18-64y) High risk of bias	Levothyroxine (3 months)	30	-	-	-	-	-	-	-	-
			-		-		-		-	
	Placebo (3 months)	30	-		-		-		-	
			-		-		-		-	
Jorde et al., 2006 [48] (Tromsø; mix; >29y) High risk of bias	Levothyroxine (12 months)	36	21.4 (6.1)‖	*ns***	17.7 (6.6)	*ns***	25.8 (5.7)	*ns***	424 (236)‖	*ns***
			23.3 (5.8)¶		19.7 (7.5)		26.3 (5.6)¶		372 (108)¶	
	Placebo (12 months)	33	19.8 (5.2)		16.6 (6.3)		23.7 (6.8)		399 (132)#	
			22.7 (5.5)		16.9 (7.9)		25.2 (9.5)		409 (114)	

Keys: NR = not reported; *ns* = not significant, RCT = randomised controlled trial, SD = standard deviation, y = years. * Total memory quotient of the 7 Wechsler's memory subtests. Higher scores indicating better memory; † Interpretation of results not specified; ‡ p-value change between groups; §p-value between groups at end of trial; ‖ Sample size: n=34; ¶ Sample size: n=35; # Sample size: n=32;** Full p-value not available. Statistically significant results

Table 24. Efficacy results on levothyroxine treatment: Cognitive processing tests - Part I

First author, year (study name; gender; age) Risk of bias RCT	Intervention (duration) Comparator (duration)	Sample size	CalCAP (speed of information processing)* Mean (SD) at baseline, score Mean (SD) at end of trial, score	p-value§	Letter-digit coding test† Mean (SD) at baseline, score Mean (SD) at end of trial, score	Difference in change in adjusted mean score (95% CI) between treatment arms ; p-value	Digit symbol test* Mean (SD) at baseline, score Mean (SD) at end of trial, score	p-value§	SCOLP test‡ Mean (SE) at baseline, score Adjusted mean (SE) at end of trial, score	Difference in change of adjusted mean score (95% CI) between treatment arms; p-value‖	Controlled word association test* Mean (SD) at baseline, score Mean (SD) at end of trial, score	p-value§	Vocabulary (WAIS)* Mean (SD) at baseline, score Mean (SD) at end of trial, score	p-value§
Jorde et al., 2006 [48] (Tromsø; mix; >29y) High risk of bias	Levothyroxine (12 months)	36	1573 (313)¶	*ns*††	-	-	45.7 (12.1)	*ns*††	-	-	37.9 (12.4)	*ns*††	18.9 (3.7)	*ns*††
			1483 (302)#		-		48.2 (14.4)		-		41.7 (15.0)		19.8 (3.5)	
	Placebo (12 months)	33	1633 (348)**		-		39.5 (14.2)		-		35.0 (12.1)		18.1 (5.4)	
			1562 (331)		-		40.7 (15.8)		-		41.6 (14.1)		19.7 (4.2)	
Mooijaart et al., 2019 [50] (TRUST and IEMO; mix; ≥80y) Moderate risk of bias	Levothyroxine (12 months)	85	-	-	22.0 (7.1)	1.24 (−0.30 to 2.78); 0.11	-	-	-	-	-	-	-	-
			-		21.6 (9.2)		-		-		-		-	
	Placebo (12 months)	109	-		21.9 (7.7)		-		-		-		-	
			-		20.5 (7.1)		-		-		-		-	
Stott et al., 2017 [4] (original TRUST study; mix; ≥65y) Moderate risk of bias	Levothyroxine (12 months)	358	-	-	28.0 (10.2)	-0.1 (-0.9 to 0.7); *ns*††	-	-	-	-	-	-	-	-
		302	-		27.5 (10.5)		-		-		-		-	
	Placebo (last follow-up between 12-36 months)	366	-		25.2 (8.3		-		-		-		-	
		298	-		27.1 (11.2)		-		-		-		-	
Parle et al., 2010 [26] (The Birmingham	Levothyroxine (12 months)	NR	-	-	-	-	-	-	1.91 (0.46)	0.44 (-0.46 to 1.36); 0.59	-	-	-	-
		49	-		-		-		1.29 (0.30)		-		-	
	Placebo	NR	-		-		-		0.71 (0.57)		-		-	

First author, year (study name; gender; age) Risk of bias RCT	Intervention (duration) Comparator (duration)	Sample size	CalCAP (speed of information processing)* Mean (SD) at baseline, score Mean (SD) at end of trial, score	p-value§	Letter-digit coding test† Mean (SD) at baseline, score Mean (SD) at end of trial, score	Difference in change in adjusted mean score (95% CI) between treatment arms ; p-value	Digit symbol test* Mean (SD) at baseline, score Mean (SD) at end of trial, score	p-value§	SCOLP test‡ Mean (SE) at baseline, score Adjusted mean (SE) at end of trial, score	Difference in change of adjusted mean score (95% CI) between treatment arms; p-value‖	Controlled word association test* Mean (SD) at baseline, score Mean (SD) at end of trial, score	p-value§	Vocabulary (WAIS)* Mean (SD) at baseline, score Mean (SD) at end of trial, score	p-value§
elderly thyroid study; mix; ≥65y) Moderate risk of bias	(12 months)	36	-		-		-		0.84 (0.35)		-		-	

Keys: CalCAP = California Computerised Assessment Package, CI = confidence interval, IEMO = Institute for Evidence-Based Medicine in Old Age, *ns* = not significant, RCT = randomised controlled trial, SCOLP = Speed and Capacity of Language Processing test, SE = standard error, SD = standard deviation, TRUST = Thyroid Hormone Replacement for Subclinical Hypothyroidism, WAIS = Wechsler Adult Intelligence Scale, y = years. * Interpretation of results not specified; † Higher scores indicating better executive cognitive function; ‡ Higher discrepancy score indicates greater cognitive dysfunction (≥4 indicates a severe problem); § p-value change between groups; ‖ p-value group x time, calculated for baseline score, intermittent score (not presented in table) and end of trial score; ¶ Sample size: n=34; # Sample size: n=35; ** Sample size: n=32; †† Full p-value not available.

Table 25. Efficacy results on levothyroxine treatment: Cognitive processing tests - Part II

First author, year (study name; gender; age) Risk of bias RCT	Intervention (duration) Comparator (duration)	Sample size	Trail Making B-A* Mean (SE) at baseline, score Adjusted mean (SE) at end of trial, score	Difference in change of adjusted mean score (95% CI) between treatment arms; p-value	Trail making A* Mean (SE) at baseline, score Adjusted mean (SE) at end of trial, score	Mean (SD) at baseline, score Mean (SD) at end of trial, score	Difference in change of adjusted mean score (95% CI) between treatment arms; p-value	Trail making B* Mean (SE) at baseline, score Adjusted mean (SE) at end of trial, score	Mean (SD) at baseline, score Mean (SD) at end of trial, score	Difference in change of adjusted mean score (95% CI) between treatment arms; p-value	Stroop Color-Word test, part 1 and 2† Mean (SD) at baseline, score Mean (SD) at end of trial, score	p-value	Stroop Color-Word test, part 3† Mean (SD) at baseline, score Mean (SD) at end of trial, score	p-value
Parle et al., 2010 [26] (The Birmingham elderly thyroid study; mix; ≥65y) Moderate risk of bias	Levothyroxine (12 months)	NR	65.76 (14.81)	-12.72 (-33.50 to 8.06); 0.86‡	45.72 (2.32)	-	-1.44 (-9.42 to 6.53); 0.52‡	110.57 (15.89)	-	-13.46 (-37.15 to 10.22); 0.95‡	-	-	-	-
		48	54.55 (6.80)		45.33 (2.63)	-		100.65 (7.75)	-		-		-	
	Placebo (12 months)	NR	82.63 (18.38)		50.29 (2.81)	-		131.46 (19.72)	-		-		-	
		34	67.27 (7.97)		46.78 (3.05)§	-		114.11 (9.07)	-		-		-	
Jorde et al., 2006 [48] (Tromsø; mix; >29y) High risk of bias	Levothyroxine (12 months)	36	-	-	-	45.5 (17.0)	NR; *ns*‖	-	92 (44)	NR; *ns*‖	48.0 (9.0)**	*ns*‖	*63.1 (21.0)*	*ns*‖
			-		-	39.0 (14.8)		-	94 (62)		47.1 (8.9)		61.1 (27.3)	
	Placebo (12 months)	33	-		-	50.0 (19.7)		-	111 (48)¶		56.2 (27.7)††		74.3 (54.4)††	
			-		-	44.1 (17.7)		-	103 (49)#		50.7 (13.2)††		66.7 (28.2)††	

Keys: CI = confidence interval, SD = standard deviation, SE = standard error, NR; = not reported, *ns* = not significant, RCT = randomised controlled trial, y = years. * Slower times to complete the test indicate greater cognitive dysfunction; † Interpretation of results not specified; ‡ p-value group x time, calculated for baseline score, intermittent score (not presented in table) and end of trial score; § Sample size: n=36; ‖ p-value change between groups, full p-value not available; ¶ Sample size: n=31; # Sample size: n=30; ** Sample size: n=35; †† Sample size: n=31.

Table 26. Efficacy results on levothyroxine treatment: Anxiety

First author, year (study name; gender; age) Risk of bias RCT	Intervention (duration) Comparator (duration)	Sample size	HADS: Anxiety*				Hamilton scale: Anxiety†	
			Change in case status, n (%)			p-value difference in the change in case status between treatment arms	Mean variation (SD) at end of trial, score	p-value
			Improved: from case to non-case	Worse: from non-case to case	Unchanged			
Kong et al., 2002 [61] (NR; women; ≥18y) Moderate risk of bias	Levothyroxine (6 months)	20	2 (10%)	3 (15%)	15 (75%)	0.5	-	-
	Placebo (6 months)	14	3 (21%)	1 (7%)	10 (72%)		-	
Reuters et al., 2012 [71] (NR; mix; NA‡) High risk of bias	Levothyroxine (6 months)	25	-	-	-	-	-3.6 (5.1)	0.83
	Placebo (6 months)	32	-	-	-		-4.1 (7.6)	

Keys: GHQ = general health questionnaire, HADS = Hospital Anxiety and Depression Scale, n = number, NA = not applicable, NR = not reported, SD = standard deviation, RCT = randomised controlled trial, y = years. * Scores of 8 on the Hospital Anxiety and Depression Scale used to identify clinical case status. † Negative score indicates an improvement, i.e. a reduction in the level of anxiety symptoms. ‡ Age was not an inclusion or exclusion criterion; the mean (SD) age at the analysis was 51.7 (9.2) years in the intervention group and 48.3 (11.7) years in the placebo group.

Table 27a. Efficacy results on levothyroxine treatment: Depression

First author, year (study name; gender; age) Risk of bias RCT	Intervention (duration) Comparator (duration)	Sample size	HADS: depression* Mean (SE) at baseline, score Adjusted mean (SE) at end of trial, score	Difference in change of adjusted mean score (95% CI) between treatment arms; p-value	Change in status, n (%) Improved: from case to non-case	Worse: from non-case to case	Unchanged	p-value difference in the change in case status between treatment arms	BDI total score (1-21): depressed mood† Mean (SD) at baseline, score Mean (SD) at end of trial, score	Mean variation (SD) at end of trial, score	p-value
Parle et al., 2010 [26] (The Birmingham elderly thyroid study; mix; ≥65y) Moderate risk of bias	Levothyroxine (12 months)	52	3.38 (0.37)	0.18 (-0.64 to 1.00); 0.82	-	-	-	-	-	-	-
		49	3.55 (0.27)						-		
	Placebo (12 months)	42	2.88 (0.45)		-	-	-		-	-	
		36	3.37 (0.31)						-		
Kong et al., 2002 [61] (NR; women; ≥18y) Moderate risk of bias	Levothyroxine (6 months)	20	-	-	3 (15%)	1 (5%)	16 (80%)	0.8	-	-	-
			-						-		
	Placebo (6 months)	14	-		1 (7%)	1 (7%)	12 (86%)		-	-	
			-						-		
Reuters et al., 2012 [71] (NR; mix; NA[#]) High risk of bias	Levothyroxine (6 months)	25	-	-	-	-	-	-	-	-2.4 (5.8)	0.834
			-						-		
	Placebo (6 months)	32	-		-	-	-		-	-2.1 (4.8)	
			-						-		
Jorde et al., 2006 [48] (Tromsø; mix; >29y) High risk of bias	Levothyroxine (12 months)	36	-	-	-	-	-	-	4.4 (3.7)	-	*ns***
			-						4.3 (3.6)		
	Placebo (12 months)	33	-		-	-	-		3.7 (3.8)	-	
			-						3.3 (4.0)		
		211	-	-	-	-	-	-	-	-	-

First author, year (study name; gender; age) Risk of bias RCT	Intervention (duration) Comparator (duration)	Sample size	HADS: depression* Mean (SE) at baseline, score Adjusted mean (SE) at end of trial, score	Difference in change of adjusted mean score (95% CI) between treatment arms; p-value	Change in status, n (%) Improved: from case to non-case	Worse: from non-case to case	Unchanged	p-value difference in the change in case status between treatment arms	BDI total score (1-21): depressed mood† Mean (SD) at baseline, score Mean (SD) at end of trial, score	Mean variation (SD) at end of trial, score	p-value
Wildisen et al., 2021[53] (nested study in TRUST in the Netherlands, Switzerland, Ireland included incidence/recovery analyses; mix; ≥65y) Low risk of bias	Levothyroxine (12 months)		-						-		
	Placebo (12 months)	216	- -	-	-	-	-		- -	-	

Table 27b continued. Efficacy results on levothyroxine treatment: Depression

First author, year (study name; gender; age) Risk of bias RCT	Intervention (duration) Comparator (duration)	Sample size	Hamilton scale: depression‡ Mean variation (SD) at end of trial, score	p-value	GDS-15 score§ Mean (SD) at baseline, score Mean (SD) at end of trial, score	Adjusted mean difference (95% CI) between treatment arms††; p-value	GDS-15 and CESD-20: Incidence of mild depression‖ n/N, %	Adjusted OR (95% CI); p-value	GDS-15 and CESD-20: Recovery of mild depression¶ n/N, %	Adjusted OR (95% CI); p-value
Parle et al., 2010 [26] (The Birmingham elderly thyroid study; mix; ≥65y) Moderate risk of bias	Levothyroxine (12 months)	52	-	-	-	-	-	-	-	-
		49	-							
	Placebo (12 months)	42	-		-		-		-	
		36	-							
Kong et al., 2002 [61] (NR; women; ≥18y) Moderate risk of bias	Levothyroxine (6 months)	20	-	-	-	-	-	-	-	-
	Placebo (6 months)	14	-		-		-		-	
Reuters et al., 2012 [71] (NR; mix; NA#) High risk of bias	Levothyroxine (6 months)	25	-1.6 (2.8)	0.25	-	-	-	-	-	-
	Placebo (6 months)	32	-0.6 (2.9)		-		-		-	
Jorde et al., 2006 [48] (Tromsø; mix; >29y) High risk of bias	Levothyroxine (12 months)	36	-	-	-	-	-	-	-	-
	Placebo (12 months)	33	-		-		-			
Wildisen et al., 2021[53] (nested study in TRUST in the Netherlands, Switzerland, Ireland included incidence/recovery analyses; mix; ≥65y) Low risk of bias	Levothyroxine (12 months)	211	-	-	1.26 (1.85)	0.15 (-0.15 to 0.46); 0.33	12/235 = 5.1%	1.06 (0.45 to 2.49); 0.89	9/23 = 39.1%	0.20 (0.02 to 1.93); 0.16
			-		1.39 (2.13)					
	Placebo (12 months)	216	-		0.96 (1.58)		12/246 = 4.9%		6/10 = 60.0%	
			-		1.07 (1.67)					

Keys: BDI = Beck Depression Inventory, CESD-20 = Center for Epidemiologic Studies Depression 20-item scale, CI = confidence interval, GDS = Geriatric Depression Scale, HADS = Hospital Anxiety and Depression Scale, NA = not applicable, ns = not significant, OR = odds ratio, RCT = randomised controlled trial, SD = standard deviation, SE = standard error, TRUST = Thyroid Hormone Replacement for Subclinical Hypothyroidism, y = years. * Parle et al., 2010: Score range from 0-21 with a score of 11 or above indicating caseness (diagnosis of depression). Kong et al., 2002: Scores of 8 on the Hospital

Anxiety and Depression Scale used to identify clinical case status. † Interpretation of results not specified. ‡ Negative score indicates an improvement, i.e. a reduction in the level of depression symptoms. § Higher score indicates more severe depressive symptoms. Positive mean difference in GDS-15 score indicate benefit of placebo. ‖Incidence depression: GDS-15 >3, CESD-20 >20. Inclusion of participants no mild depression at baseline (GDS-15 ≤3, CESD-20 ≤20). OR <1 indicate benefit of levothyroxine therapy. ¶ Recovery: GDS-15 ≤3, CESD-20 ≤20. Inclusion of participants with mild depression at baseline (GDS-15 > 3, CESD-20 > 20). OR >1 indicates benefit for levothyroxine. # Age was not an inclusion or exclusion criterion; the mean (SD) age at the analysis was 51.7 (9.2) years in the intervention group and 48.3 (11.7) years in the placebo group; ** Full p-value not available; †† Not further specified.

Table 28. Efficacy results on levothyroxine treatment: Stratified results of depression

First author, year (study name; gender; age) Risk of bias RCT	Stratification	Intervention (duration) Comparator (duration)	Sample size	GDS-15 score* Adjusted mean difference (95% CI) between treatment arms†	 p-value between stratified groups‡
Wildisen et al., 2021 [53] (nested study in TRUST in the Netherlands and Switzerland; mix; ≥65y) Low risk of bias	Stratified by gender				
	Women	Levothyroxine (12 months)	118	0.03 (-0.42 to 0.48)	0.29
		Placebo (12 months)	121		
	Men	Levothyroxine (12 months)	93	0.27 (-0.13 to 0.66)	
		Placebo (12 months)	95		
	Stratified by age				
	65-75 years	Levothyroxine (12 months)	128	0.02 (-0.41 to 0.45)	0.36
		Placebo (12 months)	116		
	>75 years	Levothyroxine (12 months)	83	0.23 (-0.21 to 0.67)	
		Placebo (12 months)	100		
	Stratified by TSH level				
	TSH 4.5 to <7.0 µU/ml	Levothyroxine (12 months)	157	0.21 (-0.18 to 0.61)	0.5
		Placebo (12 months)	153		
	TSH 7.0 to <10 µU/ml	Levothyroxine (12 months)	39	-0.02 (-0.55 to 0.51)	
		Placebo (12 months)	50		
	TSH 10.0 to 10.0 µU/ml	Levothyroxine (12 months)	15	-0.23 (-1.13 to 0.67)	
		Placebo (12 months)	13		

Keys: CI = confidence interval, GDS = Geriatric Depression Scale, RCT = randomised controlled trial, TRUST = Thyroid Hormone Replacement for Subclinical Hypothyroidism, TSH = thyroid stimulating hormone, y = years.* Higher scores indicating higher likelihood of depression. Positive mean difference in GDS-15 score indicate benefit of placebo † Not further specified; ‡ p-value applies to adjusted between group difference.

Table 29. Efficacy results on levothyroxine treatment: GHQ with GHQ scoring method

First author, year (study name; gender; age) Risk of bias RCT	Intervention (duration) Comparator (duration)	Sample size	GHQ-30* Mean (SD) at baseline, score Mean (SD) at end of trial, score	p-value	Change in status, n (%) Improved: from case to non-case	Worse: from non-case to case	Unchanged	p-value difference in the change in case status between treatment arms
Kong et al., 2002 [61] (NR; women; ≥18y) Moderate risk of bias	Levothyroxine (6 months)	20	-	-	5 (25%)	3 (15%)	12 (60%)	0.2
			-					
	Placebo (6 months)	12	-		7 (58%)	1 (8%)	4 (33%)	
			-					
Jorde et al., 2006 [48] (Tromsø; mix; >29y) High risk of bias	Levothyroxine (12 months)	36	1.5 (2.3)	*ns*†	-	-	-	-
		36	1.9 (3.3)					
	Placebo (12 months)	33	0.7 (1.3)		-	-	-	
		32	1.2 (2.0)					

Keys: CI = confidence interval, GHQ = general health questionnaire, n = number of cases, *ns* = not significant, NR = not reported, RCT = randomised controlled trial, SD = standard deviation, y = years. * Case status an indicator general psychological disorder, was defined as a General Health Questionnaire score of 4 or more; † Full p-value not available.

Table 30. Efficacy results on levothyroxine treatment: GHQ with Likert-scoring method

First author, year (study name; gender; age) Risk of bias RCT	Intervention (duration) Comparator (duration)	Sample size	GHQ-30 with Likert-scoring method* Mean (SD) at baseline, score Mean (SD) at end of trial, score	p-value	GHQ-30 with Likert-scoring method Factor A (anxiety)* Mean (SD) at baseline, score Mean (SD) at end of trial, score	p-value	GHQ-30 with Likert-scoring method Factor B (feeling of incompetence)* Mean (SD) at baseline, score Mean (SD) at end of trial, score	p-value	GHQ-30 with Likert-scoring method Factor C (depression, hopelessness)* Mean (SD) at baseline, score Mean (SD) at end of trial, score	p-value	GHQ-30 with Likert-scoring method Factor D (difficulty in coping)* Mean (SD) at baseline, score Mean (SD) at end of trial, score	p-value	GHQ-30 with Likert-scoring method Factor E (social dysfunction)* Mean (SD) at baseline, score Mean (SD) at end of trial, score	p-value
Jorde et al., 2006 [48] (Tromsø; mix; >29y) High risk of bias	Levothyroxine (12 months)	36	22.6 (5.5)	*ns*†	4.8 (2.9)	*ns*†	4.0 (0.7)	*ns*†	1.8 (1.2)	*ns*†	3.8 (1.2)	*ns*†	3.1 (1.0)	*ns*†
		36	23.9 (7.2)		5.7 (3.1)		4.0 (0.9)		2.1 (1.6)		4.0 (1.4)		3.1 (1.2)	
	Placebo (12 months)	33	21.6 (5.2)		4.6 (2.5)		3.9 (0.7)		2.1 (1.6)		3.6 (0.8)		2.8 (1.1)	
		32	23.5 (6.1)		5.5 (3.0)		4.0 (0.6)		2.2 (1.6)		4.0 (1.2)		3.0 (0.5)	

Keys: GHQ = general health questionnaire, NR = not reported, *ns* = not significant, RCT = randomised controlled trial, SD = standard deviation, y = years. * Interpretation of results not specified; † Full p-value not available.

14.3.5 *Musculoskeletal outcomes*

Table 31. Efficacy results on levothyroxine treatment: Fractures and osteoporosis

First author, year (study name; gender; age) Risk of bias RCT	Intervention (duration) Comparator (duration)	Sample size	Fracture			New diagnosis of osteoporosis	
			n (%) at end of trial	Adjusted HR (95% CI)	Estimated risk difference (95% CI)*	n (%) at end of trial	Estimated risk difference (95% CI)
Stott et al., 2017 [4] (original TRUST study; mix; ≥65y) Moderate risk of bias	Levothyroxine (12 months)	368	9 (2.4%)	1.06 (0.41 to 2.76)	-	3 (0.8%)	NR†
	Placebo (12 months)	369	8 (2.2%)			4 (1.1%)	
Mooijaart et al., 2019 [50] (TRUST and IEMO; mix; ≥80y) Moderate risk of bias	Levothyroxine (last visit, >12 months)	112	4 (3.6%)	-	0.00 (-0.04 to 0.03)	-	-
	Placebo (last visit, >12 months)	139	5 (3.6%)			-	

Keys: CI = confidence interval, HR = hazard ratio, IEMO = Institute for Evidence-Based Medicine in Old Age, n = number, NR = not reported, RCT = randomised controlled trial, TRUST = Thyroid Hormone Replacement for Subclinical Hypothyroidism, y = years. * Estimated risk difference were estimated using Cox proportional hazard regression models, presented as risk differences and 95%CIs, obtained through bootstrap resampling in 1000 iterations; † Hazard ratios were not calculated for new diagnosis of osteoporosis owing to the small number of events; however, the authors reported in the text that the incidence of new diagnosis of osteoporosis (i.e. a serious adverse event of special interest) was similar in the two groups.

14.3.6 *Cardiovascular outcomes*

Table 32. Efficacy results on levothyroxine treatment: atrial fibrillation and heart failure

First author, year (study name; gender; age) Risk of bias RCT	Intervention (duration) Comparator (duration)	Sample size	New onset atrial fibrillation			Heart failure		
			n (%)	Estimated risk difference (95% CI)*	Adjusted HR (95% CI)	n (%)	Estimated risk difference (95% CI)*	Adjusted HR (95% CI)
Stott et al., 2017 [4] (original TRUST study; mix; ≥65y) Moderate risk of bias	Levothyroxine (12 months)	368	11 (3.0%)	NR	0.80 (0.35 to 1.80)	3 (0.8%)	NR	NR†
	Placebo (12 months)	369	13 (3.5%)			6 (1.6%)		
Mooijaart et al., 2019 [50] (TRUST and IEMO; mix; ≥80y) Moderate risk of bias	Levothyroxine (last visit, >12 months)	112	4 (3.6%)	0.00 (-0.02 to 0.03)	NR	3 (2.7%)	0.01 (-0.03 to 0.05)	NR
	Placebo (last visit, >12 months)	139	6 (4.3%)			6 (4.3%)		
Zijlstra et al., 2021 [52] (TRUST & IEMO; mix; ≥65y or ≥80y) Low risk of bias	Levothyroxine (extended follow-up, >12 months)	NR	11 (2.6%)	NR	0.69 (0.32 to 1.52)	4 (1.0%)	NR	0.41 (0.13 to 1.35)
	Placebo (extended	NR	15 (3.6%)			9 (2.1%)		

First author, year (study name; gender; age) Risk of bias RCT	Intervention (duration) Comparator (duration)	Sample size	New onset atrial fibrillation			Heart failure		
			n (%)	Estimated risk difference (95% CI)*	Adjusted HR (95% CI)	n (%)	Estimated risk difference (95% CI)*	Adjusted HR (95% CI)
	follow-up, >12 months)							

Keys: CI = confidence interval, HR = hazard ratio, IEMO = Institute for Evidence-Based Medicine in Old Age, n = number, RCT = randomised controlled trial, TRUST = Thyroid Hormone Replacement for Subclinical Hypothyroidism, y = years. * Estimated risk difference were estimated using Cox proportional hazard regression models, presented as risk differences and 95%CIs, obtained through bootstrap resampling in 1000 iterations; † Hazard ratios were not calculated for heart failure owing to the small number of events. However, the authors reported in the text that the incidence of heart failure (i.e. a serious adverse event of special interest) was similar in the two groups.

14.3.7 *Activities of daily living and HRQoL*

Table 33. Efficacy results on levothyroxine treatment: activities of daily living

First author, year (study name; gender; age) Risk of bias RCT	Intervention (duration) Comparator (duration)	Sample size	Barthel index*		Instrumental activities of daily living†	
			Mean (SD), score at baseline Mean (SD), score at end of trial	Difference in change in adjusted mean score (95% CI) between treatment arms; p-value	Mean (SD), score at baseline Mean (SD), score at end of trial	Difference in change in adjusted mean score (95% CI) between treatment arms; p-value
Mooijaart et al., 2019 [50] (TRUST and IEMO; mix; ≥80y) Moderate risk of bias	Levothyroxine (last visit, >12 months)	90	19.3 (1.5)‡	0.09 (-0.33 to 0.52); 0.66	13.3 (1.3)‡	-0.40 (-0.92 to 0.13); 0.14
		93	19.0 (1.9)‡		12.4 (2.4)‡	
	Placebo (last visit, >12 months)	122	19.4 (1.2)§		13.2 (1.5)§	
		124	19.1 (2.1)§		12.7 (2.3)§	

Keys: CI = confidence interval, IEMO = Institute for Evidence-Based Medicine in Old Age, RCT = randomised controlled trial, SD = standard deviation; TRUST = Thyroid Hormone Replacement for Subclinical Hypothyroidism, y = years. * Range 0-20. Score of 20 indicates fully independent in activities of daily living and mobility, score range: 15-19 moderately to fully independent, score range: 10-14 needing help but capable of own activities, score range: 5-9 severely dependent, score range: 0-4 totally dependent; † Range 0-14, higher scores indicating better performance in instrumental activities of daily living; ‡ Sample size: n=91; § Sample size: n=117.

Table 34. Efficacy results on levothyroxine treatment: HRQoL – Part I

First author, year (study name; gender; age) Risk of bias RCT	Intervention (duration) Comparator (duration)	Sample size	EQ-5D descriptive index*		EQ visual-analogue scale score†	
			Mean (SD), score at baseline Mean (SD), score at end of trial	Difference in change in adjusted mean score (95% CI) between treatment arms; p-value	Mean (SD), score at baseline Mean (SD), score at end of trial	Difference in change in adjusted mean score (95% CI) between treatment arms; p-value
Stott et al., 2017 [4] (original TRUST study; mix; ≥65y)	Levothyroxine (12 months)	368	0.846 (0.187)	-0.025 (-0.050 to 0.000); 0.05	78.4 (15.3)	-1.3 (-3.2 to 0.6); 0.18
		318	0.833 (0.212)		77.3 (15.6)	
	Placebo	369	0.847 (0.171)		76.5 (16.3)	

First author, year (study name; gender; age) Risk of bias RCT	Intervention (duration) Comparator (duration)	Sample size	EQ-5D descriptive index*		EQ visual-analogue scale score†	
			Mean (SD), score at baseline Mean (SD), score at end of trial	Difference in change in adjusted mean score (95% CI) between treatment arms; p-value	Mean (SD), score at baseline Mean (SD), score at end of trial	Difference in change in adjusted mean score (95% CI) between treatment arms; p-value
Moderate risk of bias	(12 months)	320	0.853 (0.191)		77.4 (13.7)	
Mooijaart et al., 2019 [50] (TRUST and IEMO; mix; ≥80y) Moderate risk of bias	Levothyroxine (last visit, >12 months)	90	0.785 (0.199)	-0.024 (-0.097 to 0.049); 0.52	75.27 (14.59)	-1.60 (-6.16 to 2.96); 0.49
		93	0.763 (0.228)		74.10 (11.97)	
	Placebo (last visit, >12 months)	122	0.811 (0.210)		73.98 (14.26)	
		124	0.806 (0.213)		74.28 (14.32)‖	

Keys: CI = confidence interval, EQ-5D = EuroQol-5 Dimension, IEMO = Institute for Evidence-Based Medicine in Old Age, RCT = randomised controlled trial, SD = standard deviation; TRUST = Thyroid Hormone Replacement for Subclinical Hypothyroidism, y = years. * Range 0.50-1.00, higher scores indicate a better quality of life; † Range 0-100, higher scores indicate better health; ‡ Range 0-20. Score of 20 indicates fully independent in activities of daily living and mobility, score range: 15-19 moderately to fully independent, score range: 10-14 needing help but capable of own activities, score range: 5-9 severely dependent, score range: 0-4 totally dependent; § Range 0-14, higher scores indicating better performance in instrumental activities of daily living; ‖ Sample size: n=123; ¶ Sample size: n=91; # Sample size: n=117.

Table 35. Efficacy results on levothyroxine treatment: HRQoL Part II

First author, year (study name; gender; age) Risk of bias RCT	Intervention (duration) Comparator (duration)	Sample size	SF-36: General health*		SF-36: Physical function*		SF-36: Role physical*		SF-36: Social function*		SF-36: Role-emotional*		SF-36: Mental health*		SF-36: Vitality*		SF-36: Role pain*	
			Mean variation (SD) at end of trial, score	p-value	Mean variation (SD) at end of trial, score	p-value	Mean variation (SD) at end of trial, score	p-value	Mean variation (SD) at end of trial, score	p-value	Mean variation (SD) at end of trial, score	p-value	Mean variation (SD) at end of trial, score	p-value	Mean variation (SD) at end of trial, score	p-value	Mean variation (SD) at end of trial, score	p-value
Reuters et al., 2012 [71] (NR; mix; NA†) High risk of bias	Levothyroxine (6 months)	25	6.5 (13.2)	0.863	3.7 (17.2)	0.799	22.1 (47.5)	0.023	1.3 (24.4)	0.939	27.7 (47.5)	0.07	0.2 (28.2)	0.476	-2.3 (22.9)	0.719	19.7 (15.2)	0.046
	Placebo (6 months)	32	7.4 (18.3)		1.9 (21.6)		-8.0 (35.1)		0.3 (33.1)		2.6 (40.8)		5.6 (22.1)		0.2 (21.1)		-4.6 (16.9)	

Keys: CI = confidence interval, SF-36 = 36-item short form survey, NA = not applicable, NR = not reported, RCT = randomised controlled trial, SD = standard deviation, y = years. * Range: 0-100 points. Higher score indicates better satisfaction with QoL; † Age was not an inclusion or exclusion criterion; the mean (SD) age at the analysis was 51.7 (9.2) years in the intervention group and 48.3 (11.7) years in the placebo group.
Statistically significant results

14.4 Search strategy cost-effectiveness analysis

Table 36. 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 37. 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 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 "hor-mone 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 38. Search strategy for cost-effectiveness systematic literature search NHS EED & CEA Registry

Database	NHS EED	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

Table 39. Drug acquisition costs for levothyroxine in Swiss Franc (CHF)

Brand	Dosage (ug)	Units	Price (pack)	Price (unit)	Price DDD (50 ug)
Eltroxin LF ®	50.00	100.00	9.35 CHF	0.09 CHF	0.09 CHF
Eltroxin LF ®	100.00	100.00	14.6 CHF	0.15 CHF	0.07 CHF
Euthyrox ®	100.00	100.00	14.7 CHF	0.15 CHF	0.07 CHF
Euthyrox ®	125.00	100.00	14.7 CHF	0.15 CHF	0.06 CHF
Euthyrox ®	150.00	100.00	14.7 CHF	0.15 CHF	0.05 CHF
Euthyrox ®	175.00	100.00	14.7 CHF	0.15 CHF	0.04 CHF
Euthyrox ®	200.00	100.00	14.7 CHF	0.15 CHF	0.04 CHF
Euthyrox ®	25.00	100.00	14.7 CHF	0.15 CHF	0.29 CHF
Euthyrox ®	50.00	100.00	14.7 CHF	0.15 CHF	0.15 CHF
Euthyrox ®	75.00	100.00	14.7 CHF	0.15 CHF	0.10 CHF
Novothyral ® Tabl	100.00	100.00	18.95 CHF	0.19 CHF	0.09 CHF
Tirosint ® Solution, Lös, Amp 1ml	100.00	30.00	8.8 CHF	0.29 CHF	0.15 CHF
Tirosint ® Solution, Lös, Amp 1ml	100.00	90.00	20.65 CHF	0.23 CHF	0.11 CHF
Tirosint ® Solution, Lös, Amp 1ml	112.00	30.00	8.8 CHF	0.29 CHF	0.13 CHF
Tirosint ® Solution, Lös, Amp 1ml	112.00	90.00	20.65 CHF	0.23 CHF	0.10 CHF
Tirosint ® Solution, Lös, Amp 1ml	125.00	30.00	8.8 CHF	0.29 CHF	0.12 CHF
Tirosint ® Solution, Lös Amp 1ml	125.00	90.00	20.65 CHF	0.23 CHF	0.09 CHF
Tirosint ® Solution, Lös, Amp 1ml	13.00	30.00	6.45 CHF	0.21 CHF	0.83 CHF
Tirosint ® Solution, Lös, Amp 1ml	13.00	90.00	14.4 CHF	0.16 CHF	0.62 CHF

Tirosint ® Solution, Lös, Amp 1ml	137.00	30.00	8.8 CHF	0.29 CHF	0.11 CHF
Tirosint ® Solution, Lös, Amp 1ml	137.00	90.00	20.65 CHF	0.23 CHF	0.08 CHF
Tirosint ® Solution, Lös, Amp 1ml	150.00	30.00	8.8 CHF	0.29 CHF	0.10 CHF
Tirosint ® Solution, Lös, Amp 1ml	150.00	90.00	20.65 CHF	0.23 CHF	0.08 CHF
Tirosint ® Solution, Lös, Amp 1ml	175.00	30.00	8.8 CHF	0.29 CHF	0.08 CHF
Tirosint ® Solution, Lös, Amp 1ml	175.00	90.00	20.65 CHF	0.23 CHF	0.07 CHF
Tirosint ® Solution, Lös, Amp 1ml	200.00	30.00	8.8 CHF	0.29 CHF	0.07 CHF
Tirosint ® Solution, Lös, Amp 1ml	200.00	90.00	20.65 CHF	0.23 CHF	0.06 CHF
Tirosint ® Solution, Lös, Amp 1ml	25.00	30.00	8.8 CHF	0.29 CHF	0.59 CHF
Tirosint ® Solution, Lös, Amp 1ml	25.00	90.00	20.65 CHF	0.23 CHF	0.46 CHF
Tirosint ® Solution, Lös, Amp 1ml	50.00	30.00	8.8 CHF	0.29 CHF	0.29 CHF
Tirosint ® Solution, Lös, Amp 1ml	50.00	90.00	20.65 CHF	0.23 CHF	0.23 CHF
Tirosint ® Solution, Lös, Amp 1ml	75.00	30.00	8.8 CHF	0.29 CHF	0.20 CHF
Tirosint ® Solution, Lös, Amp 1ml	75.00	90.00	20.65 CHF	0.23 CHF	0.15 CHF
Tirosint ® Solution, Lös, Amp 1ml	88.00	30.00	8.8 CHF	0.29 CHF	0.17 CHF
Tirosint ® Solution, Lös, Amp 1ml	88.00	90.00	20.65 CHF	0.23 CHF	0.13 CHF

Tirosint, ® Kaps	100.00	100.00	26.1 CHF	0.26 CHF	0.13 CHF
Tirosint,® Kaps	100.00	50.00	16.05 CHF	0.32 CHF	0.16 CHF
Tirosint, ® Kaps	112.00	100.00	26.1 CHF	0.26 CHF	0.12 CHF
Tirosint, ® Kaps	112.00	50.00	16.05 CHF	0.32 CHF	0.14 CHF
Tirosint,® Kaps	125.00	100.00	26.1 CHF	0.26 CHF	0.10 CHF
Tirosint, ® Kaps	125.00	50.00	16.05 CHF	0.32 CHF	0.13 CHF
Tirosint, ® Kaps	13.00	100.00	15.1 CHF	0.15 CHF	0.58 CHF
Tirosint, ® Kaps	13.00	50.00	8 CHF	0.16 CHF	0.62 CHF
Tirosint, ® Kaps	137.00	100.00	26.1 CHF	0.26 CHF	0.10 CHF
Tirosint, ® Kaps	137.00	50.00	16.05 CHF	0.32 CHF	0.12 CHF
Tirosint, ® Kaps	150.00	100.00	26.1 CHF	0.26 CHF	0.09 CHF
Tirosint, ® Kaps	150.00	50.00	16.05 CHF	0.32 CHF	0.11 CHF
Tirosint, ® Kaps	175.00	100.00	26.1 CHF	0.26 CHF	0.07 CHF
Tirosint, ® Kaps	175.00	50.00	16.05 CHF	0.32 CHF	0.09 CHF
Tirosint ® Kaps	200.00	100.00	26.1 CHF	0.26 CHF	0.07 CHF
Tirosint ® Kaps	200.00	50.00	16.05 CHF	0.32 CHF	0.08 CHF
Tirosint ® Kaps	25.00	100.00	26.1 CHF	0.26 CHF	0.52 CHF
Tirosint ® Kaps	25.00	50.00	16.05 CHF	0.32 CHF	0.64 CHF
Tirosint ® Kaps	50.00	100.00	26.1 CHF	0.26 CHF	0.26 CHF
Tirosint ® Kaps	50.00	50.00	16.05 CHF	0.32 CHF	0.32 CHF
Tirosint ® Kaps	75.00	100.00	26.1 CHF	0.26 CHF	0.17 CHF
Tirosint ® Kaps	75.00	50.00	16.05 CHF	0.32 CHF	0.21 CHF
Tirosint ® Kaps	88.00	100.00	26.1 CHF	0.26 CHF	0.15 CHF
Tirosint ® Kaps	88.00	50.00	16.05 CHF	0.32 CHF	0.18 CHF

Keys: CHF: Swiss Franc, DDD: defined daily dose