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	Thyroid function tests for the diagnosis of suspected primary or secondary thyroid dysfunction
Author/Affiliation	**Remziye Zaim, Ingelin Kvamme, Matthijs Versteegh** Institute for Medical Technology Assessment, Erasmus University, Rotterdam, the Netherlands **Anouk Oordt, Femke van Kessel** Pallas Health Research and Consultancy, Rotterdam, the Netherlands
Technology	(free) Triiodothyronine ((f)T3), (free) Thyroxine ((f)T4), and Thyroid Stimulating Hormone (TSH) testing
Type of Technology	Thyroid function tests
Date	31 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: Thyroid function tests are used for the diagnosis and monitoring of patients with thyroid disorders. The most common thyroid function tests include thyroid stimulating hormone (TSH) tests, total and/or free thyroxine (T4/fT4) tests, and total, free and/or reverse triiodothyroxine (T3, fT3, rT3) tests. Depending on the clinical situation, the choice of test(s) can be modified. For the diagnosis of suspected primary or secondary thyroid dysfunction two typical test approaches include a one-step test approach (i.e. measurement of TSH and (f)T4/(f)T3 simultaneously) and a two-step test approach (i.e. testing TSH first, only followed by the measurement of (f)T4/(f)T3 when TSH is out of the reference range). For suspected primary hypothyroidism TSH is complemented with (f)T4 and for suspected primary hyperthyroidism TSH is complemented with (f)T4 and (f)T3.

OBJECTIVE: Although most guidelines recommend a two-step test approach for the diagnosis of suspected primary or secondary thyroid dysfunction, it is observed that TSH and (f)T4/(f)T3 are often measured simultaneously (one-step test approach). The focus of this HTA is to evaluate the clinical and financial consequences of the two-step as compared to the one-step thyroid function test approach for the diagnosis of patients with suspected primary or secondary thyroid dysfunction. The difference in the number of missed cases, the diagnostic subcategory distribution of missed cases, and the costs and budget impact of thyroid function tests performed with the one-step versus the two-test test approach are addressed. Ethical, legal, social, and organisational issues are also addressed.

METHODS: For the clinical review, a systematic literature search was performed in PubMed (MEDLINE) and Embase.com. The search was worldwide, in English, German, French, and Italian with the publication period ranging from 1990 to September 2021. Studies in adults tested for suspected primary or secondary thyroid dysfunction were included. Studies on thyroid disorder screening and monitoring, and those including only pregnant females were excluded. The included studies were critically appraised, and the extracted data were summarised in evidence tables and narrative text. For the economic review, the costs and budget impact literature search followed the principles of the clinical systematic literature search. After applying the quality control measures, data synthesis was done using descriptive comparisons. A budget impact model was developed using Swiss inputs, publicly available sources, and expert opinion. Evidence found with the clinical and costs and budget impact systematic literature searches was used to inform the ethical, legal, social, and organisational

domains of the report. In addition, grey literature searches were conducted on these HTA domains.

RESULTS: For the clinical systematic literature search 9'474 unique records were identified, 53 studies were screened in full-text, and 3 studies in populations tested for suspected thyroid disorders were included for full analysis. The included studies were case series conducted in Asia, all with a high risk of bias. One large case series analysed 2'768 blood samples with both TSH and fT4 assayed in inpatients and outpatients with clinically suspected primary or secondary thyroid dysfunction from a general hospital. In 5.3% of the thyroid function tests, the TSH level was within the reference range and the fT4 level abnormal. These discordant test results, reflecting missed diagnoses, would have been missed when a two-step test approach would have been applied. The study authors state that in daily praxis this percentage of patients with thyroid disease will likely be lower when combined with information on known clinical history of these patients and other external factors. The two other case series did not report discordant test results for the TSH and T3/T4 serum levels, but reported the distribution of thyroid disorder diagnoses in clinically suspected patients. The economic systematic literature search retrieved ten studies after screening 353 records. All studies shared a consistent opinion on thyroid function tests likely being unnecessary used or overused, supporting the potential for cost-savings if a step-wise approach is adopted. The included studies investigating different test approaches concluded the two-step test approach the most cost-saving approach as compared to the one-step test approach. The studies conducted on hypothetical goals of the fT4/TSH ratio concluded that the reduction of unnecessary tests was deemed economically favourable. As these studies did not provide relevant cost data for constructing a Swiss budget impact model, additional sources were used for the inputs of the budget impact model.

A budget impact model, over a time horizon of five years (2020-2024), was developed to assess the potential for cost-savings in Switzerland by accounting the number of thyroid function tests at the population-level. Compared to the one-step thyroid function test approach the two-step test approach (i.e. including second blood draw) was cost-saving (approximately 6.1 million CHF in year 2020). The impact of implementing reflex testing (i.e. using the initial blood sample without the need for a second blood draw) in the two-step test approach was also taken into account. Reflex testing was considered the most favourable (cost-saving) test approach compared to the two-step test approach (approximately 2.1 million CHF in year 2020) and one-step test approach (approximately 8.2 million CHF in year 2020), respectively. The results for year 2021, year 2022, year 2023 and year 2024 were similar.

Nine observational studies and two guidelines informed the ethical, legal, social, and organisational

issues related to the use of thyroid function tests in adults with suspected thyroid dysfunction. Physicians expressed a degree of fear of possibly missing positive diagnoses, as well as experiencing pressure from patients requesting the complete diagnostic profile, when applying a two-step instead of a one-step test approach. The importance of good instructions to physicians were flagged as pivotal for a successful implementation of a two-step test approach.

CONCLUSION:

The clinical evidence was sparse, and the low quality does not permit to draw conclusions on a concrete number of cases missed when the two-step instead of the one-step test approach is applied in adults with suspected primary or secondary thyroid dysfunction in the Swiss setting. The budget impact analysis showed that the two-step test approach can be considered cost-saving compared to the one-step test approach. From a clinical perspective the two-step test approach seems to be justified, given the likely low risk of missed cases when applying this approach. The impact of reflex testing was also assessed and was more cost-saving compared to the one-step and two-step test approach, since the initial blood sample is used for possible additional thyroid function tests without the need for a second blood draw. When applying a two-step or reflex test approach instead of the one-step test approach, the possible impact of missing or misclassifying diagnoses and practical aspects of how to successfully implement the reflex testing approach should be considered.

Table of Contents

Abbreviations and acronyms

AACE	American Association of Clinical Endocrinologists
ATA	American Thyroid Association
BFS	Bundesamt für Statistik
CHF	Swiss Franc
ETA	European Thyroid Association
EUnetHTA	European Network for Health Technology Assessment
FOPH	Federal Office of Public Health
fT3	free Triiodothyronine
fT4	free Thyroxine
GRADE	Grading of Recommendations, Assessment, Development, and Evaluations
HTA	Health Technology Assessment
ICD-10	International Statistical Classification of Diseases 10
ISPOR	Professional Society for Health Economics and Outcomes Research
JBI	Joanna Briggs Institute
LC-MS/MS	Liquid Chromatography with tandem Mass Spectrometry
NICE	National Institute for Health and Care Excellence
NHS EED	National Health Service Economic Evaluation Database
NR	Not reported
OECD	Organisation for Economic Cooperation and Development
PICO	Population Intervention Comparator Outcome
P.P.	Per person
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RCT	Randomised Controlled Trial
SR	Systematic Review

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 this HTA report include the clinical outcomes, costs and budget impact, and the 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. Thyroid function tests are used for the diagnosis and monitoring of patients with thyroid disorders, i.e. patients with hypo- or hyperthyroidism. For the diagnosis of thyroid disorders, most guidelines recommend a two-step test approach: TSH should be measured first, only followed by a measurement of (free(f)) T4/(f)T3 if TSH is out of the reference range or if there are clinical symptoms indicating a thyroid disorder. However, it is observed that TSH and (f)T4/(f)T3 are often measured together in a one-step test approach. The argument in favour of the one-step test approach is prevention of missed diagnoses of patients with normal TSH levels and thyroid hormone levels outside the reference range. The argument in favour of the two-step test approach is the saving potential, as most T3/T4 tests can be prevented when TSH levels are found to be normal. Therefore, the policy objective of this HTA report is to evaluate the available clinical and economic evidence associated with the one-step and the two-step test approaches in adults with suspected primary or secondary thyroid dysfunction.

2 Research question(s)

To answer a policy question, the research question must 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. The central research questions for this topic focus mainly on the clinical and economic evidence in favour or against applying one of the two test approaches. In general, a systematic literature search for the best available evidence on the efficacy/effectiveness/safety, cost-effectiveness/budget impact, legal/social/ethical/organisational issues provide the first step towards the classical HTA. However, with respect to the clinical evidence the standard HTA approach is not suited for this HTA as neither the analytical and clinical validity nor the clinical utility of the tests are the objective of this report. Information regarding the number of missed adults with suspected primary or secondary thyroid dysfunction when applying the two-step test approach, the type of thyroid dysfunction diagnoses that are potentially missed, and the clinical and economic consequences of the one-step compared to the two-step test approach is more relevant to address in this HTA topic. To comply with this need the research questions regarding the clinical evidence are adapted accordingly.

HTA Research Questions

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

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

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

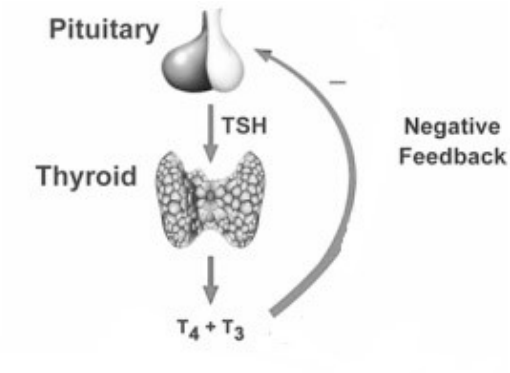
3) Are there any ethical, legal, social, or organisational issues related to the one-step versus two-step thyroid function test approach in adults with suspected primary or secondary thyroid dysfunction?

3 Medical background

Thyroid hormone signalling pathway

The thyroid is a butterfly-shaped gland, located in the front of the neck. Major thyroid hormones secreted by the thyroid gland are 80% thyroxine (T4) and 20% triiodothyronine (T3), which mainly regulate body metabolism. After its release from the thyroid gland, T4 is converted to T3, which is an active thyroid hormone, or to reverse T3, an inactive form. The amount of T4 and T3 produced by the thyroid gland is regulated by the thyroid stimulating hormone (TSH), which is secreted in the bloodstream by the pituitary gland. [1–3] Most thyroid hormones (99%) bind to proteins, and only 1% circulates freely in the blood, known as free T4/T3 (fT4/fT3), and is able to enter and affect body tissue. The release of TSH by the pituitary gland is regulated by the concentration of T4 and T3 in the blood in a converging pathway; when T4/T3 concentrations are low, the TSH production is increased (absence of negative feedback loop), and when T4/T3 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 of thyroid stimulating hormone (TSH), thyroxine (T4), and triiodothyronine (T3)*



*** Other forms of thyroid hormones are not included (e.g. T2 and rT3). Minus indicates a negative feedback loop. Reproduced with modifications from Freitas et al. 2012.** [4]

TSH is secreted during lifetime and increases significantly in periods of rapid growth, therefore TSH levels vary based on age, sex, and stage of life. In addition, external factors, such as stress, diet, and medication can result in fluctuation of TSH levels. A normal thyroid function with serum levels of TSH and (f)T4/(f)T3 within the reference ranges is referred as euthyroidism. [5,6]

Thyroid disorders

Thyroid disorder is a medical condition that affects the function of the thyroid gland. The most common thyroid problems involve abnormal production of thyroid hormones. Disorders of thyroid function are frequently diagnosed, with the prevalence in Europe varying between 2% and 6% in large population-based studies and the incidence calculated in a meta-analysis being 259 per 100'000 per year. [7–10] The prevalence of thyroid disorders depends on a large number of factors, of which the most important include: sex, age, geographic factors, and ethnicity. There is a clear female preponderance; females are five to eight times more likely than males to have thyroid problems. [6–9] Thyroid disorders, especially subclinical thyroid dysfunction, are common in the elderly, as the prevalence of thyroid disorders increases over age. [11,12] Also, the prevalence of thyroid disorders is higher in countries with iodine deficiency. [13] A simplified overview of thyroid disorders based on TSH and (f)T4/(f)T3 serum levels is listed in ***Table 1***, followed by a brief explanation on the two main types of thyroid disorder (i.e. hyperthyroidism and hypothyroidism). Thereafter, the difference between primary, secondary, subclinical, and central hyperthyroidism and hypothyroidism, respectively, is described.

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

<table>
<tr><th colspan="2" rowspan="2"></th><th colspan="3">TSH</th></tr>
<tr><th>Low</th><th>Normal</th><th>High</th></tr>
<tr><td rowspan="3">(f)T4 or (f)T3</td><td>High</td><td rowspan="2">Primary hyperthyroidism</td><td></td><td>Secondary hyperthyroidism</td></tr>
<tr><td colspan="2">Central hyperthyroidism</td></tr>
<tr><td>Normal</td><td>Subclinical hyperthyroidism</td><td>Euthyroidism</td><td>Subclinical hypothyroidism</td></tr>
<tr><td>Low</td><td>Secondary hypothyroidism</td><td></td><td rowspan="2">Primary hypothyroidism</td></tr>
<tr><td colspan="2">Central hypothyroidism</td></tr>
</table>

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

The two main types of thyroid disorders are hyperthyroidism and hypothyroidism:

- Hyperthyroidism (overactive thyroid) occurs when the thyroid gland produces too much of T3 and/or T4. [14]
- Hypothyroidism (underactive thyroid) occurs when the thyroid gland does not produce enough hormones (T3 and/or T4) to meet the metabolic demand of the body.

Four main categories can be defined, based on the origin and severity of the thyroid disorder: primary, secondary, subclinical, and central thyroid disorder:

Primary thyroid disorder: failure of the thyroid gland.

- Primary hyperthyroidism is defined as high levels of blood thyroid hormone (fT3 and/or fT4) and low levels of TSH (see ***Table 1***). Primary hyperthyroidism originates within the thyroid gland (e.g. caused by autoimmune thyroid disease).
- Primary hypothyroidism is defined as low levels of blood thyroid hormone (fT3 and/or fT4) due to destruction of the thyroid gland. The TSH levels are high (see ***Table 1***). This destruction is usually caused by autoimmunity, or an intervention such as surgery, radioiodine, or radiation. [15]

Secondary thyroid disorder: change in the thyroid hormones as a result of a disease not in the thyroid gland.

- In secondary hyperthyroidism there is excessive TSH in the circulation not caused by the thyroid gland but for example by TSH-producing pituitary adenomas, which, in turn, stimulate the thyroid gland to secrete thyroid hormones in excessive amounts. Both TSH and (f)T4/(f)T3 levels are high (see ***Table 1***). [3,16]
- In secondary hypothyroidism there is insufficient production of TSH to stimulate the thyroid gland to produce thyroid hormones and the (f)T4/(f)T3 levels go down. It is mostly due to pituitary or hypothalamic disease. [16] Both TSH and (f)T4/(f)T3 levels are low (see ***Table 1***).

Central thyroid disorder: rare disorder with discordant TSH level and thyroid hormones, a sub-entity of secondary thyroid disorder. Potentially missed when two-step test approach is applied.

- In central hyperthyroidism there is primary TSH overproduction by the pituitary gland with subsequent thyroid enlargement and hyperfunction. Two probable causes of central hyperthyroidism are TSH-producing pituitary adenomas and the syndrome of pituitary thyroid hormone resistance (PRTH). [17] Central hyperthyroidism is characterised by TSH levels high or within the reference range, and elevated circulating levels of (f)T4 and (f)T3 (see ***Table 1***).
- Central hypothyroidism is defined by a normal functioning thyroid gland and a defect of thyroid hormone production due to insufficient stimulation by TSH, caused by a pituitary gland or hypothalamus disorder. [18] In central hypothyroidism, the TSH levels are low or within the reference range, but concentration of (f)T4 and (f)T3 are below the reference range (see ***Table 1***).

Subclinical thyroid disorder: mild form of thyroid disorder.

- In subclinical hyperthyroidism TSH level is below the reference range, but concentrations of (f)T4/(f)T3 are within the reference range (see ***Table 1***). [19] The most common causes of subclinical hyperthyroidism include Graves' disease, autonomous functioning thyroid adenoma, and multinodular toxic goitre. [20]
- In subclinical hypothyroidism TSH level is above the reference range, but concentrations of (f)T4/(f)T3 are within the reference range (see ***Table 1***). [19] The most common cause of subclinical hypothyroidism is a condition known as Hashimoto thyroiditis. [21]

Testing for thyroid disorders

Thyroid disorders affect many body systems, the symptoms are often non-specific, and most single symptoms alone are not predictive of thyroid dysfunction, therefore there is a wide range of possible indications for testing. [22,23] The best way to diagnose thyroid dysfunction is through blood tests. The thyroid hormone levels in the blood (TSH, total and free T4/T3, and reverse T3) indicate if the thyroid gland is functioning properly. Not all combinations of the thyroid hormone levels are equally appropriate and useful in every diagnosis. TSH level assessment in a blood sample is usually the first test performed. The addition of T3 and T4 serum assessments, may help categorising a potential thyroid disease. Because T4 more accurately reflects how the thyroid gland is functioning, most changes in the functioning of the thyroid gland show up in T4 first, the free T4 is the most important hormone to measure. T3 tests are useful to identify hyperthyroidism or to determine the severity of hyperthyroidism. T3 testing is rarely helpful to identify hypothyroidism. Measurement of free T3 is possible, but the results of the tests are considered less reliable than those of bound T3. The results are affected by alterations in serum binding proteins that occur in many physiologic and disease states. Blood test for reverse T3 (rT3) is not performed in routine clinical practice but there are distinct entities (e.g. consumptive hypothyroidism) where the measurement of rT3 has an important role in characterising the phenotype. A normal TSH value usually means that the thyroid is functioning properly. [6] However, TSH levels of patients with a central thyroid disorder are mostly within the reference range and the concentrations of (f)T4/(f)T3 are outside the reference range. [24] Therefore central hyper- or hypothyroidism can be diagnosed rarely when TSH levels are within the reference range, and testing TSH first (i.e. with the two-step test approach) may miss diagnosis of central hyper- or hypothyroidism. Thyroid function tests are used for the diagnosis, screening, and monitoring of treatment of thyroid disorders. The focus of this HTA is on the application of thyroid function tests for diagnosis in adults with suspected primary or secondary thyroid dysfunction, for example adults with symptoms of thyroid disease (although one symptom alone may not be indicative of thyroid disease), or adults with depression or unexplained anxiety. [23] Thyroid function tests performed for screening (i.e. presumptive identification of unrecognised disease with tests in people who do not

have the symptoms of early disease) and monitoring or follow-up of thyroid dysfunction are out of the scope. [25]

Populations targeted for thyroid dysfunction testing

Professional societies have tried to define the indications which populations should be tested for thyroid dysfunction, because symptoms and indications may vary widely between subjects with suspected thyroid disorders. The American Association of Clinical Endocrinologists (AACE) and the American Thyroid Association (ATA) published an evidence-based guideline in 2012 [26] and an evidence-based white paper in 2016 [25] including a detailed list with diseases (e.g. diabetes mellitus type 1), symptoms (e.g. malaise and fatigue), medical history (e.g. history of thyroid dysfunction), and specific medication (e.g. lithium) as indications for thyroid dysfunction aggressive case finding. In 2019, the National Institute for Health and Care Excellence (NICE) published an evidence-based guideline [27] with a short list of recommendations on who should be tested for thyroid disease: persons with diabetes mellitus type 1, depression or unexplained anxiety, and new onset atrial fibrillation. Furthermore, NICE reported two specific contraindications for thyroid dysfunction testing: do not test for thyroid dysfunction during an acute illness unless it is suspected that the acute illness is due to thyroid dysfunction, because the acute illness may affect the test results; and do not offer testing for thyroid dysfunction solely, because a person has diabetes mellitus type 2. No Swiss or European-level guidelines with data on which populations should be targeted for thyroid dysfunction testing were found (See ***Supplemental Document***).

4 Technology description

Thyroid function tests quantify serum levels of TSH and the circulating thyroid hormones fT4/fT3, reflecting the ability of the thyroid gland to produce and regulate thyroid hormone production. [23,27] Thyroid function tests have high sensitivity and specificity, therefore these tests are amongst the most widely used blood tests when a thyroid disorder is suspected. [22] Since 1995, third-generation assays, which represent an extra 10-fold increase in sensitivity compared to second-generation assays, have been used. [9] Two possible test approaches can be applied: the one-step test approach (i.e. measurement of TSH and thyroid hormone (f)T4/(f)T3 simultaneously) or two-step test approach (i.e. testing TSH first, followed by the measurement of thyroid hormone (f)T4/(f)T3 when TSH is out of the reference range). When hypo- or hyperthyroidism is suspected in adults, clinical guidelines recommend the two-step test approach [27]: first test TSH alone and then a) if the TSH is above the reference range, test fT4, or b) if the TSH is below the reference range, test fT4 and fT3. This second testing can be done from an additional blood sample or determined from the first sample. Furthermore, second testing from the first blood sample can be assessed automatically on the basis of algorithms (i.e. reflex testing) or by laboratory professionals (i.e. reflective testing).[28] For the one-step test approach only one blood sample is taken for the analysis.

Because thyroid hormones and TSH test results are variable over time, current guidelines emphasise to have an interval of two to three months before the diagnosis can be established. To have an accurate blood sampling, two complete sets of thyroid hormone tests should be analysed for comparable results. [29] Thyroid function tests are not standardised or harmonised. The reference intervals of thyroid function tests are hampered by wide variability of commercial immunoassays. Clinical laboratories may address this issue by using assay-specific reference intervals. Some laboratories may use Liquid Chromatography with tandem Mass Spectrometry (LC-MS/MS), while others prefer immunometric assays for thyroid function tests. [30] Immunoassay tests deliver quantitative or semi-quantitative results, and may not always be very accurate, whilst LC-MS/MS analysis enables the determination of several compounds in a single analysis.

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

5 Population, Intervention, Comparator, Outcome (PICO)

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

Table 2: PICO for the clinical and costs and budget impact systematic review

P: Adults with suspected primary or secondary thyroid dysfunction

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

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

O: **Primary clinical outcome**:

Number and percentage of patients classified in nine categories:

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

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

* Number of cases potentially missed with two-step test approach.

Secondary clinical outcome:

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

Economic outcomes

- Medical costs
- Budget impact

Keys: fT3=free triiodothyronine, fT4=free thyroxine, TSH=thyroid-stimulating hormone.

6 Efficacy, effectiveness, and safety

Summary statement efficacy, effectiveness, and safety

With a broad systematic literature search for clinical evidence on the one-step test approach in adults with suspected primary or secondary thyroid dysfunction, three case series were included in this HTA. Most studies were excluded, because they were not conducted in clinically suspected populations, did not define their study population in sufficient detail, or the data could not be categorised in one of the nine outcome categories (see PICO).

The case series were conducted in non-western countries. One case series analysed levels of TSH and fT4 in inpatients and outpatients with clinically suspected primary or secondary thyroid dysfunction from a general hospital in Papua New Guinea. In 5.3% of the thyroid function tests the TSH level was within the reference range and the fT4 level abnormal. These discordant test results, reflecting missed diagnoses, would have been missed when a two-step test approach (i.e. testing TSH first, followed by the measurement of fT4 when TSH is out of the reference range) had been applied. The study authors state that in daily praxis this percentage of patients with thyroid disease will likely be lower when combined with clinical information on known thyroid disease on treatment, sick-euthyroid syndrome, or patients using drugs of which it is known these interfere with thyroid hormone levels. The two other case series did not report discordant test results for the TSH and T3/T4 serum levels but reported the distribution of thyroid disorder diagnoses in clinically suspected patients.

Based on the sparse, low quality clinical evidence, no valid conclusions can be drawn on the number of cases missed when adults with suspected primary or secondary thyroid dysfunction are tested with the two-step test approach. Nonetheless the sparse evidence, in combination with insight from clinical experts indicates that the percentage of not detected thyroid disease patients when applying the two-step test approach is likely to be very low.

6.1 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). [31,32]

The SR process consists of the following fundamental steps:

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

In the following sections, the SR methodology is described in detail.

6.1.1 Databases and search strategy

6.1.1.1 Search strategy

PubMed (MEDLINE) and Embase.com databases were searched for peer-reviewed scientific literature. Since there is considerable overlap in studies included in other literature databases (such as Cochrane Library), it was decided to search in these two main databases. The search strategy was built using the PICO-framework (see ***Chapter 5***). A general search string was made for the 'intervention' (i.e. thyroid function test). We did not include search strings for 'population' and 'outcome', because population details are not always described in the title or abstract of these articles and the outcomes of studies on thyroid function tests are very diverse. Including the search strings for 'population' and 'outcome' to the search strategy will lower the number of hits and increase the chance of missing pertinent articles. Animal studies, case reports, irrelevant publication types (e.g. editorials, expert opinions), and congress abstracts were excluded with additional search strings. We searched in four languages: English, German, French, and Italian. The publication period was limited from 1990 to 9 September 2021. This time period is based on a combination of expert input and publication dates found with a preliminary search for key articles. The details of the search strategies are included in ***Appendix 13.1.1***. 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.

6.1.1.2 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 study 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 6.1.1.3***).
3. Screening during data extraction phase: further scrutiny of the article during the data extraction phase might lead to exclusion, for example for studies with unexplained errors in their patient flow or studies based on duplicate data. No studies were excluded in this third selection step.

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 6.2.1***) and an overview table with the reasons for exclusion per excluded study (***Appendix 13.1.2***). The implemented quality control during the selection process is described in ***Section 6.1.4***.

6.1.1.3 Inclusion and exclusion criteria

The inclusion and exclusion criteria applied during the selection processes are presented in ***Table 4***. Although the intervention of interest is the two-step test approach, for answering the research question only studies applying the one-step test approach are potentially relevant: The one-step test approach combines TSH and (f)T4/(f)T3 data and thus fills out all nine outcome categories (see ***Table 3***, copied from PICO). Studies providing data on the two-step test approach on the other hand can only provide data on (f)T4/(f)T3 if TSH is outside the reference range (see ***Table 3***, cells in blue). If TSH is normal, the second step to analyse (f)T4/(f)T3 will not be performed. However, for this HTA we are interested in the number of missed cases (see ***Table 3***, cells in red) not identified by the two-step test approach. These cases can only be found by studies that applied at least the one-step test approach.

Table 3: Nine outcome categories based on TSH level and (f)T4/(f)T3 level

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

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

Table 4: Inclusion and exclusion criteria for the clinical outcomes systematic literature search

	Inclusion	**Exclusion**
Period publication	1990 to 9 September 2021	<1990
Language of publication	• English • German • French • Italian	All other languages
Country of study	Worldwide	-
Study design/type	• Randomised controlled trials • Comparative non-randomised trials (e.g. cohort studies, retrospective observational studies) • Non-comparative studies (e.g. case series)	• Systematic reviews* • Narrative reviews • Case reports • Abstract only (e.g. conference abstract) • Non-pertinent publication types (e.g. expert opinion, letter to editor, editorial, comment)
Study quality	Sufficient methodological quality and coherent reporting of the results (e.g. data reported in text and tables are coherent without unexplainable errors and interpretation of the study results is not hampered)	Major insufficient methodological quality or incoherent reporting of the results (e.g. unexplained errors in patient flow)
Study population	Adults with suspected primary or secondary thyroid dysfunction	• Children • Populations screened for thyroid disorders • Populations monitored for the treatment of thyroid disorders • Studies only including females with specific female sex hormonal states, e.g. pregnant females, non-pregnant females on fertility-related treatment, or menopausal females
Study intervention	Thyroid function tests performed with the two-step test approach (i.e. testing TSH first, followed by the measurement of thyroid hormone (f)T4/(f)T3 when TSH is out of the reference range) or no intervention (only studies investigating the one-step test approach)	All other study interventions
Comparison	Thyroid function tests performed with the one-step test approach (i.e. measurement of TSH and thyroid hormone (f)T4/(f)T3 simultaneously)	All other comparators
Study outcomes	See PICO	Other outcomes

Keys: (f)T3=(free) triiodothyronine, (f)T4=(free) thyroxine, PICO=Population Intervention Comparator Outcome, TSH=thyroid-stimulating hormone. *The proposed approach for SRs was: Relevant SRs will be selected during the screening of title and abstract phase; during the full-text screening phase, reference lists of good quality SRs will be checked for possibly missed relevant individual articles; no data extraction will be performed for SRs, only for relevant individual articles. However, we did not find SRs on this topic.

6.1.2 Assessment of quality of evidence

For this HTA, it was not possible to determine the method to assess the risk of bias beforehand. After inclusion of the studies and definition of the study designs, the most appropriate critical appraisal method was chosen and discussed with the FOPH. All included studies were classified as case series and critically appraised by the Joanna Briggs Institute (JBI) checklist for case series. [33]

The JBI checklist for case series consists of ten questions with the answers being classified as yes, no, unclear, or not applicable (see ***Table 6***). The risk of bias of the study design was categorised as: low risk of bias, moderate/unclear risk of bias, or high risk of bias based on the crucial limitations for one or more of the criteria. The risk of bias was assessed by two independent researchers. In case of discrepancy a third researcher was consulted to reach consensus.

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

Data from the included studies found in the peer-reviewed literature was extracted using standardised tables in Word to present relevant information for the review objectives per included article. The data was further summarised and presented in this HTA report in a study characteristics table, risk of bias table, summary tables, and narrative text. A summary table was compiled for the distribution of the thyroid status of blood samples classified in nine categories based on low/normal/high TSH combined with low/normal/high fT4. In each cell we provided a short description of the TSH and fT4 combination (i.e. primary hyper/hypothyroidism, secondary hyper/hypothyroidism, subclinical hyper/hypothyroidism, euthyroidism, and missed cases with two-step test approach). When data in the article was lacking to apply the classification in nine categories of thyroid status, a separate summary table was made for the distribution of the diagnosis of thyroid disorders in patients with both TSH and T3/T4 assayed.

The options for clinically relevant data merging/stratification were assessed. It was not possible to calculate pooled estimates, since at least two studies did not report on the same outcome and/or were defined in the same way. For the included non-comparative studies and outcomes of interest it was not

possible to apply the Grading of Recommendations, Assessment, Development, and Evaluations (GRADE) approach. [34]

6.1.4 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.
    - 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. However, we did not find SRs on this topic.
    - The supplementary search technique citation chasing (i.e. backward by finding other studies cited within the selected articles) was applied in addition to the database searches. No additional studies 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 a study, this was discussed. If the differences remained after discussion, the study 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.

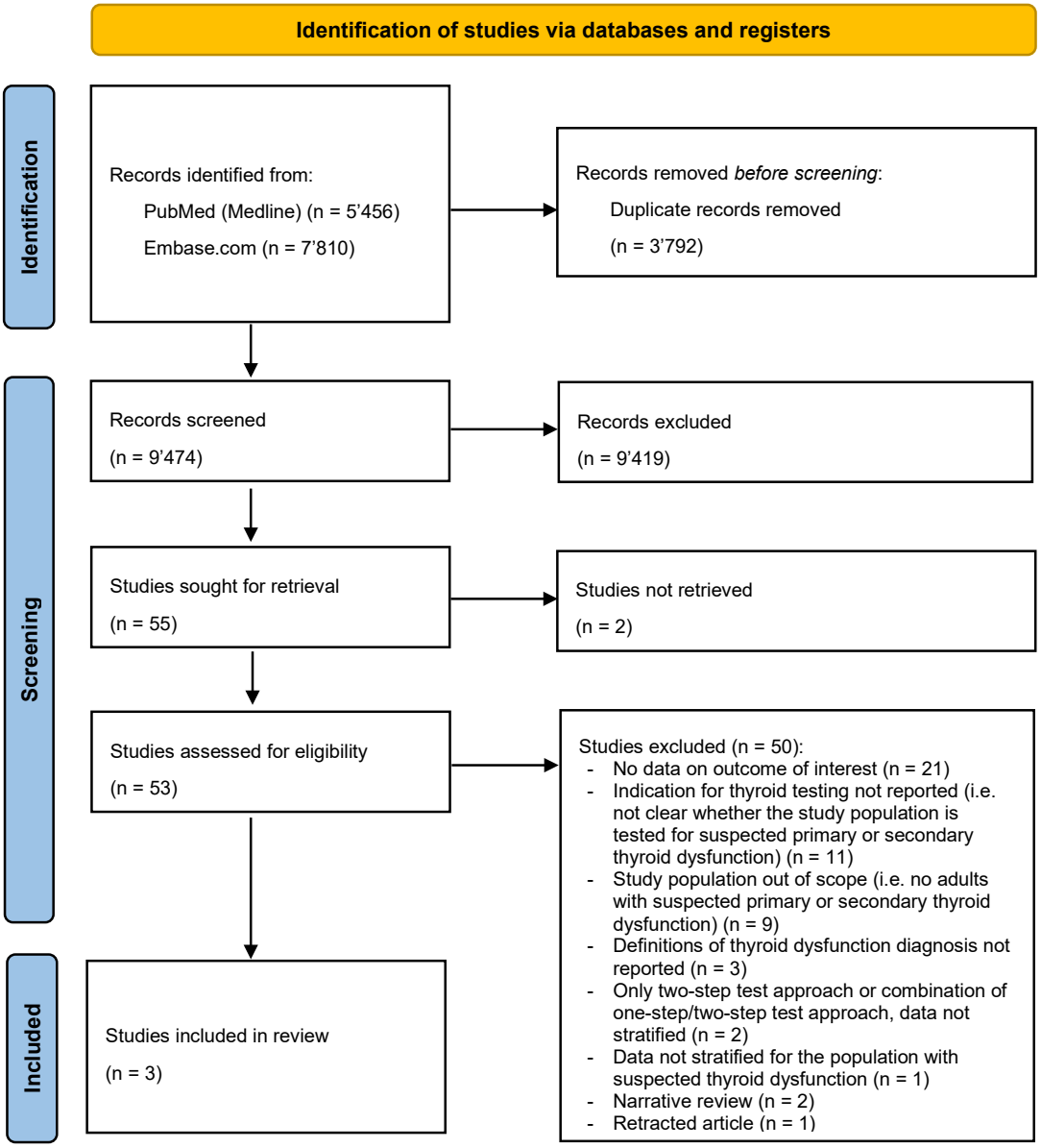
- Data extraction and synthesis
  - The critical appraisal of included studies was done in duplicate. In case of discrepancy a third researcher was consulted to reach consensus.
  - The data extraction tables were fully checked with the original articles by a second researcher.
  - The summary tables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

6.2 Results efficacy, effectiveness, and safety

6.2.1 PRISMA flow diagram

In total, 9'474 unique records were identified in PubMed (MEDLINE) and Embase.com with our search strategy. Of those, 9'419 records were excluded based on their title and abstract, because these records did not contain relevant data for the objectives of this HTA. Many articles were aimed at populations screened for thyroid disorders, populations monitored for the treatment of thyroid disorders, or assessing TSH reference values for specific populations. This resulted in 55 studies selected to be screened in full text. We were not able to retrieve two studies in full-text (see footnote below the PRISMA flowchart in ***Figure 2***. After applying the inclusion and exclusion criteria, three studies were included for full analysis. The main reasons for exclusion were no data on outcome of interest (n=21 studies), indication for thyroid testing not reported (i.e. not clear whether the study population is tested for suspected primary or secondary thyroid dysfunction; n=11 studies), study population out of scope (i.e. no adults with suspected primary or secondary thyroid dysfunction; n=9 studies), and definitions of thyroid dysfunction diagnosis not reported (n=3 studies). 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 study is enclosed in ***Appendix 13.1.2***.

Figure 2: PRISMA flowchart of the clinical outcomes systematic literature search



References of studies not retrieved: 1) Hijam D, Hidangmayum D, Singh WG. Thyroid dysfunction among referred cases for thyroid function test. JMS - Journal of Medical Society 2004;18(3):91-92. 2) Qavi A, Khokhar GN, Zahoor A, et al. Significance of hypercholesterolemia in hypothyroid patients. Medical Forum Monthly 2013;24(12):86-88.

6.2.2 Study characteristics and quality assessment of included studies

Only a very limited number of studies investigate thyroid function tests performed with the one-step test approach in a population of adults with suspected primary or secondary thyroid dysfunction. Three case series from non-western countries were selected for full analysis in this HTA report. The study characteristics are summarised in ***Table 5*** and the risk of bias in ***Table 6***.

Kende et al. 2002 [35] analysed the records of thyroid function tests performed between January 1996 and May 2000 for clinically suspected inpatients and outpatients from a general hospital in Papua New Guinea. This study focussed on records of thyroid function tests (i.e. not patients), information on age and gender was not available. The blood samples with both TSH and fT4 assayed were classified in nine categories of thyroid status (***Table 7***). The risk of bias of this case series was high (see ***Table 6***). It was unclear whether the participants were included consecutively. The number of blood samples was not identical to the number of patients, because of inclusion of data from patients measured on multiple occasions. No details were reported on the study population or the setting (e.g. iodine status in the region).

In India, two case series were conducted in patients with a clinical suspicion of thyroid disorders evaluated with thyroid function tests for TSH, T3, and T4 serum levels. Between April and September 2012, Bansal et al. 2014 [36] investigated 96 ambulatory patients at a tertiary care centre. The patients were aged between 15 and 55 years and 22% was male. The iodine consumption in the tribal Bastar region was relatively low and the ground water of some areas was rich in fluorine, which may cause fluorosis and is associated with thyroid dysfunction. From January to December 2017, Natasha and Badiger, 2019 [37] included 100 elderly outpatients aged 60+ years at a tertiary care teaching hospital. One third of the patiens was male. No additional details were reported on the setting. Both case series reported the distribution of thyroid disorder diagnoses, however detailed definitions were lacking and combined data for TSH and T3/T4 serum levels were not provided. It was unclear whether the participants were included consecutively and if the inclusion of participants was complete. The risk of bias of both case series was high (see ***Table 6***).

6.2.3 Evidence table

Table 5: Study characteristics of the included studies

First author, year	Study design	Country Study period	Study population/ data source	Thyroid function tests & reference range Type of assay performed	Sample size	Outcomes & definitions
Kende et al. 2002 [35]	Case series	Papua New Guinea Jan 1996-May 2000	Records of thyroid function tests for clinically suspected inpatients and outpatients from a general hospital in Port Moresby* *Age (mean±SD)* NA *Sex (% male)* NA	- TSH 0.3-5.0 µU/ml - fT4 9-25 pmol/L Assays performed by microparticle immunoassay technique (ABBOTT IMX) using hormone-specific antibodies.	n=2'768 blood samples† with both TSH & fT4 assayed	n (%) samples classified as: - hyperthyroidism: low TSH/high fT4 - not further specified: low TSH/normal fT4 - hypothyroidism: low TSH/low fT4 - not further specified: normal TSH/high fT4 - euthyroidism: normal TSH/normal fT4 - not further specified: normal TSH/low fT4 - hyperthyroidism: high TSH/high fT4 - not further specified: high TSH/normal fT4 - hypothyroidism: high TSH/low fT4
Bansal et al. 2014 [36]	Case series	India April-Sept 2012	Ambulatory patients aged 15-55 years with a clinical suspicion of thyroid disorders at a tertiary care centre in Bastar area (tribal area) of Chhattisgarh *Age (mean±SD)* Mean: NR 15-35 years: 49% 35-55 years: 51% *Sex (% male)* 21.9%	- TSH 0.3-6.3 µU/ml - T3 0.5-1.9 ng/ml - T4 4.8-11.6 pg/dl (females) 4.4-10.8 pg/dl (males) Assays performed by enzyme-linked immunosorbent assay method (Omega diagnostics)	n=96 patients	n (%) patients diagnosed as: - euthyroidism: not further specified (based on reference ranges and levels of TSH/T3/T4 among patients) - hypothyroidism: not further specified (based on reference ranges and levels of TSH/T3/T4 among patients) - hyperthyrodism: not further specified (based on reference ranges and levels of TSH/T3/T4 among patients)
Natasha and Badiger, 2019 [37]	Case series	India Jan-Dec 2017	Elderly outpatients aged >60 years with a clinical suspicion of thyroid disorders at a tertiary care teaching hospital in Belagavi (Karnataka) *Age (mean±SD)* 67.7±7.2 years *Sex (% male)*	- TSH reference NR - T3 reference NR - T4 Reference NR Type of assay performed NR	n=100 patients	% patients diagnosed as: - subclinical hypothyroidism: not further specified - hypothyroidism: not further specified (described as overt hypothyroidism in the discussions section) - subclinical hyperthyroidism: not further specified - hyperthyroidism: not further specified

			33%			

Keys: fT4=free thyroxine, NA=not applicable, NR=not reported, TSH=thyroid-stimulating hormone, T3=triiodothyronine, T4=thyroxine. * Most of the samples were from clinics around Port Moresby city, but test samples were received also from all parts of Papua New Guinea. † Not identical to the number of patients, because of inclusion of data from patients measured on multiple occasions.

Table 6: Risk of bias of the included studies

	Kende et al. 2002 [35]				Bansal et al. 2014 [36]				Natsha and Badiger, 2019 [37]			
Checklist for case series*	**Yes**	**No**	**Unclear**	**Not applicable**	**Yes**	**No**	**Unclear**	**Not applicable**	**Yes**	**No**	**Unclear**	**Not applicable**
Were there clear criteria for inclusion in the case series?												
Was the condition measured in a standard, reliable way for all participants included in the case series?												
Were valid methods used for identification of the condition for all participants included in the case series?												
Did the case series have consecutive inclusion of participants?												
Did the case series have complete inclusion of participants?												
Was there clear reporting of the demographics of the participants in the study?												
Was there clear reporting of clinical information of the participants?												
Were the outcomes or follow up results of cases clearly reported?												
Was there clear reporting of the presenting site(s)/clinic(s) demographic information?												
Was statistical analysis appropriate?												
Risk of bias	*High risk of bias*				*High risk of bias*				*High risk of bias*			

* Joanna Briggs Institute Checklist for Case Series [33].

6.2.4 Findings clinical outcomes

Primary clinical outcome

In the case series of Kende et al. 2002 [35] in total 3'188 thyroid function tests were assayed in Papua New Guinea over a 4 years and 5 months period and in 2'768 blood samples both TSH and fT4 were examined. The requests for thyroid function tests came from clinically suspected patients from the outpatient as well as inpatient departments. The assay results showed primary hyperthyroidism in 22.4%, primary hypothyroidism in 4.1%, and secondary hyperthyroidism or secondary hypothyroidism in 1.1% of the blood samples (***Table 7***). Euthyroidism was found in half of the thyroid function test requests. In 5.3% of the thyroid function tests the TSH level was within the reference range and the fT4 level abnormal. These discordant test results would have been missed when a two-step test approach (i.e. testing TSH first, followed by the measurement of fT4 when TSH is out of the reference range) would have

been applied. The authors state that the blood TSH and fT4 patterns were not indicative of the true distribution of various thyroid disorder states in patients since they did not know the full clinical history of the patients. The laboratory results were not related to clinical findings by the authors.

Table 7: Distribution of the thyroid status of 2'768 blood samples assayed with TSH and fT4 (Kende et al. 2002 [35])

		TSH n (%)		
		Low	**Normal**	**High**
fT4 n (%)	**High**	620 (22.4%) *primary hyperthyroidism*	55 (2.0%) *missed cases with two-step test approach**	31 (1.1%) *secondary hyperthyroidism*
	Normal	212 (7.7%) *subclinical hyperthyroidism*	1'378 (49.8%) *Euthyroidism*	238 (8.6%) *subclinical hypothyroidism*
	Low	31 (1.1%) *secondary hypothyroidism*	90 (3.3%) *missed cases with two-step test approach**	113 (4.1%) *primary hypothyroidism*

Keys: fT4=free thyroxine. * Cases in the grey box have a TSH level within the reference range and abnormal fT4 level and would not have been detected when a two-step test approach was applied, that is first testing TSH, followed by the measurement of thyroid hormone (f)T4/(f)T3 when TSH is out of the reference range.

With the data reported in the other two included case series it was not possible to make a comparable table like ***Table 7*** with the combination of TSH and T3/T4 serum levels. ***Table 8*** shows the distribution of thyroid disorder diagnoses in patients with a clinical suspicion of thyroid disorders as reported in the case series of Bansal et al. 2014 and Natasha and Badiger, 2019. [36,37] In the case series of Bansal et al. 2014 [36] a thyroid disorder was diagnosed in 64 of the 96 patients (66.7%) in the tribal Bastar region of India. Hypothyroidism was diagnosed in 37.5% of the patients and hyperthyroidism in 29.2%. In elderly patients in India, hypothyroidism was the most common thyroid abnormality reported in the case series of Natsha and Badiger, 2019. [37] Hypothyroidism was diagnosed in 12% of the patients and subclinical hypothyroidism in 10% of the patients. Hyperthyroidism and subclinical hyperthyroidism were found in 2% and 4%, respectively, of the elderly patients with a clinical suspicion of thyroid disorder.

Table 8: Distribution of thyroid disorder diagnoses in patients with both TSH and T3/T4 assayed

Reference (number of patients)	Thyroid disorder diagnosis n (%)				
	Hyperthyroidism	Subclinical hyperthyroidism	Euthyroidism	Subclinical hypothyroidism	Hypothyroidism
Bansal et al. 2014 [36] (n=96)	28 (29.2%)	NR	32 (33.3%)	NR	36 (37.5%)
Natasha and Badiger, 2019 [37] (n=100)	2 (2%)	4 (4%)	NR	10 (10%)	12 (12%)

Keys: NR=not reported.

Secondary clinical outcome

The study of Kende et al. 2002 [35] provided a short narrative description of the 5.3% of thyroid function tests which would have been missed when a two-step test approach were to be applied. The authors state that only a small percentage of patients with true thyroid disease will be included in the 5.3% discordant test results; patients missed by TSH-only testing will include patients with known thyroid disease on treatment, sick-euthyroid syndrome, or patients using drugs of which it is known these interfere with thyroid hormone levels. Furthermore, they highlight that many patients in Papua New Guinea present with clear clinical features of thyroid disease. This will make it unlikely that patients with significant thyroid dysfunction will be missed by assaying TSH alone, since diagnosis will be clearly evident from clinical and laboratory tests. The authors emphasise that when normal TSH values are found in patients with strong clinical features of thyroid disease, further tests for fT4 and fT3 can be requested after consulting the laboratory staff.

6.2.5 GRADE Summary of Findings Table

Not applicable.

7 Costs and budget impact

Summary statement costs and budget impact

The systematic literature search for studies on costs and budget-impact related to thyroid function testing in adults with suspected primary or secondary thyroid dysfunction retrieved ten cost studies. All studies shared a consistent opinion on thyroid function tests likely being unnecessary used or overused, supporting the potential for cost-savings if a stepwise approach is adopted. The included studies investigating different test approaches concluded the two-step test approach (i.e. testing TSH first, followed by the measurement of thyroid hormone (f)T4/(f)T3 when TSH is out of the reference range) the most cost-saving approach as compared to the one-step test approach (i.e. measurement of TSH and (f)T4/(f)T3 simultaneously). The studies conducted on hypothetical goals of the fT4/TSH ratio concluded that the reduction of unnecessary tests were deemed economically favourable. As these studies did not provide relevant cost data for the Swiss budget impact model, Swiss databases, publicly available sources, and expert opinion were sourced for the inputs of the budget impact model.

A budget impact model, over a time horizon of five years (2020-2024), was developed to assess the potential for cost-savings in Switzerland by accounting the number of thyroid function tests at the population-level. Compared to the one-step thyroid function test approach the two-step test approach (i.e. including second blood draw) was cost-saving (approximately 6.1 million CHF in year 2020). The impact of implementing reflex testing (i.e. using the initial blood sample without the need for a second blood draw) in the two-step test approach was also taken into account. Reflex testing was considered the most favourable (cost-saving) test approach compared to the two-step test approach (approximately 2.1 million CHF in year 2020) and one-step test approach (approximately 8.2 million CHF in year 2020), respectively. The results for year 2021, year 2022, year 2023 and year 2024 were similar.

7.1 Methodology costs and budget impact

7.1.1 Databases and search strategy

7.1.1.1 Search strategy

The costs and budget impact systematic literature search followed the principles of the systematic literature search outlined in ***Chapter 6.1***. PubMed (MEDLINE) and Embase.com databases were searched for peer-reviewed scientific literature. In addition, the NHS EED economic database was searched. The

searches were built using the PICO-framework (see ***Chapter 5***). In PubMed (MEDLINE) and Embase.com, the search terms of the clinical outcome literature search were combined with costs and budget impact search terms. The details of the search strategy are presented in ***Appendix 13.1.1***.

7.1.1.2 Selection procedure

From the articles retrieved from PubMed (MEDLINE), Embase.com, and NHS EED, the 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 on relevancy for the objectives by the title and abstract. In this step, articles that seemed 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 article:

   The full-text articles, selected in the first step, were assessed based on the inclusion and exclusion criteria as defined in ***Table 9***. Articles were included if they fulfilled the inclusion criteria and excluded when they did not.

The process of selection and inclusion and exclusion of articles was recorded in both Microsoft Excel and Rayyan (www.rayyan.ai). This method provides transparency regarding all selection steps and assures reproducibility. The selection procedure is presented in a PRISMA flow chart (***Figure 3***).

7.1.1.3 Inclusion and exclusion criteria

The inclusion and exclusion criteria applied during the selection processes are listed in ***Table 9***.

Table 9: Inclusion and exclusion criteria for costs and budget impact analyses studies

	Inclusion	**Exclusion**
Period publication	No restriction/from database inception	-
Country of study	Worldwide	-
Study design/type	Budget impact analysis Costs analysis	Resource use measurement
Study population	Adults with suspected primary or secondary thyroid dysfunction	• Children • Populations screened for thyroid disorders •Populations monitored for the treatment of thyroid disorders • Studies only including females with specific female sex hormonal states, e.g. pregnant females, non-pregnant females on fertility-related treatment, or menopausal females
Study intervention	Thyroid function tests performed with the two-step test approach (i.e. testing TSH first, followed by the measurement of (f)T4/(f)T3 when TSH is out of the reference range) or no intervention (only studies investigating the one-step test approach)	All other study interventions
Study comparison	Thyroid function tests performed with the one-step test approach (i.e. measurement of TSH and (f)T4/(f)T3 simultaneously)	All other comparators
Study outcomes	See PICO	Other outcomes

Keys: (f)T3=(free) triiodothyronine, (f)T4=(free) thyroxine, PICO=Population Intervention Comparator Outcome, TSH=Thyroid-stimulating hormone.

7.1.1.4 Quality control

- Search strategy

The search terms for the costs and budget impact were developed together with a medical information specialist. The 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

Studies were selected in duplicate by two independent researchers based on title and abstract and in full text. When there were differences between the findings of the two researchers, these differences were discussed. In case of discrepancy or disagreements, a third researcher was consulted.

7.1.2 Methodology data extraction, analysis and synthesis of health economic data

7.1.2.1 Data extraction

Relevant data from the included articles found in the peer-reviewed literature were summarised using a data-extraction spreadsheet in Microsoft Excel, and reported the following information in ***Table 10***:

- First author, year
- Country
- Study population
- Study design
- Intervention
- Comparator
- Economic outcomes
- Conclusions of the study authors

7.1.2.2 Data synthesis

Data synthesis was done using descriptive comparisons of the study question, methods, and results. An evidence table is included which presents key information described in the data extraction ***Subchapter 7.2.3***. The analytical approaches used in the studies were compared and their robustness was discussed. The published studies did not provide sufficient information on the costs of thyroid function tests, therefore Swiss databases and publicly available sources were sourced. A new budget impact model was developed for Switzerland. The general approach to the budget impact modelling and analysis is summarised in ***Subchapter 7.2.5***.

7.1.2.3 Quality control

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

- The data filled into the data extraction spreadsheet were checked by two reviewers.
- The data synthesis files and evidence table were checked by two reviewers.

7.1.3 Assessment of quality of evidence

Not applicable. There are no quality assessment tools specifically for the included costs or budget impact studies.

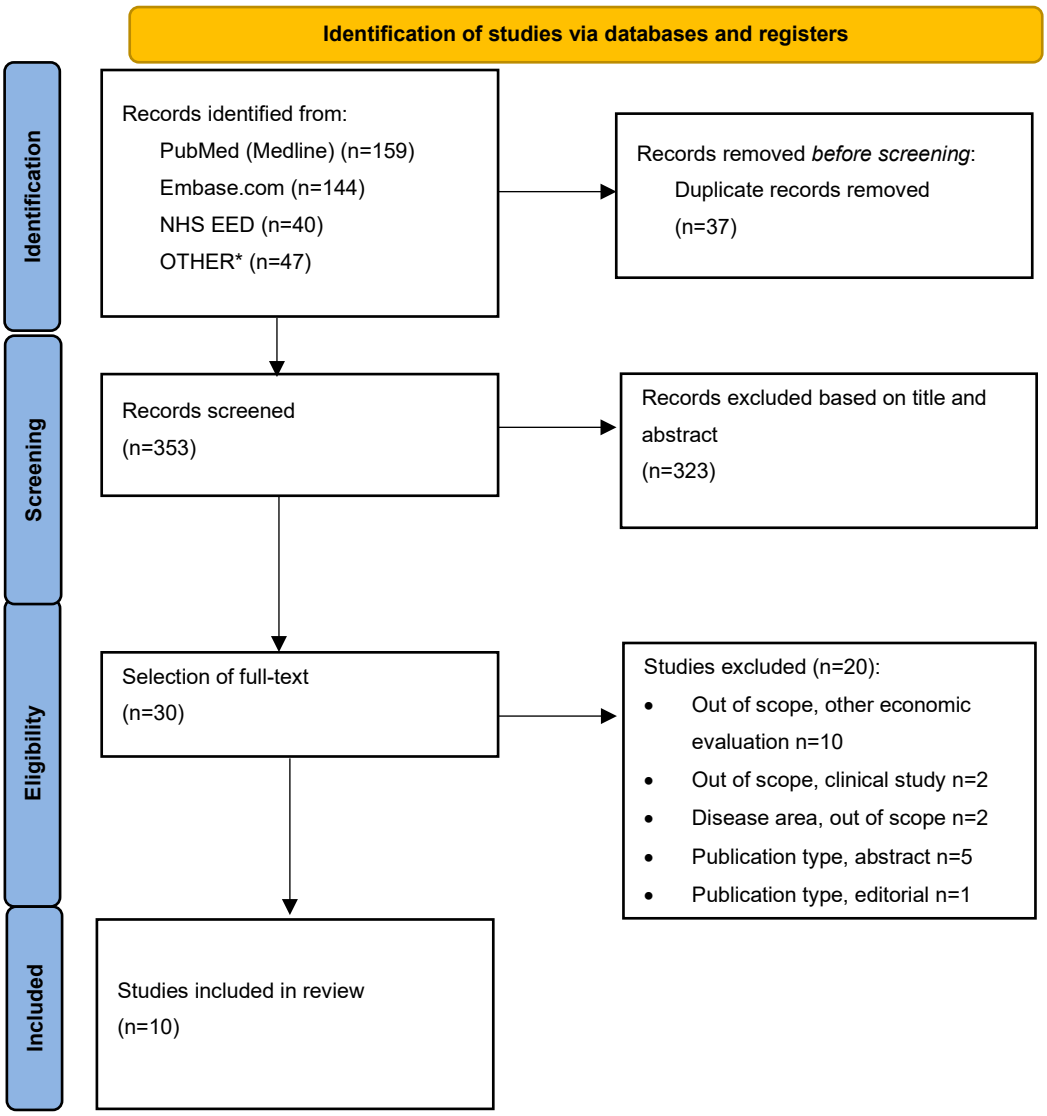
7.2 Results costs and budget impact

7.2.1 PRISMA flow diagram

Our search strategy yielded 390 unique records, which were identified in PubMed (MEDLINE), Embase.com, NHS EED, and titles of interest gathered during the clinical systematic literature review. In total, 37 titles were excluded due to duplications and 323 titles were excluded based on title and abstract using the eligibility criteria for this HTA. After the full-text review, 20 studies were excluded. The main reasons for exclusion were studies being out of scope (other economic evaluation n=10, clinical study n=2, and disease area n=2) and publication types (abstract n=5 and editorial n=1).

An overview of the reasons for exclusion is presented in the PRISMA flow chart (***Figure 3***) and a detailed overview of the reasons for exclusion for each excluded study is included in ***Appendix 13.2.2***.

Figure 3: PRISMA flowchart of the costs and budget impact systematic literature search



Keys: NHS EED=National Health Service Economic Evaluation Database *Other refers to titles obtained in the clinical systematic literature search review.

7.2.2 Study characteristics of included studies

All included studies were classified as cost studies, and the study characteristics can be viewed in ***Table 10***. While most of the studies included patients of the general population subjected to thyroid function tests, one study, Nordyke et al. 1998 [38], reviewed a population of patients tested for (suspected) thyroid disease. The sample size of the studies varied, both in terms of number of patients included in studies and number of tests reviewed. Kiel et al. 2020 [39] included 5'552 patients, Nordyke et al. 1998 [38] included 2'000 patients, and Vanderpump et al. 1997 [40] included 550 samples of TSH tests. Several studies looked at the total number of thyroid function tests analysed at a selection of laboratories and hospitals, and not all reported on the number of tests or patients. The majority of studies were conducted in Western countries, mostly in the United States and Spain. The interventions reviewed mostly included a two-step test approach. Comparisons varied according to the research question of the study: interventions mostly focused on a version of the two-step algorithm test or a hypothetical fT4/TSH ratio goal with corresponding comparators. The studies by Caldarelli et al. 2017 [43] and Feldkamp et al. 1996 [41] compare a similar two-step test approach, and report subsequent savings. Salinas et al. 2017 reviewed the two-step test approach by limiting unnecessary tests to achieve a hypothetical appropriate tests ratio. [42,44,45] Two studies that did not include a comparator were Kiel et al. 2020 [39] and Salinas et al. 2017. [44]

7.2.3 Evidence table

Table 10: Included cost studies

First author, year	Study design	Country	Study population/ data source	Intervention	Comparator	Cost items and Outcomes	Currency and year	Economic findings	Study Conclusions
Caldarelli et al. 2017 [43]	Cost study	Italy	Test prescriptions in the clinical laboratory of a hospital in 2012 (TSH 34'985, fT4 29'283, fT3 28'260) and 2015 (TSH 31'544, fT4 24'548, fT3 22'423, and 3'386 TSH reflex tests)	TFT prescriptions and TSH reflex testing* in 2015	Pre-reflex TFT prescription frequency in 2012	Reagent cost	€ 2015	Annual savings 11'079	Authors concluded that the reduction in tests resulted in cost-savings.
Feldkamp et al. 1996 [41]	Cost study	USA	A 750 bed hospital and 22 satellite clinics	TSH algorithm testing† in 1993	Pre-algorithm test frequency in 1989	Reagent and labour costs	€ 1993	Annual savings 74'000	The algorithm resulted in cost-savings compared to pre-algorithm average orderings.
Kende et al. 2002 [35]	Cost study	Papua New Guinea	Clinical laboratory of a hospital (3'089 TSH and 2'867 fT4 tests)	TSH only (subsequent fT4 in abnormal TSH) (January 1996 to May 2000)	TSH, fT4	Reagent cost	K (Papua New Guinean Kina). Cost year not reported	Annual savings 4'000	Reagent costs were estimated to be K10.36 for TSH and K11.17 for fT4. Authors conclude TSH testing alone was adequate in 95% of test requests and reduces the cost of TFT reagent by 25% and could save K4000 annually.
Kiel et al. 2020 [39]	Cost Study	Germany	5'552 patients	TSH, fT3/fT4	No clear comparison	TSH, fT3/fT4 test costs per person (p.p.) in diagnosed patients‡	€ based on billing codes 2002-2006	Cost of TSH € 3.00; Cost of fT3/fT4 € 3.70/€ 4.10	Authors find potential overuse of TSH, fT3, and fT4 tests, and suggests diagnostic testing should be used rationally with regards to costs.
Kluesner et al. 2018 [46]	Cost Study	USA	25'142 laboratory tests at military	TSH	TSH + fT4	Department of Defence prices for	US$ 2016§	TSH alone cost 2.02; TSH + fT4	Authors suggests majority of TFT were unnecessary, and assumes potential cost

			hospital			TFT		cost 3.71	savings of $120 000 (within the San Antonio Military Health System).
Nordyke et al. 1998 [38]	Cost study	USA	2'000 patients tested for (suspected) thyroid disease	TSH, and then fT4 if TSH is outside normal range	TSH + fT4	Not reported	US$ 1995	Savings per test 34.24	Authors favoured TSH test, followed by a fT4, if TSH test results are abnormal.
Salinas et al. 2016[42]	Cost study	Spain	370 bed hospital serving a population of 234'551 (Number of tests not reported)	Hypothetical fT4/TSH ratio goal of 0.25‖	fT4/TSH ratio of 0.4	Direct costs, reagents and labour (personnel)	$¶ 2012/14	Savings 9,826#	Reduction of unnecessary tests in hospitalised patients resulted in cost savings.
Salinas et al. 2016 [45]	Cost study	Spain	76 laboratories serving a population of 17'679'195 (Number of tests: TSH 3'268'925, fT4 1'225'306, fT3 78'590)	Hypothetical goal of fT4/TSH ratio of 0.25‖	Mean fT4/TSH ratio of 0.37	Reagent costs	€ 2012	Annual savings 937,261	Potential of cost saving if the hypothesised fT4/TSH ratio was reached.
Salinas et al. 2017 [44]	Cost study	Spain	110 laboratories serving a population of 27'798'262 (Number of tests: TSH 5'923'116, fT4 1'644'350, fT3 141'988)	Hypothetical goal of fT4/TSH ratio of 0.25‖	Not reported	Average price per TSH test (only reagent)	Not reported	Not reported	Significant over-request of thyroid laboratory tests in primary care in Spain, resulting in a high economic impact to society.
Vanderpump et al. 1997 [40]	Cost study	UK	550 samples with second-gen serum TSH	Testing with third gen serum TSH	Two-step testing with second gen serum TSH	Reagent costs	£ 1995	Annual savings 12'500 (rounded)	3-gen TSH better sensitivity reduced the number of fT4 tests, but at a higher reagent cost.

Keys: USA=United States of America, UK=United Kingdom, (f)T3=(free) triiodothyronine, (f)T4=(free) thyroxine, TFT=thyroid function tests, TSH=Thyroid-stimulating hormone. * TSH reflex testing is defined in the study as an automated diagnostic algorithm following the rule of "If…Then…" in which if an TSH test is outside the normal range then it is followed by a fT4, and possibly fT3 [40] † The algorithm is described in the study as starting with TSH testing, followed by a same day free thyroid index (T4 x Resin Uptake Ratio) test performed on the same sample.‡ Patients diagnosed only in the Study of Health In Pomerania (SHIP) assumed to be diagnosed patients, not monitored patients. § 2016 is the study period, no clear description of cost year. ‖ Authors describe the hypothetical fT4/TSH ratio as a suitable fT4 to TSH ratio reported in literature, fT4/TSH ratio of 0.25 and fT3/TSH ratio of 0.1. ¶ Costs are denoted as unit $ in the study, however, it is not specified what cost year nor whether the $ is US$ or some other $. # The study states reducing unnecessary tests in hospitalised patients resulted in savings of 9,826 from January 2012 to December 2014.

7.2.4 Findings costs

Table 10 shows the conclusions of all included studies by the study authors. Whilst the table presents the magnitude of potential savings as presented in the published studies, readers are urged to avoid direct comparisons of the results between studies, given the heterogeneity of interventions, comparators, and study designs. Overall, the costs and budget impact systematic literature results suggest potential for cost-savings. The included studies investigating different test approaches concluded that the two-step test approach was the most cost-saving approach as compared to the one-step test approach. Similarly, the studies conducted on hypothetical goals of the fT4/TSH ratio concluded that the reduction of unnecessary tests were deemed economically favourable.

The included studies did not provide relevant cost data for the Swiss budget impact model. Swiss databases, publicly available sources, and expert opinion were used for the cost inputs in the budget impact model. The unit costs used in the budget impact model are reported in ***Table 11***.

Table 11: Unit costs used in the budget impact model

Cost inputs	Swiss unit cost (CHF), until 31 July 2022	Swiss unit cost (CHF), from 1 August 2022	Swiss description/Source
Price per each thyroid function test • TSH • (f)T4 • (f)T3 • Tax per requested testing order	 TSH: 9CHF fT4/T4: 9CHF fT3/T3: 10.40 CHF 24 CHF	 TSH: 8.10 CHF fT4/T4: 8.10 CHF fT3/T3: 9.40 CHF 21.60 CHF	Analysis list, FOPH
Outpatient visit costs*	91.51 CHF		Expert opinion/TARMED
Follow-up costs†	91.51 CHF		Expert opinion/TARMED

* Outpatient visit costs are calculated as follows. Two-step test approach (1st step): A typical consultation is about 20 min. Depending on whether the blood is taken before or after the consultation there are costs for the laboratory order in advance or costs for a phone call to inform the patient about the results afterwards. The total sum is 102.82 tax points. The tax points are multiplied by a Swiss average tax point value of 0.89. The final cost after tax is 91.51 CHF. Two-step test approach (2nd step): A second test would again require a consultation (by phone or in person) and a blood draw, as well as a report of findings to the patient. The costs are 91.51 CHF.

† In the case of "borderline" thyroid values, a check/follow-up after 3 months can be assumed. Follow-up costs were not included for reflex testing.

7.2.5 Findings budget impact

A budget impact model was developed according to the ISPOR principles of good practice guidelines.[47,48] The analysis was performed from the Swiss healthcare payer perspective. Costs of healthcare services covered by the Swiss mandatory health insurance were analysed, irrespective of the actual payer

(mandatory health insurance, other social insurance, government, out-of-pocket). The analysis did not include indirect costs due to informal care or productivity losses and additional non-medical costs for patients, such as travel costs. The costs were reported in Swiss Francs (CHF). Model input parameters on epidemiology were based on the published literature. [9,49] Clinical expert opinion and real-world data were used whenever published data were not available from the literature. Additional searches on Tarifpool SASIS AG database and analysis list from the FOPH were conducted to determine healthcare resource use, and unit costs. When the required data could not be identified from these sources, assumptions were made by consulting expert opinion.

Epidemiology estimates

Accurate Swiss thyroid disease incidence and prevalence rates were not available. European epidemiological data for this report were extracted from the meta-analysis conducted by Garmendia Madariaga et al. 2014. [9] The reported mean prevalence of thyroid dysfunction in Europe is 3.82%. [9] The incidence rate of thyroid dysfunction was estimated at 259.12 per 100'000 per year (226.2 hypothyroidism and 51 hyperthyroidism, respectively). [9]

Estimation of thyroid function tests in Switzerland

Real-world total number of thyroid function tests (between years 2015-2019) were obtained from Tarifpool SASIS AG database. From year 2015 until 2019, the average number of TSH tests in the specified years were 2'307'026. For fT4/T4 and fT3/T3, the average number of tests were 814'303, and 517'412, respectively (***Appendix 13.3***).

Quantification of TSH tests in Switzerland

The Swiss adult population was estimated at 7'114'731 persons in 2020 (Bundesamt für Statistik, BFS). The incidence rate of 259.12 cases per 100'000 per year was obtained from a published study by Garmendia Madariaga et al. 2014. [9] After applying this incidence rate to the adult Swiss population, there were approximately 18'436 new cases with a thyroid hormone disorder per year. According to Werhun et al. 2015 [50], the positivity rate of all thyroid hormone tests performed in a year in one medical practice in England was 2.1%. At this positivity rate of 2.1%, and by taking 18'436 confirmed new cases into account, 877'905 TSH tests were estimated for the base case (year 2020) in Switzerland. This estimate suggests that 38.1% of all TSH tests performed in Switzerland in year 2020 (out of 2'307'026 TSH tests in total) corresponds to the percentage of TSH tests performed for the identification of suspected thyroid dysfunction, while 61.9% of the tests corresponds to the percentage of TSH tests not performed for the identification of suspected thyroid dysfunction (1'429'121 TSH tests out of 2'307'026 TSH tests in total).

Disorders of thyroid function are frequently diagnosed, with the prevalence rate in Europe varying between 2% and 6%. Taking the mean prevalence of thyroid dysfunction in Europe (3.82%) [9] as a proxy for the prevalence in Switzerland, we assume there were 271’783 already known thyroid dysfunction cases in Switzerland. After the assumption of one test per already known patient with thyroid dysfunction in one year, we estimated a TSH over-testing of 1’157’338 TSH tests (i.e. 1’429’121 TSH tests minus 271’783 TSH tests) which corresponds to approximately 50.2% of all TSH tests performed in Switzerland.

Quantification of T4, fT4, T3, and fT3 thyroid function tests in Switzerland

The average total numbers of TSH tests (2015-2019) was 2'307'026. The average total numbers of T4, fT4, T3 and fT3 tests (2015-2019) was 1'331'715, of which 814’303 (61% of all T4, fT4, T3 and fT3 tests) were T4, fT4 tests and 517’412 (39% of all T4, fT4, T3 and fT3 tests) were T3, fT3 tests. Therefore, approximately 57.7% of the TSH tests were followed by a T4, fT4, T3, or fT3 thyroid function test. As estimated above, 877’905 TSH tests are performed to identify suspected thyroid dysfunction and 271’783 TSH tests are performed per year to test already known thyroid function cases. Hence, there were 1’149’688 TSH tests per year (approximately 49.8% of all TSH tests) that would require subsequent T4, fT4, T3, and fT3 testing versus the 1’331’715 T4, fT4, T3, and fT3 tests that were actually performed. This estimation corresponds to an assumed T4, fT4, T3, or fT3 over-testing of 182’027 tests per year (13.7% of all T4, fT4, T3, and fT3 tests) in Switzerland.

Base case analysis: Swiss budget impact estimates

A budget impact model was developed by accounting the number of thyroid function tests at the population-level. Test costs (9 CHF for TSH; 9 CHF for fT4/T4; 10.4 CHF for fT3/T3; plus 24 CHF tax per each analysis order up to 31 July 2022; 8.10 CHF for TSH; 8.10 CHF for fT4/T4; 9.40 CHF for fT3/T3; plus 21.60 CHF tax per each analysis from 1 August 2022) and outpatient visits (91.51 CHF) were accounted, by the analysis list and by consulting expert opinion. The model was programmed in Microsoft Excel. The average number of tests from years 2015 until 2019 from Tarifpool SASIS AG data procured by the FOPH, was used as a proxy.

Base case results showed that the two-step test approach was cost-saving (6’074’949 CHF) compared to the one-step test approach. The assumption of the two-step test approach, i.e. second blood draw and consultation, was applied in this base case analysis.

After review of the draft version of this HTA report, clinical experts suggested to also include reflex testing (i.e. using the initial blood sample without the need for a second blood draw or consultation) in

the budget impact analysis. Therefore, the impact of implementing reflex testing in the two-step test approach was taken also into account. Reflex testing was the most favourable (cost-saving) compared to the two-step test approach (2'129'542 CHF) and one-step test approach (8'204'491 CHF).

Table 12 shows the base case (year 2020) results of the budget impact model.

Similarly, yearly percentage increases of TSH and fT3/fT4 tests were obtained from Tarifpool SASIS AG data procured by the FOPH (***Appendix 13.3***) and used as a proxy to calculate budget impact figures from year 2021 until year 2024. The budget impact results for year 2021, year 2022, year 2023, and year 2024 were similar and are included in the appendix of this report. (***Appendix 13.4***)

Table 12: Base case (year 2020) results of the Swiss budget impact model

		Thyroid function test		
		One-step test approach		
Estimated number of tests	**Thyroid function tests unit costs**	**One-step test costs**	**Outpatient visits**	**Total Costs**
877'905† **535'522** (=877'905*61%)‡ **342'383** (=877'905*39%)§	TSH: 9 + 24 CHF fT4/T4: 9 CHF fT3/T3: 10.40 CHF	**28'970'865 CHF** (=9+24*877'905) **4'819'698 CHF** (=9*535'522) **3'560'783 CHF** (=10.40*342'383)	**80'337'087 CHF** (=91.51*877'905)	**117'688'433 CHF**
		Two-step test approach		
Estimated number of tests	**Thyroid function tests unit costs**	**Two-step test costs**	**Outpatient/Follow-up visits**	**Total Costs**
877'905† **11'246** (=18'436*61%)‡ **7'190** (=18'436*39%)§	TSH: 9 + 24 CHF fT4/T4: 9 + 24 CHF fT3/T3: 10.40 + 24 CHF	**28'970'865 CHF** (=9+24*877'905) **371'118 CHF** (=9+24*11'246) **247'336 CHF** (=10.40+24*7'190)	**80'337'087 CHF** (=91.51*877'905) **1'029'121 CHF** (=91.51*11'246) **657'957 CHF** (=91.51*7'190)	**111'613'484 CHF**
		Reflex test approach		
Estimated number of tests	**Thyroid function tests unit costs**	**Reflex test costs**	**Outpatient visits**	**Total Costs**
877'905† **11'246** (=18'436*61%)‡ **7'190** (=18'436*39%)§	TSH: 9 +24 CHF fT4/T4: 9 CHF fT3/T3: 10.40 CHF	**28'970'865 CHF** (=9+24*877'905) **101'214 CHF** (=9*11'246) **74'776 CHF** (=10.40*7'190)	**80'337'087 CHF** (=91.51*877'905)	**109'483'942 CHF**
Reflex test and two-step test approaches are cost-saving in Switzerland compared to one-step test approach			**Incremental Costs**	
			Two-step test vs. one-step test approach	**- 6'074'949 CHF**
			Reflex test vs. two-step test approach	**- 2'129'542 CHF**
			Reflex test vs. one-step test approach	**- 8'204'491 CHF**

Keys: CHF=Swiss franc, fT3=free triiodothyronine, fT4=free thyroxine, TSH=thyroid-stimulating hormone, T3=free triiodothyronine, T4=free thyroxine

24 CHF: Swiss tax

†At a positivity rate of 2.1%, to find all 18’436 confirmed cases with a thyroid hormone disorder per year in Switzerland, 877’905 tests would have to be performed. 18’436 cases with a thyroid hormone disorder per year: at an incidence rate of 259.12 cases per 100’000 per year [9] , adult Swiss population (18 years and above) of 7’114’731 million (Bundesamt für Statistik, BFS).

‡ Proportion of fT4/T4 test calculated from average number of tests over 5 years: 814’303/1’331’715 (***Appendix 13.3***)

§ Proportion of fT3/T3 test calculated from average number of tests over 5 years: 517’412/1’331’715 (***Appendix 13.3***)

8 Ethical, legal, social, and organisational issues

Summary statement ethical, legal, social, and organisational issues

Nine observational studies and two guidelines informed the ethical, legal, social, and organisational issues related to the use of thyroid function tests in adults with suspected thyroid dysfunction. Ethical issues associated with thyroid function tests relate to a potential pressure felt by the physician from the patients requesting the complete diagnostic profile and the inherent fear of the physician of missing a diagnosis. Another issue raised in the literature relates to uncertainty regarding normal range thresholds for TSH, fT3, and fT4 tests and the resulting potential for missed diagnoses when applying the two-step test approach. A societal issue identified in the studies was the risk that a change in the national thyroid test approach might discriminate against specific subgroups that are more sensitive/prone to have a misclassification. As implementing a new diagnostic test approach can require alterations in organisation, features like the automation setup of the two-step test approach, physician education, and requirements of justification are discussed.

8.1 Methodology ethical, legal, social, and organisational issues

8.1.1 Databases and search strategy

When conducting the systematic literature searches for clinical outcomes and costs and budget impact, articles discussing issues pertaining to other HTA domains were included from the literature. In addition, grey literature searches were performed on these HTA domains using the following publicly available sources:

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

8.1.2 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.

8.2 Results ethical, legal, social, and organisational issues

8.2.1 PRISMA Flow diagram

Not applicable.

8.2.2 Study characteristics of included studies

The study characteristics of the included studies are qualitatively discussed in the ***Subchapters 8.2.3*** to ***8.2.5***. The search for studies covering ethical, legal, social, and organisational issues relating to consequences of switching the approach of thyroid function testing resulted in nine peer-reviewed observational studies and two European Thyroid Association guidelines. [18,21,43,51–58] The two guidelines and six studies covered issues in the ethical domain, two studies the social domain, and three studies the organisational domain. No studies were identified covering legal issues. Six studies reviewing laboratory thyroid function tests included either data on all patients within a laboratory system or the data on thyroid function tests, with study periods varying from six months to several years. Bauer et al. 1996 [51], Schneider et al. 2018 [52], and Viera et al. 2003 [53] relied on thyroid function test databases gathered in the 1990s, while Asban et al. 2018 [54], Caldarelli et al. 2017 [43], Livingston et al. 2015 [21] were based on thyroid function test data post-2010. The studies were mostly set in the U.S. or European context.

8.2.3 Findings ethical issues

In this section, the ethical issues of applying the two-step test instead of the one-step test approach will be discussed. One issue associated with applying the two-step instead of the one-step test approach relates to the uncertainty around normal range thresholds for TSH, fT3, and fT4 tests. [55] Different studies use different TSH thresholds and ranges. The choice of a TSH threshold that dictates what test outcome will be followed-up by a fT4 or fT3 affects the number of potentially missed diagnoses. For example, a broad normal range for the TSH test in a two-step test approach could miss diagnoses that could have been registered by a fT3 or fT4 test. In such test range grey zones, clinical context, and decision making is important and must be emphasised. [51] Livingston et al. 2015 [21] reiterates the issue in a study looking

into missed cases of secondary hypothyroidism for which the TSH results are considered within normal range. The authors discovered eight cases of pituitary insufficiency that the two-step test approach would have missed, arguing for the one-step test approach. The 2018 European Thyroid Association (ETA) Guidelines on the Diagnosis and Management of Central Hypothyroidism [59] warn for major false negative results applying the two-step test approach. The diagnosis of central hypothyroidism is usually indicated by the combination of low fT4 and low to normal TSH concentrations. The ETA guideline recommends the combination be confirmed on two separate determinations, as well as after several conditions have been excluded to avoid misdiagnoses.

In the guideline for diagnosis and treatment of endogenous subclinical hyperthyroidism, authors Biondi et al. 2015 [56] highlight the issue of TSH assay sensitivity and choice of normal range thresholds. The prevalence of subclinical thyroidism can vary, with an estimated prevalence of 0.7% for a TSH threshold of <0.1 mU/L and 1.8% for a TSH threshold of <0.4 mU/L. Therefore, the choice of TSH assay and their respective thresholds affect the incidence of the diagnosis of thyroid dysfunction.

Yet, Schneider et al. 2018 [52] acknowledge the risk of missing a diagnosis of central hypothyroidism or hyperthyroidism, or impaired sensitivity to thyroid hormone when applying the two-step test approach. In these cases where the two-step test approach cannot capture the diagnoses, clinical signs, and context need to be accounted for in a clinical decision making. [55] Schneider and colleagues further suggest fear of false negative test results may direct physicians to prefer the one-step test approach. Another major argument in the eyes of the physician in favour of the one-step test approach is convenience of simultaneous testing. These arguments may overrule the argument of the additional test costs, according to the authors. [52]

Patient preference is another ethical issue that physicians may face. In a paper by Esfandiari et al. 2019 [57], mismanagement of tests and treatment due to patient requests in the context of hypothyroidism is discussed. The authors argue that physicians experiencing pressure from patients are more likely to honour the patients' request. They further suggest this represents a barrier to optimal patient care in other clinical settings.

In their retrospective, cross-sectional analysis of 4'843 participants Schneider et al. 2018 [52] showed that 93% of the participants had a normal TSH, 0.8% had overt hypothyroidism, 1.9% had subclinical hypothyroidism, 3.8% subclinical hyperthyroidism, and 0.5% overt hyperthyroidism. They argued that 93% of the fT4 testing could be prevented when a two-step instead of a one-step test approach were to be used. The results also showed that 3.8% of all participants had normal TSH levels, while fT4 levels were outside the reference range (false negatives). However, 85% of the potentially missed diagnoses were

within 2 pmol/L of the limits for the fT4 reference range, which the authors considered to be euthyroid, and suggests the patients missed by the two-step test approach would not have a true thyroid dysfunction. Similar statements were made regarding the low likelihood of an abnormal fT4 after a normal TSH test by Bauer et al. 1996 and Viera et al. 2003. [51,53]

8.2.4 Findings legal issues

While no legal issues were identified specifically in the literature searches, the rights of a patient in the case of missing diagnoses due to a change in the test approach may warrant further inspection.

8.2.5 Findings social issues

Physiological factors affect serum TSH concentration levels, as levels rise with age and are particularly increased past the age of 70 years. Further ethnicity is found to play a role, as Jonklaas et al. 2019 [55] reports black individuals to have lower serum TSH concentrations than white individuals. These variations affect subgroups at risk of being misclassified by a two-step test approach, as a uniform reference range risk misclassifying more black than white individuals as having low TSH values, and less as having high TSH values. [55] It follows that a move to two-step test approach as a national thyroid test approach may result in a disproportional distribution of TSH misclassifications due to demographic-related variations within a population. Addressing similar concerns, Koulouri et al. 2013 [58] suggests that age, sex, and pregnancy-specific reference ranges for the thyroid tests – TSH, T4, and T3 – should be developed to address these issues.

8.2.6 Findings organisational issues

Asban et al. 2018 [54] studied underdiagnosed and undertreated hyperthyroidism among 174’011 patients of the University of Alabama at Birmingham in the United States, using serum TSH levels. In 3’336 patients, TSH levels were below 0.005 mU/L, indicating hyperthyroidism. In almost 67% of the patients TSH testing was not followed-up by thyroid hormone testing, indicating a systemic shortcoming in timely diagnosis and treatment according to the authors. The rationale for the missed diagnoses were explained by human and system errors. The authors suggest the initiation of the second step be automatic in nature when the TSH values are outside the reference range in the two-step test approach, i.e. an automatic two-step process in which the suppressed TSH results are automatically flagged, and further tested. If the two-step test approach is implemented with an automatic process as a strategy feature, the type of human and system errors listed by the authors are less likely to occur. [54]

When implementing a new strategy, organisational issues may occur. Caldarelli et al. 2017 [43] investigated the introduction, and implementation of the two-step test approach, called the “TSH reflex”. While the results of the study showed the TSH reflex improving the efficiency and appropriateness of prescriptions, the study also points to hurdles of implementation. Of particular interest, the authors state the implementation required commitment of laboratory personnel and education of physicians. Schneider et al. 2018 [52] echoes the same sentiment and refers to a French study, in which physicians were required to justify all laboratory tests to increase awareness of thyroid function testing, resulting in a simultaneous decrease in TSH and fT4 testing from 77% to 51%.

9 Additional issues

9.1 Guideline Recommendations

Most guidelines with recommendations on thyroid function tests were aimed at establishing a diagnosis of thyroid disorder. The NICE guideline 145 “Thyroid disease: assessment and management” [23] provided recommendations specifically on which thyroid function tests should be used in a population suspected of thyroid dysfunction. NICE did not find clinical evidence to underpin the recommendations on thyroid function tests and the committee used their experience to develop the recommendations (i.e. expert opinion based recommendations).

Citing the recommendation of the guideline: *TSH alone was in general an appropriate first test for people in whom thyroid dysfunction is suspected. Subsequent tests (cascading) are only needed if TSH is abnormal (with fT4 if the TSH suggests hypothyroidism and both fT4 and fT3 if the TSH suggests hyperthyroidism). This approach reduces unnecessary testing compared with simultaneous TSH, fT4 and fT3 testing for all people. However, tests should be done in a way to minimise potential delays and the need for additional appointments, for example, by laboratories keeping original samples and performing subsequent tests on the same samples. The committee agreed, based on their experience that this approach did not apply to adults in whom secondary thyroid dysfunction is suspected where both TSH and fT4 are needed by default because of the differing likely causes of dysfunction. The committee further agreed that tests may need repeating when new symptoms develop or worsen, but that this should not be within six weeks of the last test because this is unlikely to provide new information.*

In this NICE guideline, they also discuss how the recommendation might affect practice: *The recommendations broadly reflect current practice, although not all laboratories currently follow the cascading*

approach to testing. Where fT4 is currently a routine test for thyroid dysfunction, cascading will reduce National Health Service (NHS) costs by avoiding extra tests for people with a TSH within the reference range. In areas where fT3 is not currently being measured, cascading will mean a cost increase. But this will be offset by the benefits of correctly diagnosing and managing thyrotoxicosis.

Two additional guidelines provided information on how to establish the diagnosis of central hypothyroidism and subclinical hyperthyroidism, respectively. In the ETA Guidelines on the Diagnosis and Management of Central Hypothyroidism, it was recommended to use the combined determination of serum fT4 and TSH in order to evaluate the presence of central hypothyroidism. [59] To establish the diagnosis of persistent subclinical hyperthyroidism, the ETA recommended in the Guidelines on Diagnosis and Treatment of Endogenous Subclinical Hyperthyroidism that serum TSH should be measured as an initial screening test. ETA stated that if serum TSH is low, thyroid hormones (fT4 and T3 or fT3) should be measured. [56]

No recommendations on the one-step or two-step test approach in a population suspected of thyroid dysfunction were found in other leading guidelines.

9.2 Ongoing Trials

A search on clinicaltrial.gov did not reveal any ongoing trials that investigated our research question.

10 Discussion

The clinical systematic literature search for studies informing on the number of missed cases and the diagnostic subcategory distribution of these missed cases in adults with suspected primary or secondary thyroid dysfunction when applying a two-step test approach instead of a one-step test approach retrieved three case series conducted in Asia, all with a high risk of bias. Although the number of missed cases is likely to be very low, the studies did not provide sufficient evidence to define a number of missed cases when applying the two-step test approach. According to the study authors of the one study reporting on missed diagnosis, the most likely missed diagnoses included patients with known thyroid disease on treatment, sick-euthyroid syndrome, or patients using drugs of which it is known these interfere with thyroid hormone level. [35]

Cases missed when the two-step test approach is applied have to live longer with their unexplained symptoms. Studies on missed diagnosis report symptoms on a variety of body systems like slowed

metabolism, tiredness, and weight gain for those with central hypothyroidism and more energy weight loss and anxiousness for those with missed diagnosis of central hyperthyroidism. The symptoms of central hyperthyroidism or hypothyroidism are usually milder than the symptoms observed in cases with primary hyperthyroidism or hypothyroidism. [60] Correct diagnosis and subsequent treatment led to resolution of numerous symptoms and biochemical abnormalities. [24,61]

The search for scientific evidence to inform the research questions was highly challenging in this HTA report. Two search strategies were developed to inform research question 1 and 2 respectively. A rigorous methodology, adhering to international methodological standards such as Cochrane and PRISMA, was applied to identify, critically appraise, analyse, and summarise pertinent evidence on predefined outcomes of interest in order to minimise bias. Despite a broad systematic literature search for clinical evidence only a very limited number of studies could be included, because only few studies were conducted specifically in clinically suspected populations or defined their study population in sufficient detail. In addition, very few studies reported data that could be categorised in one of the nine categories. Data based on populations screened or monitored for thyroid disorders will not be applicable directly to populations suspected of thyroid disorders. Screening is the presumptive identification of unrecognised disease with tests in people who do not have the symptoms of early disease and testing for monitoring or follow-up of thyroid dysfunction is applied in patients already diagnosed with thyroid disorders.

Only one included case series reported the number of discordant test results, reflecting missed diagnoses when a two-step test approach would have been applied. It is unclear whether these results are representable for Switzerland, given potential demographic variations in thyroid disorder subcategory distributions. Moreover, this study reported blood sample results instead of individual patient results (i.e. the number of blood samples was not identical to the number of patients, because of inclusion of data from patients measured on multiple occasions) and no patient characteristics were reported. The applied reference ranges for TSH level, (f)T3, and (f)T4 were reported in two of the three case series and differed between the two case series.

Several study authors and Swiss clinical experts stressed that test results alone will never inform a final diagnosis of thyroid disorder. Additional clinical findings and patient characteristics like age, gender, and ethnicity all have to be considered. Although the clinical literature assessment results in this HTA report are not sufficient to define a definitive number of missed cases when applying a one-step instead of a two-step thyroid function test approach, the findings indicate that this number is likely to be low, while the costs of preventing these cases are high. The systematic literature search for studies on costs and budget impact studies related to thyroid function testing in adults with suspected primary or secondary thyroid dysfunction, retrieved ten cost studies. All studies shared a consistent opinion on thyroid function

tests likely being unnecessary used or overused, supporting the potential for cost-savings if a step-wise approach is adopted. The included studies investigating different test approaches concluded the two-step test approach the most cost-saving approach as compared to the one-step test approach. The studies conducted on hypothetical goals of the fT4/TSH ratio concluded that the reduction of unnecessary tests was deemed economically favourable. As these studies did not provide relevant cost data for constructing a Swiss budget impact model, Swiss databases, publicly available sources, and expert opinion were sourced for the inputs of the budget impact model.

A budget impact model, over a time horizon of five years (2020-2024), was developed to assess the potential for cost-savings in Switzerland by accounting the number of thyroid function tests at the population-level. Compared to the one-step thyroid function test approach the two-step test approach (i.e. including second blood draw) was cost-saving (approximately 6.1 million CHF in year 2020). The impact of implementing reflex testing (i.e. using the initial blood sample without the need for a second blood draw) in the two-step test approach was also taken into account. Reflex testing was considered the most favourable (cost-saving) test approach compared to the two-step test approach (approximately 2.1 million CHF in year 2020) and one-step test approach (approximately 8.2 million CHF in year 2020), respectively. The results for year 2021, year 2022, year 2023 and year 2024 were similar. The cost difference between the two-step test approaches is caused by the Swiss tax paid for each blood draw.

There were several assumptions that were influential in the budget impact model. The average number of tests in Switzerland from years 2015 until 2019 was used as a proxy. Similarly, yearly percentage increases of TSH and fT3/fT4 tests were used as a proxy to calculate the year 2021 until year 2024 budget figures. Outpatient visits and test calculations were informed by expert opinion.

Other HTA domains (ethical, legal, social, and organisational) were researched to identify issues related to the approach of thyroid function tests in adults with suspected thyroid dysfunction and resulted in the inclusion of nine studies and two guidelines. Prominent issues identified included physicians inherent fear of missing diagnoses as well as pressure from patients requesting the complete diagnostic profile, and that a change in the national thyroid test approach might risk missing specific diagnostic subgroups that are more prone to have a false negative diagnosis. Further organisational issues were identified relating to education of physicians as well as the specific features required for a successful implementation of a two-step test approach.

11 Conclusions

The clinical evidence was sparse, and the low quality does not permit to draw conclusions on a concrete number of cases missed when the two-step instead of the one-step test approach is applied in adults with suspected primary or secondary thyroid dysfunction in the Swiss setting. The budget impact analysis showed that the two-step test approach can be considered cost-saving compared to the one-step test approach. From a clinical perspective the two-step test approach seems to be justified, given the likely low risk of missed cases when applying this approach. The impact of reflex testing was also assessed and was more cost-saving compared to the one-step and two-step test approach, since the initial blood sample is used for possible additional thyroid function tests without the need for a second blood draw. When applying a two-step or reflex test approach instead of the one-step test approach, the possible impact of missing or misclassifying diagnoses and practical aspects of how to successfully implement the reflex testing approach should be considered.

12 References

1 How Your Thyroid Works. https://www.endocrineweb.com/conditions/thyroid/how-your-thyroid-works (accessed 20 Dec 2021).

2 Thyroid Gland: Overview. https://www.endocrineweb.com/conditions/thyroid-nodules/thyroid-gland-controls-bodys-metabolism-how-it-works-symptoms-hyperthyroi (accessed 20 Dec 2021).

3 Pirahanchi Y, Toro F, Jialal I. Physiology, Thyroid Stimulating Hormone. In: *StatPearls*. Treasure Island (FL): : StatPearls Publishing 2021. http://www.ncbi.nlm.nih.gov/books/NBK499850/ (accessed 10 Jan 2022).

4 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).

5 European Thyroid Association (ETA). https://www.eurothyroid.com (accessed 19 Dec 2021).

6 American Thyroid Association (ATA). https://www.thyroid.org (accessed 19 Dec 2021).

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

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

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

10 Wouters HJCM, Slagter SN, Kobold ACM, *et al.* Epidemiology of thyroid disorders in the Lifelines Cohort Study (the Netherlands). *PLOS ONE* 2020;**15**:e0242795. doi:10.1371/journal.pone.0242795

11 British Thyroid Foundation. Older patients and thyroid disease. . https://www.btf-thyroid.org/older-patients-and-thyroid-disease (accessed 19 Dec 2021).

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

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

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

15 Nygaard B. Hypothyroidism (primary). *BMJ Clin Evid* 2014;**2014**:0605.

16 Hammett-Stabler CA, Maygarden SJ. Pathology of the Endocrine System. In: Reisner HM, ed. *Pathology: A Modern Case Study*. New York, NY: : McGraw-Hill Education 2015. accessmedicine.mhmedical.com/content.aspx?aid=1115281093 (accessed 20 Dec 2021).

17 McDermott MT, Ridgway EC. CENTRAL HYPERTHYROIDISM. *Endocrinol Metab Clin North Am* 1998;**27**:187–203. doi:10.1016/S0889-8529(05)70306-6

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

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

20 Donangelo I, Suh SY. Subclinical Hyperthyroidism: When to Consider Treatment. *Am Fam Physician* 2017;**95**:710–6.

21 Livingston M, Twomey PJ, Basu A, *et al.* Should Free Thyroxine Go Back into the Routine Thyroid Profile? *Exp Clin Endocrinol Diabetes Off J Ger Soc Endocrinol Ger Diabetes Assoc* 2015;**123**:594–7. doi:10.1055/s-0035-1559782

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

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

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

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

26 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

27 National Institute for Health and Care Excellence. NICE guideline. Thyroid disease: assessment and management. UK: 2019.

28 Plebani M, Giovanella L. Reflex TSH strategy: the good, the bad and the ugly. *Clin Chem Lab Med CCLM* 2019;**58**:1–2. doi:10.1515/cclm-2019-0625

29 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

30 Van Houcke SK, Van Uytfanghe K, Shimizu E, *et al.* IFCC international conventional reference procedure for the measurement of free thyroxine in serum: International Federation of Clinical Chemistry and Laboratory Medicine (IFCC) Working Group for Standardization of Thyroid Function Tests (WG-STFT) [1]. *Clin Chem Lab Med CCLM* 2011;**49**:1275–81. doi:10.1515/CCLM.2011.639

31 Higgins J, Thomas J, Chandler J, *et al.* Cochrane Handbook for Systematic Reviews of Interventions version 6.2 (updated February 2021). Cochrane, 2021. www.training.cochrane.org/handbook

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

33 Munn Z, Barker TH, Moola S, *et al.* Methodological quality of case series studies: an introduction to the JBI critical appraisal tool. *JBI Evid Synth* 2020;**18**:2127–33. doi:10.11124/JBISRIR-D-19-00099

34 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

35 Kende M, Kandapu S. Evaluation of thyroid stimulating hormone (TSH) alone as a first-line thyroid function test (TFT) in Papua New Guinea. *P N G Med J* 2002;**45**:197–9.

36 Bansal A, Kaushik A, Sarathe H. Effect of thyroid on lipid profile and renal function: an observational study from tertiary care centre of tribal region of Bastar. *Ann Med Health Sci Res* 2014;**4**:140. doi:10.4103/2141-9248.138035

37 Natasha, Badiger R. A Prospective Study of Thyroid Function Test in Geriatric Population and its Clinical Correlation. *J Assoc Physicians India* 2019;**67**:33–6.

38 Nordyke RA, Reppun TS, Madanay LD, *et al.* Alternative Sequences of Thyrotropin and Free Thyroxine Assays for Routine Thyroid Function Testing: Quality and Cost. *Arch Intern Med* 1998;**158**:266. doi:10.1001/archinte.158.3.266

39 Kiel S, Ittermann T, Völzke H, *et al.* Frequency of thyroid function tests and examinations in participants of a population-based study. *BMC Health Serv Res* 2020;**20**:70. doi:10.1186/s12913-020-4910-7

40 Vanderpump MPJ, Neary RH, Manning K, *et al.* Does an Increase in the Sensitivity of Serum Thyrotropin Assays Reduce Diagnostic Costs for Thyroid Disease in the Community? *J R Soc Med* 1997;**90**:547–50. doi:10.1177/014107689709001006

41 Feldkamp CS, Carey JL. An Algorithmic Approach to Thyroid Function Testing in a Managed Care Setting: *3-Year Experience*. *Am J Clin Pathol* 1996;**105**:11–6. doi:10.1093/ajcp/105.1.11

42 Salinas M, López-Garrigós M, Flores E, *et al.* Managing Inappropriate Requests of Laboratory Tests: From Detection to Monitoring. 2016;:8.

43 Caldarelli G, Troiano G, Rosadini D. Adoption of TSH Reflex algorithm in an Italian clinical laboratory. *Ann Ig Med Prev E Comunità* 2017;:317–22. doi:10.7416/ai.2017.2158

44 Salinas M, Lopez-Garrigos M, Flores E, *et al.* Uncritical Request of Thyroid Laboratory Tests May Result in a Major Societal Economic Burden: Results from a Large Population Study in Spain. *Clin Lab* 2017;**63**. doi:10.7754/Clin.Lab.2017.170101

45 Salinas M, López-Garrigós M, Pomares FJ, *et al.* Request of thyroid function tests from Primary Care in Spain. *Endocrinol Nutr Engl Ed* 2016;**63**:19–26. doi:10.1016/j.endoen.2016.01.001

46 Kluesner JK, Beckman DJ, Tate JM, *et al.* Analysis of current thyroid function test ordering practices. *J Eval Clin Pract* 2018;**24**:347–52. doi:10.1111/jep.12846

47 Mauskopf JA, Sullivan SD, Annemans L, *et al.* Principles of Good Practice for Budget Impact Analysis: Report of the ISPOR Task Force on Good Research Practices—Budget Impact Analysis. *Value Health* 2007;**10**:336–47. doi:10.1111/j.1524-4733.2007.00187.x

48 Weinstein MC, O'Brien B, Hornberger J, *et al.* Principles of Good Practice for Decision Analytic Modeling in Health-Care Evaluation: Report of the ISPOR Task Force on Good Research Practices—Modeling Studies. *Value Health* 2003;**6**:9–17. doi:10.1046/j.1524-4733.2003.00234.x

49 Taylor PN, Albrecht D, Scholz A, *et al.* Global epidemiology of hyperthyroidism and hypothyroidism. *Nat Rev Endocrinol* 2018;**14**:301–16. doi:10.1038/nrendo.2018.18

50 Werhun A, Hamilton W. Thyroid function testing in primary care: overused and under-evidenced? A study examining which clinical features correspond to an abnormal thyroid function result. *Fam Pract* 2015;**32**:187–91. doi:10.1093/fampra/cmv010

51 Bauer DC. Sensitive thyrotropin and free thyroxine testing in outpatients. Are both necessary? *Arch Intern Med* 1996;**156**:2333–7. doi:10.1001/archinte.156.20.2333

52 Schneider C, Feller M, Bauer DC, *et al.* Initial evaluation of thyroid dysfunction - Are simultaneous TSH and fT4 tests necessary? *PLOS ONE* 2018;**13**:e0196631. doi:10.1371/journal.pone.0196631

53 Viera AJ. Thyroid Function Testing in Outpatients: Are Both Sensitive Thyrotropin(sTSH) and Free Thyroxine(FT4) Necessary? *Fam Med* 2003;:3.

54 Asban A, Chung SK, Tresler MA, *et al.* Hyperthyroidism is Underdiagnosed and Undertreated in 3336 Patients: An Opportunity for Improvement and Intervention. *Ann Surg* 2018;**268**:506–12. doi:10.1097/SLA.0000000000002922

55 Jonklaas J, Razvi S. Reference intervals in the diagnosis of thyroid dysfunction: treating patients not numbers. *Lancet Diabetes Endocrinol* 2019;**7**:473–83. doi:10.1016/S2213-8587(18)30371-1

56 Biondi B, Bartalena L, Cooper DS, *et al.* The 2015 European Thyroid Association Guidelines on Diagnosis and Treatment of Endogenous Subclinical Hyperthyroidism. *Eur Thyroid J* 2015;**4**:149–63. doi:10.1159/000438750

57 Esfandiari NH, Reyes-Gastelum D, Hawley ST, *et al.* Patient Requests for Tests and Treatments Impact Physician Management of Hypothyroidism. *Thyroid* 2019;**29**:1536–44. doi:10.1089/thy.2019.0383

58 Koulouri O, Moran C, Halsall D, *et al.* Pitfalls in the measurement and interpretation of thyroid function tests. *Best Pract Res Clin Endocrinol Metab* 2013;**27**:745–62. doi:10.1016/j.beem.2013.10.003

59 Persani L, Brabant G, Dattani M, *et al.* 2018 European Thyroid Association (ETA) Guidelines on the Diagnosis and Management of Central Hypothyroidism. *Eur Thyroid J* 2018;**7**:225–37. doi:10.1159/000491388

60 Beck-Peccoz P, Rodari G, Giavoli C, *et al.* Central hypothyroidism — a neglected thyroid disorder. *Nat Rev Endocrinol* 2017;**13**:588–98. doi:10.1038/nrendo.2017.47

61 Glyn T, Harris B, Allen K. Lessons learnt from a case of missed central hypothyroidism. *Endocrinol Diabetes Metab Case Rep* 2017;**2017**. doi:10.1530/EDM-17-0112

13 Appendices

13.1 Systematic review clinical outcomes

13.1.1 Search strategy for clinical outcomes systematic review

Table 1: Search strategy for the clinical systematic literature search: PubMed (MEDLINE)

Population	No search string
Intervention: thyroid function test	("Thyroid function tests"[Mesh] OR test*[tiab] OR assay*[tiab]) AND (("Thyrotropin"[Mesh] OR thyrotropin[tiab] OR thyroid stimulating hormone*[tiab] OR thyroid-stimulating hormone*[tiab] OR TSH[tiab]) AND ("Thyroxine"[Mesh] OR thyroxine[tiab] OR T4[tiab] OR freeT4[tiab] OR free-T4[tiab] OR FT4[tiab] OR "Triiodothyronine"[Mesh] OR Triiodothyronine[tiab] OR T3[tiab] OR freeT3[tiab] OR free-T3[tiab] OR FT3[tiab]))
Comparator	No search string
Outcomes	No search string
Limits	*No animal studies:* NOT (Animals[Mesh] NOT (Humans[Mesh] AND Animals[Mesh]))
	No case reports, irrelevant publication types, congress abstracts: NOT (case reports[pt] OR editorial[pt] OR letter[pt] OR news[pt] OR comment[pt] OR congress[pt])
	Publication period: 1990 to 9 September 2021
	Language: English, German, French, Italian

Table 2: Search strategy for the clinical systematic literature search: Embase.com

Population	No search string
Intervention: thyroid function test	('thyroid function test'/exp OR test*:ti,ab OR assay*:ti,ab) AND (('thyrotropin'/exp OR thyrotropin:ti,ab OR "thyroid stimulating hormone*":ti,ab OR "thyroid-stimulating hormone*":ti,ab OR TSH:ti,ab) AND ('thyroxine'/exp OR thyroxine:ti,ab OR T4:ti,ab OR freeT4:ti,ab OR free-T4:ti,ab OR FT4:ti,ab OR Triiodothyronine:ti,ab OR T3:ti,ab OR freeT3:ti,ab OR free-T3:ti,ab OR FT3:ti,ab))
Comparator	No search string
Outcomes	No search string
Limits	*No animal studies:* NOT ('animal'/exp OR 'nonhuman'/exp NOT ('animal'/exp OR 'nonhuman'/exp AND 'human'/exp))
	Relevant publication types: [article]/lim OR [article in press]/lim OR [review]/lim
	Publication period: 1990 to 9 September 2021
	Language: English, German, French, Italian

13.1.2 Excluded studies during full-text selection

Table 3: Excluded studies found in the clinical outcomes systematic literature search

Reference	Reason for exclusion
Adlan MA, Neel V, Lakra SS, et al. Targeted thyroid testing in acute illness: Achieving success through audit. Journal of Endocrinological Investigation 2011;34(8 SUPPL.):e210-e13. doi: 10.3275/7480	Study population out of scope (i.e. no adults with suspected primary or secondary thyroid dysfunction)
Afrose R, Ahasan HN, Das A, et al. A clinical study and short term outcome of the patients presented with thyroiditis. Journal of Medicine (Bangladesh) 2013;14(1):23-27. doi: 10.3329/jom.v14i1.14532	Study population out of scope (i.e. no adults with suspected primary or secondary thyroid dysfunction)
Al-Rabia MW. Correlation of thyroid antibodies with TSH, T3 and T4 hormones in patients diagnosed with autoimmune thyroid disorders. Pakistan journal of pharmaceutical sciences 2017;30(2):607-12.	Study population out of scope (i.e. no adults with suspected primary or secondary thyroid dysfunction)
Aryal M, Gyawali P, Rajbhandari N, et al. A prevalence of thyroid dysfunction in Kathmandu University Hospital, Nepal. Biomedical Research 2010;21(4):411-15.	Indication for thyroid testing not reported (i.e. not clear whether the study population is tested for suspected primary or secondary thyroid dysfunction)
Asmelash D, Tesfa K, Biadgo B. Thyroid Dysfunction and Cytological Patterns among Patients Requested for Thyroid Function Test in an Endemic Goiter Area of Gondar, North West Ethiopia. International Journal of Endocrinology 2019;2019 doi: 10.1155/2019/9106767	Definitions of thyroid dysfunction diagnosis not reported
Azam HUD, Hayat Z, Fida Z, et al. Subclinical hypothyroidism in patients with non-specific symptoms. Journal of Medical Sciences 2010;18(4):191-93.	No data on outcome of interest
Azim W, Shafi M, Farooq W, et al. The need and rationale of thyroid function tests in a tertiary care laboratory. Journal of Postgraduate Medical Institute 2011;25(4):343-46.	Only two-step test approach or combination of one-step/two-step test approach, data not stratified
Bauer DC, Brown AN. Sensitive thyrotropin and free thyroxine testing in outpatients: Are both necessary? Archives of Internal Medicine 1996;156(20):2333-37. doi: 10.1001/archinte.156.20.2333	Study population out of scope (i.e. no adults with suspected primary or secondary thyroid dysfunction)
Boon-Falleur L, Sokal E, Nightingale PG, et al. Utilization of laboratory resources: Developments in knowledge-based ordering systems. International Journal of Bio-Medical Computing 1995;40(1):17-30. doi: 10.1016/0020-7101(95)01118-X	Only two-step test approach or combination of one-step/two-step test approach, data not stratified
Bould H, Panicker V, Kessler D, et al. Investigation of thyroid dysfunction is more likely in patients with high psychological morbidity. Family Practice 2012;29(2):163-67. doi: 10.1093/fampra/cmr059	Indication for thyroid testing not reported (i.e. not clear whether the study population is tested for suspected primary or secondary thyroid dysfunction)
Caldarelli G, Troiano G, Rosadini D, et al. Adoption of TSH Reflex algorithm in an Italian clinical laboratory. Annali di igiene : medicina preventiva e di comunita 2017;29(4):317-22. doi: 10.7416/ai.2017.2158	No data on outcome of interest
Dalal S, Bhesania S, Silber S, et al. Use of Electronic Clinical Decision Support and Hard Stops to Decrease Unnecessary Thyroid Function Testing. BMJ Qual Improv Rep 2017;6(1) doi: 10.1136/bmjquality.u223041.w8346 [published Online First:	No data on outcome of interest

2017/05/05]	
Davey RX, Clarke MI, Webster AR. Thyroid function testing based on assay of thyroid-stimulating hormone: Assessing an algorithm's reliability. Medical Journal of Australia 1996;164(6):329-32. doi: 10.5694/j.1326-5377.1996.tb122046.x	Study population out of scope (i.e. no adults with suspected primary or secondary thyroid dysfunction)
Dogra P, Paudel R, Panthi S, et al. Low yield of thyroid-function tests in adult hospitalized patients — A retrospective analysis. International Journal of General Medicine 2020;13:343-49. doi: 10.2147/IJGM.S256868	No data on outcome of interest
Feldkamp CS, Carey JL. An algorithmic approach to thyroid function testing in a managed care setting: 3-Year experience. American Journal of Clinical Pathology 1996;105(1):11-16. doi: 10.1093/ajcp/105.1.11	Indication for thyroid testing not reported (i.e. not clear whether the study population is tested for suspected primary or secondary thyroid dysfunction)
Fitzgerald SP, Bean NG, Hennessey JV, et al. Thyroid testing paradigm switch from thyrotropin to thyroid hormones—Future directions and opportunities in clinical medicine and research. Endocrine 2021 doi: 10.1007/s12020-021-02851-6	Narrative review
Galia AM, Andag-Silva AA, Kho SA, et al. Validation of the UST thyroid Scoring index against ultrasensitive assays for thyroid-stimulating hormone and free thyroxine. Phillippine Journal of Internal Medicine 2010;48(1):15-23.	No data on outcome of interest
Ghoraishian SM, Moghaddam SHH, Afkhami-Ardekani M. Relationship between anti-thyroid peroxidase antibody and thyroid function test. Iranian Journal of Immunology 2006;3(3):146-49.	Definitions of thyroid dysfunction diagnosis not reported
Gibbons V, Lillis S, Conaglen JV, et al. Do general practitioners use thyroid stimulating hormone assay for opportunistic screening? New Zealand Medical Journal 2009;122(1301):25-30.	Study population out of scope (i.e. no adults with suspected primary or secondary thyroid dysfunction)
Gilmour JA, Weisman A, Orlov S, et al. Promoting resource stewardship: Reducing inappropriate free thyroid hormone testing. Journal of Evaluation in Clinical Practice 2017;23(3):670-75. doi: 10.1111/jep.12698	No data on outcome of interest
Hamburger JI, Kaplan MM. Diagnosis of thyroid dysfunction in ambulatory patients: primacy of the supersensitive thyroid-stimulating hormone assay. Compr Ther 1990;16(7):3-7. [published Online First: 1990/07/01]	Narrative review
Kasonga F, Cassinari K, Brunel V, et al. La juste prescription du bilan biologique thyroïdien dans le cadre du diagnostic d'une dysthyroïdie Étude rétrospective au CHU de Rouen. Annales de biologie clinique 2018;76(4):421-28. doi: 10.1684/abc.2018.1359	No data on outcome of interest
Kluesner JK, Beckman DJ, Tate JM, et al. Analysis of current thyroid function test ordering practices. Journal of Evaluation in Clinical Practice 2018;24(2):347-52. doi: 10.1111/jep.12846	No data on outcome of interest
Livingston M, Twomey PJ, Basu A, et al. Should Free Thyroxine Go Back into the Routine Thyroid Profile? Experimental and Clinical Endocrinology and Diabetes 2015;123(10):594-97. doi: 10.1055/s-0035-1559782	Indication for thyroid testing not reported (i.e. not clear whether the study population is tested for suspected primary or secondary thyroid dysfunction)
Malvano R, Ferdeghini M, Chiecchio A, et al. Assessing diagnostic performance of laboratory tests: The assays of thyroid hormones. Giornale Italiano di Chimica Clinica 1992;17(6):433-39.	No data on outcome of interest

Mindemark M, Larsson A. Long-term effects of an education programme on the optimal use of clinical chemistry testing in primary health care. Scandinavian Journal of Clinical and Laboratory Investigation 2009;69(4):481-86. doi: 10.1080/00365510902749123	No data on outcome of interest
Nordyke RA, Reppun TS, Madanay LD, et al. Alternative sequences of thyrotropin and free thyroxine assays for routine thyroid function testing. Archives of Internal Medicine 1998;158(3):266-72. doi: 10.1001/archinte.158.3.266	Data not stratified for the population with suspected thyroid dysfunction
Notas G, Kampa M, Malliaraki N, et al. Implementation of thyroid function tests algorithms by clinical laboratories: A four-year experience of good clinical and diagnostic practice in a tertiary hospital in Greece. European Journal of Internal Medicine 2018;54:81-86. doi: 10.1016/j.ejim.2018.03.012	No data on outcome of interest
Preiss D, Todd L, Panarelli M. Diagnosing unsuspected hypopituitarism in adults from suggestive thyroid function test results. Annals of Clinical Biochemistry 2008;45(1):70-75. doi: 10.1258/acb.2007.007100	No data on outcome of interest
Robles-Osorio ML, Zacarías-Rangel V, García-Solís P, et al. Prevalence of thyroid function test abnormalities and anti-thyroid antibodies an open population in central México. Revista de Investigacion Clinica 2014;66(2):113-20.	Study population out of scope (i.e. no adults with suspected primary or secondary thyroid dysfunction)
Rohil V, Mishra AK, Shrewastwa MK, et al. Subclinical hypothyroidism in eastern Nepal: A hospital based study. Kathmandu University Medical Journal 2010;8(30):231-37. doi: 10.3126/kumj.v8i2.3565	No data on outcome of interest
Roti E, Gardini E, Magotti MG, et al. Are thyroid function tests too frequently and inappropriately requested? Journal of Endocrinological Investigation 1999;22(3):184-90. doi: 10.1007/BF03343539	Indication for thyroid testing not reported (i.e. not clear whether the study population is tested for suspected primary or secondary thyroid dysfunction)
Sahu S, John J. Investigating Investigation (Ab)Use: Thyroid Function Test Audit in a Tertiary Care Teaching Institute in Eastern India. Indian Journal of Clinical Biochemistry 2021 doi: 10.1007/s12291-020-00955-w	No data on outcome of interest
Schneider C, Feller M, Bauer DC, et al. Initial evaluation of thyroid dysfunction - Are simultaneous TSH and fT4 tests necessary? PLoS ONE 2018;13(4) doi: 10.1371/journal.pone.0196631	Study population out of scope (i.e. no adults with suspected primary or secondary thyroid dysfunction)
Sezgin Y, Akgün AE. The evaluation of thyroid function tests in patients presenting to the family medicine clinics. Journal of Clinical and Analytical Medicine 2018;9(5):439-41. doi: 10.4328/JCAM.5812	Indication for thyroid testing not reported (i.e. not clear whether the study population is tested for suspected primary or secondary thyroid dysfunction)
Sharma NR. Thyroid dysfunction in suspected population of Kangra Valley in Himachal Pradesh, India. Biomedical and Pharmacology Journal 2013;6(2):415-19. doi: 10.13005/bpj/436	No data on outcome of interest
Sitkiewicz D, Pulawska MF, Janas J, et al. Comparative evaluation of commonly used in vitro tests in screening of thyroid function. Medical Science Monitor 1997;3(4):594-98.	Definitions of thyroid dysfunction diagnosis not reported
Snabboon T, Sridama V, Sunthornyothin S, et al. A more appropriate algorithm of thyroid function test in diagnosis of hyperthyroidism for Thai patients. Journal of the Medical Association of Thailand 2004;87(SUPPL. 2):S19-S21.	Indication for thyroid testing not reported (i.e. not clear whether the study population is tested for suspected primary or secondary thyroid dysfunction)

Taher J, Beriault DR, Yip D, et al. Reducing free thyroid hormone testing through multiple Plan-Do-Study-Act cycles. Clinical Biochemistry 2020;81:41-46. doi: 10.1016/j.clinbiochem.2020.05.004	Indication for thyroid testing not reported (i.e. not clear whether the study population is tested for suspected primary or secondary thyroid dysfunction)
Targher G, Chonchol M, Zoppini G, et al. Prevalence of thyroid autoimmunity and subclinical hypothyroidism in persons with chronic kidney disease not requiring chronic dialysis. Clinical Chemistry and Laboratory Medicine 2009;47(11):1367-71. doi: 10.1515/CCLM.2009.304	No data on outcome of interest
Tilahun K, Demissie M, Bekele T, et al. Thyroid Hormone Tests Ordering Practice and Cost-Effectiveness in Samples Referred to International Clinical Laboratories from Addis Ababa Health Facilities. Ethiopian journal of health sciences 2020;30(3):347-54. doi: 10.4314/ejhs.v30i3.5	Indication for thyroid testing not reported (i.e. not clear whether the study population is tested for suspected primary or secondary thyroid dysfunction)
Tomlin A, Dovey S, Gauld R, et al. Better use of primary care laboratory services following interventions to 'market' clinical guidelines in New Zealand: A controlled before-and-after study. BMJ Quality and Safety 2011;20(3):282-90. doi: 10.1136/bmjqs.2010.048124	No data on outcome of interest
Toubert ME, Chevret S, Cassinat B, et al. From guidelines to hospital practice: Reducing inappropriate ordering of thyroid hormone and antibody tests. European Journal of Endocrinology 2000;142(6):605-10. doi: 10.1530/eje.0.1420605	No data on outcome of interest
Van Veggel KM, Rondeel JM, Anten S. Occurrence and management of an aberrant free T4 in combination with a normal TSH. Netherlands Journal of Medicine 2018;76(7):314-21.	Study population out of scope (i.e. no adults with suspected primary or secondary thyroid dysfunction)
Van Walraven C, Goel V, Chan B. Effect of population-based interventions on laboratory utilization: A time-series analysis. Journal of the American Medical Association 1998;280(23):2028-33. doi: 10.1001/jama.280.23.2028	No data on outcome of interest
Vidal-Trécan G, Toubert ME, Coste J, et al. Reducing the number of T3 orders in the Paris hospital network: Towards better appropriatness of thyroid function test prescription. Annales d'Endocrinologie 2003;64(3):210-15.	No data on outcome of interest
Vidal-Trécan G, Toubert ME, Paycha F, et al. Thyroid function tests at the Assistance Publique - Hopitaux de Paris: Ordering, cost and opinions of physicians specialized in endocrinology. Annales d'Endocrinologie 1999;60(1):48-55.	No data on outcome of interest
Viera AJ. Thyroid function testing in outpatients: are both sensitive thyrotropin (sTSH) and free thyroxine (FT4) necessary? Fam Med 2003;35(6):408-10. [published Online First: 2003/06/24]	Indication for thyroid testing not reported (i.e. not clear whether the study population is tested for suspected primary or secondary thyroid dysfunction)
Wardle CA, Fraser WD, Squire CR. Pitfalls in the use of thyrotropin concentration as a first-line thyroid-function test. Lancet 2001;357(9261):1013-14. doi: 10.1016/S0140-6736(00)04248-3	Indication for thyroid testing not reported (i.e. not clear whether the study population is tested for suspected primary or secondary thyroid dysfunction)
Yadav RK, Magar NT, Poudel B, et al. A prevalence of thyroid disorder in western part of Nepal. Journal of Clinical and Diagnostic Research 2013;7(2):193-96. doi: 10.7860/JCDR/2013/4833.2724	Retracted article

13.2 Systematic review costs and budget impact

13.2.1 Search strategy for costs and budget impact systematic review

Table 1: Search strategy for costs and budget impact systematic literature search: PubMed (MEDLINE)

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

Table 2: Search strategy for costs and budget impact systematic literature search: Embase.com

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

Table 3: Search strategy for costs and budget impact systematic literature search: NHS EED

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

Limits	No limits

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

13.2.2 Excluded studies during full-text selection

Table 1: Excluded studies found in the costs and budget impact systematic literature search

Reference	Reason for exclusion
Arribas Mir, L., Muñoz de Benito, R. M., Alguacil Cubero, P. et al. Pruebas de función tiroidea y su utilización por el médico general [Thyroid function tests and their use by the general practitioner]. Atencion primaria, 1998;11(1), 3–7.	Out of scope: other economic evaluation*
Beckett GJ., Toft AD. First-line thyroid function tests - TSH alone is not enough. Clinical endocrinology. 2003; 58(1), 20–21. https://onlinelibrary.wiley.com/doi/abs/10.1046/j.1365-2265.2003.01690.x	Out of scope: other economic evaluation
Bradshaw AB.,Bonnecaze AK.,Burns CA. et al. Impact of an Interprofessional Collaborative Quality Improvement Initiative to Decrease Inappropriate Thyroid Function Testing. Hospital Pharmacy 2020. https://doi.org/10.1177/0018578720920795	Out of scope: other economic evaluation
Caberlotto L., Tessarolo A. Economic results of thyroid reflex testing application. Biochem. Clin. 2013; 37, S340.	Publication type: abstract only
Chekuri S., Mandel A., Hightower C., et al. Quantifying of the cost of unnecessary clinical laboratory testing for hospital systems and healthcare payers. Clinical Chemistry. 2014; 60:10 Suppl.1(S181 - S182)	Publication type: abstract only
Espinosa Rodriguez J. Use of thyroid function tests in public clinical laboratories of Catalonia. Quimica Clinica. 2002; 21(4) 254-261.	Out of scope: other economic evaluation
Fraser WD., Biggart EM., O'Reilly DS., et al. Are biochemical tests of thyroid function of any value in monitoring patients receiving thyroxine replacement? BMJ (Clinical research ed.)1986;293(6550), 808–810. https://doi.org/10.1136/bmj.293.6550.808	Out of scope: other economic evaluation
Heavey L., McKinney H., Sugrue M. Thyroid function tests: TSH alone, is it adequate? Ir. J. Med. Sci. 2012; 0021-1265(181), S100-S101	Publication type: abstract only
Kim Y. Experience of a Break-Even Point Analysis for Make-or-Buy Decision. K J Lab Med, 2006; 26(6), 460–464. https://www.annlabmed.org/journal/view.html?doi=10.3343/kjlm.2006.26.6.460	Out of scope: other economic evaluation
Ko SQ., Quah P., Lahiri M. The cost of repetitive laboratory testing for chronic disease. Intern Med J. 2019;49(9):1168-1170. https://doi: 10.1111/imj.14428	Disease area: out of scope
Ljunggren JG., Törring O., Wallin G., et al. Quality of life aspects and costs in treatment of Graves' hyperthyroidism with antithyroid drugs, surgery, or radioiodine: results from a prospective, randomized study. Thyroid: J Am Thyroid Assoc. 1998; 8(8), 653–659. https://doi.org/10.1089/thy.1998.8.653	Disease area: out of scope
Ma I., Lau CK., Ramdas Z., et al. Estimated costs of 51 commonly ordered laboratory tests in Canada. Clin Biochem. 2019;65, 58–60. https://doi.org/10.1016/j.clinbiochem.2018.12.013	Out of scope: other economic evaluation
Miller GD., Rogers JC., DeGroote SL., et al. (2008). Clinical inquiries: which lab tests are best when you suspect hypothyroidism?. J Fam	Publication type: abstract only

Prac. 2008;57(9), 613–614.	
Murphy MJ. Reflections on reflex thresholds. Annals Clin Biochem. 2012; 49(6), 515–517. https://doi.org/10.1258/acb.2012.012186	Publication type: editorial
Panareo S., Rossi R., Fabbri S., et al. A practical method for the estimation of therapeutic activity in the treatment of Graves' hyperthyroidism. Q J Nucl Med Mol Image. 2011;55(5), 576–585.	Out of scope: clinical study
Riesco PM., Barceló Bennassar A., Pérez Esteban G., et al. Valoración de un protocolo diagnóstico de disfunción tiroidea [Evaluation of a diagnostic protocol of thyroid dysfunction]. Atencion Primaria, 1997; 20(7), 355–360.	Out of scope: clinical study
Schectman JM., Pawlson LG. The cost-effectiveness of three thyroid function testing strategies for suspicion of hypothyroidism in a primary care-setting. J Gen Inter Med. 1990;5(1), 9–15. https://doi.org/10.1007/BF02602302	Out of scope: other economic evaluation
Schlienger JL., Heist D., Demangeat C., et al. Dysfonctionnement thyroïdien asymptomatique. Dépistage systématique au moyen de l'index de thyroxine libre [Asymptomatic thyroid dysfunction. Routine detection using the free thyroxine index]. Presse medicale 1983;12(3), 147–151.	Out of scope: clinical study
Schlienger JL., Sapin R., Gasser F., et al. Dosage ultra-sensible de l'hormone thyréotrope. Amélioration des performances et réduction du coût de l'exploration fonctionnelle thyroïdienne [Ultrasensitive determination of thyrotropin. Improvement in the performance and reduction of the cost of thyroid function tests]. Presse medicale 1987;16(1), 15–18.	Out of scope: other economic evaluation
Vidal-Trécan G., Toubert ME., Paycha F., et al. Les dosages hormonaux thyroïdiens à l'Assistance Publique-Hôpitaux de Paris: prescriptions, coût et opinions des médecins qualifiés en endocrinologie [Thyroid hormone determination at the Public Assistance Hospital of Paris: ordering, costs and opinions of physicians qualified in endocrinology]. Annales d'endocrinologie, 1999; 60(1), 48–55.	Out of scope: other economic evaluation
Weisman, A., Gilmour J., Orlov S., et al. Reducing Inappropriate Thyroid Function Tests at an Academic Ambulatory Hospital: Baseline Assessment for a Quality Improvement Initiative. Endocrine Reviews. 2015; 36 Supplement 2.	Publication type: abstract only

***** Other economic evaluations: Studies that did not report costs or budget impact outcomes, specifically when the inclusion and exclusion criteria of this study is considered.

13.3 Estimation of average number, increase and costs of thyroid function tests

		Treatment year					Average number of tests over 5 years	Average yearly increase
		2015	**2016**	**2017**	**2018**	**2019**		
Tarifpool Code	**Description**	**Frequency**						
1718.10	TSH	2’037’301	2’243’221	2’307’048	2’443’227	2’504’333	2’307’026	5.3%
1720.00	fT4	705’266	775’420	807’136	855’293	874’380	814’303	5.5%
1721.00	T4	10’610	9’781	10’202	11’564	11’863		
1732.00	fT3	419’810	466’596	492’907	534’256	554’201	517’412	6.6%
1733.00	T3	25’840	25’572	24’109	22’488	21’284		
		Costs (CHF)						
1718.10	TSH	18’444’384	20’260’068	20’902’871	22’157’175	22’722’330		
1720.00	fT4	6’370’663	6’993’330	7’295’314	7’736’576	7’909’929		
1721.00	T4	95’155	88’593	92’544	104’856	107’360		
1732.00	fT3	4’370’517	4’852’306	5’141’525	5’576’139	5’784’153		
1733.00	T3	269’653	266’692	251’657	235’249	222’561		

Keys: CHF=Swiss Franc, fT3=free triiodothyronine, fT4=free thyroxine, TSH=thyroid-stimulating hormone, T3=triiodothyronine, T4=thyroxine.

13.4 Additional results budget impact analysis

Table 1: Year 2021 results of budget impact analysis

		Thyroid Function Test		
		One-step test approach		
Estimated number of tests	**Thyroid function tests unit costs**	**One-step test costs**	**Outpatient visits**	**Total Costs**
924’434 564’976 364’980	TSH: 9 + 24 CHF fT4/T4: 9 CHF fT3/T3: 10.40 CHF	30’506’322 CHF 5’084’784 CHF 3’795’792 CHF	84’594’955 CHF	**123’981’853 CHF**
		Two-step test approach		
Estimated number of tests	**Thyroid function tests unit costs**	**Two-step test costs**	**Outpatient/Follow-up visits**	**Total Costs**
924’434 11’864 7’665	TSH: 9 + 24 CHF fT4/T4: 9 + 24 CHF fT3/T3: 10.40 + 24 CHF	30’506’322 CHF 391’512 CHF 263’676 CHF	84’594’955 CHF 1’085’675 CHF 701’424 CHF	**117’543’564 CHF**
		Reflex test approach		
Estimated number of tests	**Thyroid function tests unit costs**	**Reflex test costs**	**Outpatient visits**	**Total Costs**
924’434 11’864 7’665	TSH: 9+24 CHF fT4/T4: 9 CHF fT3/T3: 10.40 CHF	30’506’322 CHF 106’776 CHF 79’716 CHF	84’594’955 CHF	**115’287’769 CHF**
Reflex test and two-step test approaches are cost-saving in Switzerland compared to one-step test approach		**Incremental Costs**		
		Two-step test vs. one-step test approach		**- 6’438’289 CHF**
		Reflex test vs. two-step test approach		**- 2’255’795 CHF**
		Reflex test vs. one-step test approach		**- 8’694’084 CHF**

Keys: CHF=Swiss franc, fT3=free triiodothyronine, fT4=free thyroxine, TSH=thyroid-stimulating hormone, T3=free triiodothyronine, T4=free thyroxine.
24 CHF: Swiss tax

Table 2: Year 2022 results of budget impact analysis

		Thyroid Function Test		
		One-step test approach		
Estimated number of tests	**Thyroid function tests unit costs***	**One-step test costs**	**Outpatient visits**	**Total Costs**
973’429 596’049 389’069	TSH: 8.60 + 23 CHF fT4/T4: 8.60 CHF fT3/T3: 10 CHF	30’760’356 CHF 5’126’021 CHF 3’890’690 CHF	89’078’488 CHF	**128’855’555 CHF**
		Two-step test approach		
Estimated number of tests	**Thyroid function tests unit costs**	**Two-step test costs**	**Outpatient/Follow-up visits**	**Total Costs**
973’429 12’517 8’170	TSH: 8.60 + 23 CHF fT4/T4: 8.60 + 23 CHF fT3/T3: 10 + 23 CHF	30’760’356 CHF 395’537 CHF 269’610 CHF	89’078’488 CHF 1’145’431 CHF 747’637 CHF	**122’397’059 CHF**
		Reflex test approach		
Estimated number of tests	**Thyroid function tests unit costs**	**Reflex test costs**	**Outpatient visits**	**Total Costs**
973’429 12’517 8’170	TSH: 8.60 +23 CHF fT4/T4:8.60 CHF fT3/T3: 10 CHF	30’760’356 CHF 107’646 CHF 81’700 CHF	89’078’488 CHF	**120’028’190 CHF**
Reflex test and two-step test approaches are cost-saving in Switzerland compared to one-step test approach			**Incremental Costs**	
			Two-step test vs. one-step test approach	**- 6’458’496 CHF**
			Reflex test vs. two-step test approach	**- 2’368’869 CHF**
			Reflex test vs. one-step test approach	**- 8’827’365 CHF**

Keys: CHF=Swiss franc, fT3=free triiodothyronine, fT4=free thyroxine, TSH=thyroid-stimulating hormone, T3=free triiodothyronine, T4=free thyroxine.
* calculated as weighted average according to new prices (1 August 2022):
TSH: (7*9 + 5*8.10)/12 = 8.6 CHF; fT4/T4: (7*9 + 5*8.10)/12 = 8.6 CHF; fT3/T3:(7* 10.40 + 5*9.40)/12 = 10 CHF; 23 CHF Swiss tax: (7*24 + 5*21.60)/12

Table 3: Year 2023 results of budget impact analysis

		Thyroid Function Test		
		One-step test approach		
Estimated number of tests	**Thyroid function tests unit costs**	**One-step test costs**	**Outpatient visits**	**Total Costs**
1’025’021 628’832 414’747	TSH: 8.10 + 21.60 CHF fT4/T4: 8.10 CHF fT3/T3: 9.40 CHF	30’443’124 CHF 5’093’539 CHF 3’898’622 CHF	93’799’672 CHF	**133’234’957 CHF**
		Two-step test approach		
Estimated number of tests	**Thyroid function tests unit costs**	**Two-step test costs**	**Outpatient/Follow-up visits**	**Total Costs**
1’025’021 13’205 8’710	TSH: 8.10 + 21.60 CHF fT4/T4: 8.10 + 21.60 CHF fT3/T3: 9.40 + 21.60 CHF	30’443’124 CHF 392’189 CHF 270’010 CHF	93’799’672 CHF 1’208’390 CHF 797’052 CHF	**126’910’437 CHF**
		Reflex test approach		
Estimated number of tests	**Thyroid function tests unit costs**	**Reflex test costs**	**Outpatient visits**	**Total Costs**
1’025’021 13’205 8’710	TSH: 8.10 + 21.60 CHF fT4/T4: 8.10 CHF fT3/T3: 9.40 CHF	30’443’124 CHF 106’961 CHF 81’874 CHF	93’799’672 CHF	**124’431’631 CHF**
Reflex test and Two-step test approaches are cost-saving in Switzerland compared to one-step test approach		**Incremental Costs**		
		Two-step test vs. one-step test approach		**- 6’324’520 CHF**
		Reflex test vs. two-step test approach		**- 2’478’806 CHF**
		Reflex test vs. one-step test approach		**- 8’803’326 CHF**

Keys: CHF=Swiss franc, fT3=free triiodothyronine, fT4=free thyroxine, TSH=thyroid-stimulating hormone, T3=free triiodothyronine, T4=free thyroxine.
21.60 CHF: Swiss tax

Table 4: Year 2024 results of budget impact analysis

		Thyroid Function Test		
		One-step test approach		
Estimated number of tests	**Thyroid function tests unit costs**	**One-step test costs**	**Outpatient visits**	**Total Costs**
1’079’347	TSH: 8.10 + 21.60 CHF	32’056’606 CHF	98’771’044 CHF	**140’357’273 CHF**
663’418	fT4/T4: 8.10 CHF	5’373’686 CHF		
442’121	fT3/T3: 9.40 CHF	4’155’937 CHF		
		Two-step test approach		
Estimated number of tests	**Thyroid function tests unit costs**	**Two-step test costs**	**Outpatient/Follow-up visits**	**Total Costs**
1’079’347	TSH: 8.10 + 21.60 CHF	32’056’606 CHF	98’771’044 CHF	**133’653’853 CHF**
13’932	fT4/T4: 8.10 + 21.60 CHF	413’780 CHF	1’274’917 CHF	
9’285	fT3/T3: 9.40 + 21.60 CHF	287’835 CHF	849’670 CHF	
		Reflex test approach		
Estimated number of tests	**Thyroid function tests unit costs**	**Reflex test costs**	**Outpatient visits**	**Total Costs**
1’079’347	TSH: 8.10 + 21.60 CHF	32’056’606 CHF	98’771’044 CHF	**131’027’778 CHF**
13’932	fT4/T4: 8.10 CHF	112’849 CHF		
9’285	fT3/T3: 9.40 CHF	87’279 CHF		
Reflex test and two-step test approaches are cost-saving in Switzerland compared to one-step test approach			**Incremental Costs**	
			Two-step test vs. one-step test approach	**- 6’703’420 CHF**
			Reflex test vs. two-step test approach	**- 2’626’075 CHF**
			Reflex test vs. one-step test approach	**- 9’329’495 CHF**

Keys: CHF=Swiss franc, fT3=free triiodothyronine, fT4=free thyroxine, TSH=thyroid-stimulating hormone, T3=free triiodothyronine, T4=free thyroxine.
21.60 CHF: Swiss tax