Association of thyroid hormone therapy with quality of life and thyroid-related symptoms in patients with subclinical hypothyroidism: a systematic review and meta-analysis

Martin Feller, MD MSc[1,2,3], Marieke Snel, MD PhD[3], Elisavet Moutzouri, MD PhD[1,2], Douglas C Bauer, MD[4], Maria de Montmollin, MD[1,2], Drahomir Aujesky, MD MSc[1], Ian Ford, PhD[5], Jacobijn Gussekloo, MD PhD[3,6], Patricia M Kearney, MD PhD MPH[7], Simon Mooijaart MD PhD[3,8], Terry Quinn, MD[9], David Stott, MD[9], Rudi Westendorp, MD PhD[10], Nicolas Rodondi, MD MAS[1,2], Olaf M Dekkers MD, MA, MSc, PhD[3,11,12]

Authors affiliations:

[1]Department of General Internal Medicine, Inselspital, Bern University Hospital, University of Bern, Switzerland; [2]Institute of Primary Health Care (BIHAM), University of Bern, Switzerland; [3]Departement of Endocrinology / General Internal Medicine, Leiden University Center, The Netherlands; [4]Departments of Medicine and Epidemiology and Biostatistics, University of California, San Francisco, California, USA; [5]Robertson Centre for Biostatistics, Institute of Health and Wellbeing, University of Glasgow, UK; [6]Department of Public Health and Primary Care, Leiden University Center, The Netherlands; [7]School of Public Health, University College Cork, Ireland; [8]Institute for Evidence-based Medicine in Old Age, Leiden University Center, The Netherlands; [9]Institute of Cardiovascular Medicine, University of Glasgow, UK; [10]Department of Public Health and Center for Healthy Aging, University of Copenhagen, Denmark; [11]Department of Clinical Epidemiology, Leiden University Medical Centre, Leiden, the Netherlands; [12]Department of Clinical Epidemiology, Aarhus University Hospital, Aarhus, Denmark

Corresponding author:

Martin Feller, Department of General Internal Medicine, Inselspital, Bern University Hospital, University of Bern, Switzerland; Phone number: +41 31 632 77 47; Email: Martin.Feller@insel.ch

Date of revision: August 29, 2018

Manuscript word count: 2,929

Key Points

- Among patients with subclinical hypothyroidism, is the use of thyroid hormone therapy associated with improvements in general quality of life or thyroid-related symptoms?
- In this meta-analysis of 21 randomized clinical trials including 2,192 participants with subclinical hypothyroidism, thyroid hormone therapy was not significantly associated with improvements in general quality of life (standardized mean difference [SMD], -0.11) or thyroid-related symptoms (SMD, 0.01).
- These findings do not support the routine use of thyroid hormone therapy in adults with subclinical hypothyroidism.

Abstract

Importance: The benefit of thyroid hormone therapy for subclinical hypothyroidism is uncertain. New evidence from recent large randomized clinical trials (RCT) warrants an update of previous meta-analyses.

Objective: To conduct a meta-analysis of the association of thyroid hormone therapy with quality of life and thyroid related symptoms in adults with subclinical hypothyroidism.

Data Sources: PubMed, Embase, clinicaltrials.gov, Web of Science, COCHRANE Library, CENTRAL, Emcare and Academic Search Premier from inception until July 4, 2018.

Study Selection: RCTs that compared thyroid hormone therapy to placebo/no therapy in non-pregnant adults with subclinical hypothyroidism were eligible. Two reviewers independently evaluated eligibility based on titles and abstracts of all retrieved studies. Studies not excluded in this first step were independently assessed for inclusion after full-text evaluation by two reviewers.

Data Extraction and Synthesis: Two independent reviewers extracted data, assessed risk of bias (Cochrane risk of bias tool), and evaluated the quality of evidence (GRADE tool). For synthesis, differences in clinical scores were transformed (e.g. quality of life) into standardized mean differences (SMD, positive values indicate benefit of thyroid hormone therapy; 0.2, 0.5 and 0.8 correspond to small, moderate and large effects). Random-effects models for meta-analyses were applied.

Main Outcomes and Measures: General quality of life and thyroid-related symptoms after a minimum follow-up of three months.

Results: Overall, 21 of 3,088 initially identified publications met the inclusion criteria with 2,192 adults randomized. After treatment (range 3 to 18 months), thyroid hormone therapy was associated with lowering the mean TSH value into the normal reference range, compared with placebo (range 0.5 to 3.7mU/l vs 4.6 to 14.7mU/l), but was not associated with benefit regarding general quality of life (n=796, SMD -0.11, 95%CI -0.25 to 0.03, I^2=66.7%) and thyroid-related symptoms (n=858, SMD 0.01, 95%CI -0.12 to 0.14, I^2=0.0%). Overall, risk of bias was low and the quality of evidence assessed with the GRADE tool was judged moderate to high.

Conclusion and Relevance: Among non-pregnant adults with subclinical hypothyroidism, the use of thyroid hormone therapy was not associated with improvements in general quality of life or thyroid-related symptoms. These findings do not support the routine use of thyroid hormone therapy in adults with subclinical hypothyroidism.

Registration: The study protocol was registered on PROSPERO (CRD42017055536).

Introduction

Subclinical hypothyroidism, defined as elevated Thyroid-stimulating hormone (TSH) in combination with a normal free thyroxine (fT4),[1] is common.[2,3] According to the NHANES III report,[4] an estimated 13 million people have subclinical hypothyroidism in the United States. The prevalence is higher in women and in older people.[2,3] Subclinical hypothyroidism is often treated with thyroid hormones (levothyroxine),[5] particularly when it co-occurs with symptoms potentially attributable to hypothyroidism such as tiredness, constipation, and unexplained weight gain.[5]

Relatively limited evidence exists from randomized clinical trials (RCTs) to guide therapy of subclinical hypothyroidism. Systematic reviews have been inconclusive and clinical practice guidelines have varied regarding recommendations for managing subclinical hypothyroidism.[6-10] Two large randomized trials of levothyroxine therapy in patients with subclinical hypothyroidism were recently completed.[11,12] This meta-analysis and systematic review incorporates recent trials and evaluated whether thyroid hormone therapy was associated with improved symptoms and other benefits in non-pregnant adults with subclinical hypothyroidism.

Methods

We followed the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) statement[13] and published the protocol of this systematic review on PROSPERO (CRD42017055536).

Eligibility criteria, literature search and study selection

We considered randomized trials that included non-pregnant adults with subclinical hypothyroidism. Subclinical hypothyroidism was defined as TSH above the reference range in combination with an fT4 within the reference range (according to center-specific reference ranges). The intervention had to consist of thyroid hormone therapy (either triiodothyronine (T3), thyroxine (T4) or a combination of both) for at least one month, with a minimum follow-up of three months. The control group had to receive either placebo or no therapy. In order to be included, studies had to report quantitative data for at least one of the study's primary or secondary outcomes: general quality of life, thyroid-related quality of life/hypothyroid symptoms, depressive symptoms, fatigue/tiredness, cognitive function, pain, muscle strength, blood pressure, body-mass index, cardiovascular events (myocardial infarction, stroke, revascularization), mortality; or side effects (hyperthyroidism due to overdosing). Data had to be reported with effect estimates and measures of precision (standard deviations or standard errors). The primary outcomes were general quality of life and thyroid-related quality of life/hypothyroid

symptoms, whereas depressive symptoms, fatigue/tiredness, cognitive function, pain, muscle strength, blood pressure, body-mass index, cardiovascular events (myocardial infarction, stroke, revascularization), mortality and side effects (hyperthyroidism due to overdosing) were secondary outcomes. Studies that only included patients with subclinical hypothyroidism in combination with another specific condition (e.g. patients with diabetic nephropathy and subclinical hypothyroidism) were excluded, because this type of study population is not representative of most patients with subclinical hypothyroidism. We excluded studies that exclusively enrolled pregnant women and/or women who wanted to become pregnant. Pseudorandomization (pre-post comparisons for example) did not qualify for inclusion.

We searched MEDLINE, EMBASE, Web of Science, COCHRANE Library, CENTRAL, Emcare and Academic Search Premier from inception until July 4, 2018 in cooperation with a trained librarian. Search terms were adapted according to the syntax of each specific database, and no language restrictions were applied. We searched trial registries (clinicaltrials.gov) for upcoming (and not yet published) trials on this research topic and asked authors for the status of the trial if not published yet. We screened references of key articles for additional potentially relevant articles. Details of the search strategy are presented in the Appendix.

Two researchers (MS, MDM) evaluated eligibility independently, based on titles and abstracts of all studies retrieved in the electronic search. Studies not excluded in this first step were independently assessed for inclusion after full-text evaluation by two reviewers (MF, MS). We manually screened bibliographies of the included studies as well as guidelines and major reviews for additional studies. Discrepancies were resolved by consensus among the study team.

Data extraction and risk of bias assessment

A standard data extraction form was used, adapted from a template suggested by Cochrane (see Appendix).[14] Two researchers (MS, EM) independently extracted bibliographic details, funding source, eligibility criteria, information about the study population and setting, study design, risk of bias, intervention/control intervention, results, and independently evaluated the quality of evidence (GRADE tool).[15] In case a study reported more than one outcome measure for a specific outcome domain (e.g. more than one cognition test to assess cognitive function), we chose the most relevant measure, based on how broad a domain was assessed and international usage (by consensus among the study team). As an example, Parle and colleagues reported five different cognition tests.[16] We analyzed results from the Mini-Mental State Examination (MMSE) because it is used worldwide and because it

is a broader assessment of cognitive function than alternative tests, such as the Trail Making Test. When a study mentioned an outcome of interest without providing estimates, we contacted the author for the data (e.g. the study reported no difference in body-mass index between the intervention and the control group without providing data on mean differences and standard deviations). If studies reported results for an outcome at multiple time points during the intervention (e.g. body mass index at 6 and 12 months), only the most recent measurement was used in statistical analyses. Data were extracted in duplicate by two independent reviewers (MS, EM), and differences were resolved by consensus.

Statistical analyses

Study results were presented separately for each outcome, with estimates as reported in the original publication and transformed into standardized mean differences (SMD) when different scales were used for the same outcome domain. We coded SMDs such that positive values indicated benefit of thyroid hormone therapy, with 0.2, 0.5 and 0.8 corresponding to small, moderate and large effects.[17] In contrast, for body-mass index and blood pressure, negative values indicated benefit of thyroid hormone therapy. For estimations of treatment effects, we used mean values and their standard deviations at end of treatment in both groups, assuming balanced baseline values due to the randomised designs.

For outcomes where studies reported treatment effects at different time points, we only included the estimate at the most recent follow-up time point, thereby avoiding counting a study twice in a formal meta-analysis. Overall results were calculated using random effects models, unless less than five studies were included for a meta-analysis, as in this case the between-study variance cannot be estimated reliably and a fixed effect analysis was performed. For better clinical interpretation, overall SMDs were also back-transformed to one original scale according to a method proposed by the Cochrane Collaboration,[17] for general quality of life, thyroid-related quality of life/hypothyroid symptoms, depressive symptoms, cognitive function and muscle strength. Heterogeneity was assessed visually with forest plots and quantified with I^2 (0-40%, 40-75% and >75% for “low”, “moderate” and “high”, respectively). If substantial heterogeneity existed and a sufficient number of publications was available (n=10), we aimed to explore potential sources of heterogeneity in protocol pre-specified subgroup analyses (e.g. restricting the analysis to high-quality studies). In addition, a post-hoc sensitivity analysis was performed with the aim to evaluate heterogeneity after excluding studies showing statistically significant benefit of placebo treatment. In case of a sufficient number of

publications (n=10), publication bias would be assessed via funnel plots (visually), and more formally with the Egger test.[18] Statistical significance was tested 2-sided, and P-values of <0.05 were judged statistically significant. All analyses were conducted with Stata, release 14.

Results

The systematic literature search retrieved 3,086 studies, and two additional studies were retrieved after searching references of key articles. After removing duplicates (1,438), two reviewers (MS, MDM) independently screened 1,650 unique articles for potential eligibility based on title and abstract. Forty-nine potentially eligible studies were evaluated in full-text independently by two reviewers (MF, MS). Among these, 25 studies did not meet the inclusion criteria. In addition, three studies would have met the inclusion criteria but did not present the data in a manner to be included in the meta-analysis.[19-21] For two of these articles, the authors indicated that data were not available anymore.[19,20] One author could not be reached (eTable 1 for excluded studies).[21] Finally, 21 studies met the inclusion criteria (Flow chart depicted in eFigure 1). Among the 21 studies, a total of 2,192 adults were randomized (Table 1 for included studies). The study size ranged from 20 to 737 participants; the mean age ranged from 32 to 74 years, the proportion of women ranged from 46% to 100%, and baseline mean TSH ranged from 4.4 to 12.8mU/l. Two studies (99 participants) had a mean baseline TSH >10mU/l.[22,23] Seven studies provided information about hypothyroid symptoms at baseline and in these studies the burden of symptoms was mild to moderate (Table 1).[12,22-27] In the thyroid hormone therapy groups, mean TSH at the end of follow-up ranged between 0.5 and 3.7mU/l (eTable 2), indicating that treatment was associated with normalization of TSH levels. In contrast, mean TSH in the placebo/no intervention groups remained elevated at the end of follow-up, ranging from 4.6 to 14.7mU/l (eTable 2). The duration of the intervention (thyroid hormone therapy or placebo/no therapy) ranged from between 3 and 18 months. Three studies compared thyroid hormone therapy to no intervention and the other studies compared thyroid hormone therapy to placebo,[11,28,29] Two studies were supported by industry (Table 1).[22,30]

Thyroid hormone therapy was not associated with benefit for either of the two primary outcomes: Four studies including 796 participants evaluated general quality of life (SMD -0.11, 95%CI -0.25 to 0.03, I^2 66.7%; Figure 1).[12,26,27,31] It is estimated that on the EQ-5D scale (range -0.59 to 1.00, higher scores indicate better quality of life), this SMD would represent a difference of 0.02 (95%CI -0.01 to 0.05) in

favor of placebo. Four studies including 858 participants evaluated thyroid-related quality of life/hypothyroid symptoms (SMD 0.01, 95%CI -0.12 to 0.14, I^2 0.0%; Figure 1).[12,22,31,32] It is estimated that on the ThyPRO Hypothyroid Symptoms Score (range 0 to 100, higher scores indicate more hypothyroid symptoms), this SMD would represent a difference of 0.18 (95%CI -2.10 to 2.45) in favor of levothyroxine. Similarly, thyroid hormone therapy was not associated with benefit regarding the secondary outcomes: depressive symptoms (four studies, 278 participants, SMD -0.10, 95%CI -0.34 to 0.13, I^2 0.0%; Figure 1),[16,24,26,31] on the Hospital Anxiety and Depression Scale (range 0 to 21, higher scores indicate worse depressive symptoms), this SMD would represent a difference of 0.28 (95%CI -0.36 to 0.95) in favor of placebo; cognitive function (four studies, 859 participants, SMD 0.09, 95%CI -0.05 to 0.22, I^2 14.7%; Figure 2),[12,16,25,26] on the Letter-digit coding test scale (range 0 or higher (no upper limit), higher scores indicate better cognitive function), this SMD would represent a difference of 1.01 (95%CI -0.56 to 2.46) in favor of levothyroxine; muscle strength (two studies, 695 participants, SMD 0.1, 95%CI -0.1 to 0.2, I^2 0.0%; eTable 4),[12,31] in handgrip strength (in kg), this SMD would represent a difference of 1.12 (95%CI -1.12 to 2.24) in favor of levothyroxine; systolic blood pressure (eight studies, 1,372 participants, -0.7mmHg, 95%CI -2.6 to 1.2, I^2 0.0%; Figure 3),[11,12,29,32-36] or body-mass index (15 studies, 1,633 participants, 0.2kg/m^2, 95%CI -0.4 to 0.8, I^2 45.5%; Figure 4).[11,12,23,27-30,32-39] Only the TRUST trial (the largest included study, with 737 participants randomized) evaluated fatigue/tiredness, cardiovascular events, mortality, and side effects.[12] No beneficial or harmful association between thyroid hormone therapy and these outcomes was reported (Figure 1 & eTable 2 & 4). No study included pain as an outcome. Detailed results were summarized in eTable 2 & 4. Subgroup analyses were not performed, beacuse the number of studies for a single outcome was too small and/or there was low to moderate heterogeneity such that no exploration was indicated. As the meta-analyses for general quality of life and body-mass index showed moderate heterogeneity (I^2=66.7% and I^2=45.5%, respectively), post-hoc sensitivity analyses were performed, excluding studies showing statistically significant benefit of placebo.[23,27,30] Results remained similar, but heterogeneity was lower (general quality of life SMD -0.08, 95%CI -0.22 to 0.06, I^2 34.7%; body-mass index -0.2kg/m^2, 95%CI -0.6 to 0.2, I^2 1.6%).Further, we did not formally assess publication bias. Based on the negative results, there was no indication that positive studies were published while negative studies remained unpublished.

The overall quality of the 21 included studies was good with only nine out of 126 items judged to be at high risk of bias (eTable 3); two trials had low risk of bias for all critieria,[12,22] including the largest and

most recent one,[12] and only one trial, the second largest and second most recent, had a high risk of bias in three out of six domains.[11] Accordingly, the quality of evidence assessed with the GRADE tool was high regarding the main outcomes general quality of life and thyroid-related symptoms, as well as regarding muscle strength, blood pressure and body-mass index (eTable 4). The quality of evidence was moderate for depressive symptoms, fatigue / tiredness, cognitive function and side effects, whereas it was low for cardiovascular events and mortality (eTable 4).

Discussion

In this systematic review and meta-analysis of RCTs in non-pregnant adults with subclinical hypothyroidism, thyroid hormone therapy was not associated with benefit regarding general quality of life, thyroid-related symptoms, depressive symptoms, fatigue/tiredness, cognitive function, muscle strength, blood pressure and body-mass index.

Compared to prior systematic reviews and meta-analyses, published between 2007[8] and 2015[9], this meta-analysis included two recent randomized trials, which were the largest trials published to date on this topic.[11,12] Overall, the quality of evidence reported here was moderate to high. Quality of evidence was high regarding the primary outcomes of this review (general quality of life and thyroid-related symptoms). Results of this review consistently demonstrated no association of thyroid replacement therapy with improved outcomes, including a relatively large number of diverse outcomes. Most outcomes, except cardiovascular events and mortality, had narrow confidence intervals. In addition, this meta-analysis focused on patient-centered outcomes such as quality of life and fatigue, which are the most common symptoms that prompt therapy in general practice.[40]

Although current guidelines are, at first sight, cautious with treatment recommendations, more than 90% of persons with subclinical hypothyroidism and a TSH <10mU/l would actually qualify for treatment. [6,10,41] However, results of this meta-analysis are not consistent with these guideline recommendations. In addition to absence of an association of thyroid hormone therapy with improved outcomes, thyroid hormone therapy is associated with side effects when overtreatment occurs.[5,42,43]

Limitations

This study has several limitations. First, the RCTs included in this meta-analysis used different questionnaires and/or tests for a given outcome in combination with different treatment durations (e.g. four different cognitive tests in the four studies examining cognitive function, with treatment durations

ranging from three to 18 months). However, little heterogeneity across the study results was observed, except for general quality of life and body-mass index. For these outcomes, heterogeneity resulted from three studies that showed a statistically significant benefit of placebo.[23,27,30] After excluding these studies in post-hoc sensitivity analyses, thyroid hormone therapy remained unassociated with benefit for general quality of life and body-mass index, but heterogeneity was lower. Therefore, it seems unlikely that this meta-analysis misses a potential beneficial association between thyroid hormone therapy and any outcome analyzed due to inappropriate pooling of overly heterogeneous studies. Second, only one RCT reported on major adverse cardiovascular events. Therefore, definitive evidence is lacking regarding the association of therapy for subclinical hypothyroidism with reduced cardiovascular event rates.[12] Third, RCTs that reported results only qualitatively were excluded from analyses. Fourth, mean TSH values at baseline were <7.0mU/l in 11 out of 21 included RCTs, and only two RCTs examined participants with a mean baseline TSH >10mU/l.[22,23] Therefore, the current findings may not be generalizable to people with subclinical hypothyroidism and a TSH >10mU/l. Fifth, the highest mean age in the included studies was 74 years.[12,16] Therefore, these results may not be generalizable to people older than 80 years. Sixth, only seven of 21 trials (33%) reported hypothyroid symptoms at baseline, and the burden of symptoms was mild to moderate in these trials. The other 14 trials did not describe symptoms at baseline. It is possible that the subgroup of people with subclinical hypothyroidism and a high burden of symptoms would still benefit from treatment. Seventh, patients with subclinical hypothyroidism and "severe" symptoms of hypothyroidism may be underrepresented in clinical trials because they may be treated immediately with levothyroxine and are not included in clinical trials.[44] Therefore, results reported here may not be generalizable to patients with subclinical hypothyroidism who have severe symptoms. Eighth, two RCTs (n=831) included participants with a mean age >65 years.[12,16] Their mean TSH level at baseline was between 6.0 and 7.0mU/l. Given the possibility that the upper TSH reference limit may increase with age,[45] the two studies may have included older individuals with mildly elevated TSH levels who do not represent subclinical hypothyroidism, although current international guidelines do not use different TSH levels according to age to define subclinical hypothyroidism.[6,10,46] However, this phenomenon may have biased the results towards the null. Ninth, it is possible that thyroid hormone therapy is associated with benefit regarding outcomes that were not examined in this meta-analysis (e.g. carotid intima media thickness, various lipid fractions, etc.). Tenth, it is possible that treatment of subclinical hypothyroidism may be beneficial in study populations not included in these analyses (e.g. patients with subclinical hypothyroidism and

renal impairment). Eleventh, the largest RCT to date[12] contributed substantially to the results of this meta-analysis because of the large sample size relative to the other trials (737 of 2,192 participants (33.6%)). However, the mean age of participants in the largest trial was 74 years, while the mean age of participants in the studies included here ranged from 32 to 74 years.

Conclusions

Among non-pregnant adults with subclinical hypothyroidism, the use of thyroid hormone therapy was not associated with improvements in general quality of life or thyroid-related symptoms. These findings do not support the routine use of thyroid hormone therapy in adults with subclinical hypothyroidism.

Acknowledgments

We thank Dr Fahime Ghanbari, MD, Department of General Internal Medicine, Inselspital, Bern University Hospital, University of Bern, Switzerland, for translating an article written in Persian.[47] We thank Jan W. Schoones, MA, the librarian responsible for the literature search, Leiden University Medical Center, Leiden, The Neterlands. We also thank Dr Salman Razvi, MD, Institute of Genetic Medicine, Newcastle University, Newcastle upon Tyne, UK, who provided us with data on body-mass index such that we could include his results in the meta-analysis.[32] None of the persons mentioned in the *Acknowledgments* received compensation, and all gave written permission to be named in this section.

This article was presented orally in part at the European Congress of Endocrinology in Barcelona, Spain, May 19–22, 2018, and at the annual meeting of the Swiss Society of General Internal Medicine in Basel, Switzerland, May 30 – June 1, 2018.

Funding source: The Swiss National Science Foundation funded this study (SNSF 320030_172676 to Prof Rodondi). The funding source had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Conflict of interest: None of the authors has a conflict of interest regarding this manuscript.

Author Contributions: Drs Feller and Dekkers had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. Drs Feller and Dekkers had final responsibility for the decision to submit for publication.

Study concept and design: Feller, Bauer, Rodondi, Dekkers

Acquisition, analysis, or interpretation of data: Feller, Snel, Moutzouri, de Montmollin, Bauer, Ford, Gussekloo, Kearney, Mooijaart, Quinn, Aujesky, Stott, Westendorp, Rodondi, Dekkers

Drafting of the manuscript: Feller, Dekkers

Critical revision of the manuscript for important intellectual content: Feller, Snel, Moutzouri, de Montmollin, Bauer, Ford, Gussekloo, Kearney, Mooijaart, Quinn, Aujesky, Stott, Westendorp, Rodondi, Dekkers

Statistical analysis: Snel, Dekkers

Obtained funding: Rodondi

Study supervision: Feller, Rodondi, Dekkers

References

1. Peeters RP. Subclinical Hypothyroidism. *The New England journal of medicine.* 2017;376(26):2556-2565.
2. Biondi B, Cooper DS. The clinical significance of subclinical thyroid dysfunction. *Endocrine reviews.* 2008;29(1):76-131.
3. Cooper DS, Biondi B. Subclinical thyroid disease. *Lancet.* 2012;379(9821):1142-1154.
4. Hollowell JG, Staehling NW, Flanders WD, et al. Serum TSH, T(4), and thyroid antibodies in the United States population (1988 to 1994): National Health and Nutrition Examination Survey (NHANES III). *The Journal of clinical endocrinology and metabolism.* 2002;87(2):489-499.
5. Rodriguez-Gutierrez R, Maraka S, Ospina NS, Montori VM, Brito JP. Levothyroxine overuse: time for an about face? *The lancet. Diabetes & endocrinology.* 2017;5(4):246-248.
6. Pearce SH, Brabant G, Duntas LH, et al. 2013 ETA Guideline: Management of Subclinical Hypothyroidism. *European thyroid journal.* 2013;2(4):215-228.
7. Gharib H, Tuttle RM, Baskin HJ, Fish LH, Singer PA, McDermott MT. Subclinical thyroid dysfunction: a joint statement on management from the American Association of Clinical Endocrinologists, the American Thyroid Association, and the Endocrine Society. *The Journal of clinical endocrinology and metabolism.* 2005;90(1):581-585; discussion 586-587.
8. Villar HC, Saconato H, Valente O, Atallah AN. Thyroid hormone replacement for subclinical hypothyroidism. *The Cochrane database of systematic reviews.* 2007(3):CD003419.
9. Rugge JB, Bougatsos C, Chou R. Screening and treatment of thyroid dysfunction: an evidence review for the U.S. Preventive Services Task Force. *Annals of internal medicine.* 2015;162(1):35-45.
10. 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. *Thyroid : official journal of the American Thyroid Association.* 2012;22(12):1200-1235.
11. Zhao M, Liu L, Wang F, et al. A Worthy Finding: Decrease in Total Cholesterol and Low-Density Lipoprotein Cholesterol in Treated Mild Subclinical Hypothyroidism. *Thyroid : official journal of the American Thyroid Association.* 2016;26(8):1019-1029.
12. Stott DJ, Rodondi N, Kearney PM, et al. Thyroid Hormone Therapy for Older Adults with Subclinical Hypothyroidism. *The New England journal of medicine.* 2017;376(26):2534-2544.
13. Moher D, Liberati A, Tetzlaff J, Altman DG, Group P. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS medicine.* 2009;6(7):e1000097.
14. Effective Practice and Organisation of Care (EPOC). Data collection form. EPOC Resources for review authors. Oslo: Norwegian Knowledge Centre for the Health Services. *http://epoc.cochrane.org/sites/epoc.cochrane.org/files/public/uploads/Resources-for-authors2017/good_practice_data_extraction_form.doc (last accessed July 6, 2018).* 2017.
15. Schünemann H BJ, Guyatt G, Oxman A, editors. GRADE handbook for grading quality of evidence and strength of recommendations. 2013; guidelinedevelopment.org/handbook (last accessed July 6, 2018).
16. Parle J, Roberts L, Wilson S, et al. A randomized controlled trial of the effect of thyroxine replacement on cognitive function in community-living elderly subjects with subclinical hypothyroidism: the Birmingham Elderly Thyroid study. *The Journal of clinical endocrinology and metabolism.* 2010;95(8):3623-3632.
17. GRADE Working Group. About Re-expressing SMD. http://pregnancy.cochrane.org/sites/pregnancy.cochrane.org/files/public/uploads/re-expressing SMD new.pdf [last accessed August 7, 2018].
18. Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. *Bmj.* 1997;315(7109):629-634.
19. Abu-Helalah M, Law MR, Bestwick JP, Monson JP, Wald NJ. A randomized double-blind crossover trial to investigate the efficacy of screening for adult hypothyroidism. *Journal of medical screening.* 2010;17(4):164-169.
20. Jaeschke R, Guyatt G, Gerstein H, et al. Does treatment with L-thyroxine influence health status in middle-aged and older adults with subclinical hypothyroidism? *Journal of general internal medicine.* 1996;11(12):744-749.
21. Nystrom E, Caidahl K, Fager G, Wikkelso C, Lundberg PA, Lindstedt G. A double-blind cross-over 12-month study of L-thyroxine treatment of women with 'subclinical' hypothyroidism. *Clinical endocrinology.* 1988;29(1):63-75.

22. Meier C, Staub JJ, Roth CB, et al. TSH-controlled L-thyroxine therapy reduces cholesterol levels and clinical symptoms in subclinical hypothyroidism: a double blind, placebo-controlled trial (Basel Thyroid Study). *The Journal of clinical endocrinology and metabolism.* 2001;86(10):4860-4866.
23. Cooper DS, Halpern R, Wood LC, Levin AA, Ridgway EC. L-Thyroxine therapy in subclinical hypothyroidism. A double-blind, placebo-controlled trial. *Annals of internal medicine.* 1984;101(1):18-24.
24. Najafi L, Malek M, Hadian A, Ebrahim Valojerdi A, Khamseh ME, Aghili R. Depressive symptoms in patients with subclinical hypothyroidism--the effect of treatment with levothyroxine: a double-blind randomized clinical trial. *Endocrine research.* 2015;40(3):121-126.
25. Aghili R, Khamseh ME, Malek M, et al. Changes of subtests of Wechsler Memory Scale and cognitive function in subjects with subclinical hypothyroidism following treatment with levothyroxine. *Archives of medical science : AMS.* 2012;8(6):1096-1101.
26. Jorde R, Waterloo K, Storhaug H, Nyrnes A, Sundsfjord J, Jenssen TG. Neuropsychological function and symptoms in subjects with subclinical hypothyroidism and the effect of thyroxine treatment. *The Journal of clinical endocrinology and metabolism.* 2006;91(1):145-153.
27. Kong WM, Sheikh MH, Lumb PJ, et al. A 6-month randomized trial of thyroxine treatment in women with mild subclinical hypothyroidism. *The American journal of medicine.* 2002;112(5):348-354.
28. Cabral MD, Teixeira P, Soares D, Leite S, Salles E, Waisman M. Effects of thyroxine replacement on endothelial function and carotid artery intima-media thickness in female patients with mild subclinical hypothyroidism. *Clinics (Sao Paulo, Brazil).* 2011;66(8):1321-1328.
29. Ersoy I, Banu KK, Bagci O, et al. Effects of Levothyroxine Treatment on Cardiovascular Risk Profile and Carotid Intima Media Thickness in Patients with Subclinical Hypothyroidism. *Acta Endocrinologica-Bucharest.* 2012;8(3):433-442.
30. Teixeira PF, Reuters VS, Ferreira MM, et al. Lipid profile in different degrees of hypothyroidism and effects of levothyroxine replacement in mild thyroid failure. *Transl Res.* 2008;151(4):224-231.
31. Reuters VS, Almeida CP, Teixeira PF, et al. Effects of subclinical hypothyroidism treatment on psychiatric symptoms, muscular complaints, and quality of life. *Arq Bras Endocrinol Metabol.* 2012;56(2):128-136.
32. Razvi S, Ingoe L, Keeka G, Oates C, McMillan C, Weaver JU. The beneficial effect of L-thyroxine on cardiovascular risk factors, endothelial function, and quality of life in subclinical hypothyroidism: randomized, crossover trial. *The Journal of clinical endocrinology and metabolism.* 2007;92(5):1715-1723.
33. Monzani F, Di Bello V, Caraccio N, et al. Effect of levothyroxine on cardiac function and structure in subclinical hypothyroidism: a double blind, placebo-controlled study. *The Journal of clinical endocrinology and metabolism.* 2001;86(3):1110-1115.
34. Monzani F, Caraccio N, Kozakowa M, et al. Effect of levothyroxine replacement on lipid profile and intima-media thickness in subclinical hypothyroidism: a double-blind, placebo- controlled study. *The Journal of clinical endocrinology and metabolism.* 2004;89(5):2099-2106.
35. Yazici M, Gorgulu S, Sertbas Y, et al. Effects of thyroxin therapy on cardiac function in patients with subclinical hypothyroidism: index of myocardial performance in the evaluation of left ventricular function. *International journal of cardiology.* 2004;95(2-3):135-143.
36. Nagasaki T, Inaba M, Yamada S, et al. Decrease of brachial-ankle pulse wave velocity in female subclinical hypothyroid patients during normalization of thyroid function: a double-blind, placebo-controlled study. *European journal of endocrinology / European Federation of Endocrine Societies.* 2009;160(3):409-415.
37. Iqbal A, Jorde R, Figenschau Y. Serum lipid levels in relation to serum thyroid-stimulating hormone and the effect of thyroxine treatment on serum lipid levels in subjects with subclinical hypothyroidism: the Tromso Study. *Journal of internal medicine.* 2006;260(1):53-61.
38. Caraccio N, Natali A, Sironi A, et al. Muscle metabolism and exercise tolerance in subclinical hypothyroidism: a controlled trial of levothyroxine. *The Journal of clinical endocrinology and metabolism.* 2005;90(7):4057-4062.
39. Caraccio N, Ferrannini E, Monzani F. Lipoprotein profile in subclinical hypothyroidism: response to levothyroxine replacement, a randomized placebo-controlled study. *The Journal of clinical endocrinology and metabolism.* 2002;87(4):1533-1538.
40. Allport J, McCahon D, Hobbs FD, Roberts LM. Why are GPs treating subclinical hypothyroidism? Case note review and GP survey. *Primary health care research & development.* 2013;14(2):175-184.

41. Rosario PW, Calsolari MR. How selective are the new guidelines for treatment of subclinical hypothyroidism for patients with thyrotropin levels at or below 10 mIU/L? *Thyroid : official journal of the American Thyroid Association.* 2013;23(5):562-565.
42. Somwaru LL, Arnold AM, Joshi N, Fried LP, Cappola AR. High frequency of and factors associated with thyroid hormone over-replacement and under-replacement in men and women aged 65 and over. *The Journal of clinical endocrinology and metabolism.* 2009;94(4):1342-1345.
43. Baumgartner C, da Costa BR, Collet TH, et al. Thyroid Function Within the Normal Range, Subclinical Hypothyroidism, and the Risk of Atrial Fibrillation. *Circulation.* 2017;136(22):2100-2116.
44. Korevaar TIM, Chaker L, Peeters RP. Improving the clinical impact of randomised trials in thyroidology. *The lancet. Diabetes & endocrinology.* 2018;6(7):523-525.
45. Hennessey JV, Espaillat R. Diagnosis and Management of Subclinical Hypothyroidism in Elderly Adults: A Review of the Literature. *Journal of the American Geriatrics Society.* 2015;63(8):1663-1673.
46. Jonklaas J, Bianco AC, Bauer AJ, et al. Guidelines for the treatment of hypothyroidism: prepared by the american thyroid association task force on thyroid hormone replacement. *Thyroid : official journal of the American Thyroid Association.* 2014;24(12):1670-1751.
47. Malek M, Khamseh ME, Hadian A, Baradaran HR, Emami Z, Aghili R. The effect of L-thyroxine treatment on memory quotient in adults with subclinical hypothyroidism: A randomized double blind controlled trial. [Persian]. *Iranian Journal of Endocrinology and Metabolism.* 2012;13(6):624-629.
48. Watt T, Hegedus L, Groenvold M, et al. Validity and reliability of the novel thyroid-specific quality of life questionnaire, ThyPRO. *European journal of endocrinology / European Federation of Endocrine Societies.* 2010;162(1):161-167.
49. EuroQol G. EuroQol--a new facility for the measurement of health-related quality of life. *Health policy.* 1990;16(3):199-208.
50. Houx PJ, Shepherd J, Blauw GJ, et al. Testing cognitive function in elderly populations: the PROSPER study. PROspective Study of Pravastatin in the Elderly at Risk. *Journal of neurology, neurosurgery, and psychiatry.* 2002;73(4):385-389.
51. Beck AT, Ward CH, Mendelson M, Mock J, Erbaugh J. An inventory for measuring depression. *Archives of general psychiatry.* 1961;4:561-571.
52. Wechsler D. *Wechsler Memory Scale - Third edition manual.* San Antonio, TX: Psychological Corporation; 1997.
53. Zulewski H, Muller B, Exer P, Miserez AR, Staub JJ. Estimation of tissue hypothyroidism by a new clinical score: evaluation of patients with various grades of hypothyroidism and controls. *The Journal of clinical endocrinology and metabolism.* 1997;82(3):771-776.
54. Ciconelli RM, Ferraz MB, Santos W. Brazilian-Portuguese version of the SF-36. A reliable and valid quality of life outcome measure. *Rev Bras Rheumatol.* 1999(39):143-150.
55. Herrmann C. International experiences with the Hospital Anxiety and Depression Scale--a review of validation data and clinical results. *Journal of psychosomatic research.* 1997;42(1):17-41.
56. Folstein MF, Folstein SE, McHugh PR. "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. *Journal of psychiatric research.* 1975;12(3):189-198.
57. Golding E. *The Middlesex Elderly Assessment of Mental State.* Bury St. Edmonds, UK: Thames Valley Test; 1989.
58. Medical Research Council. *The speed and capacity of language-processing test.* Bury St. Edmunds, UK: Thames Valley Test; 1992.
59. Kortte KB, Horner MD, Windham WK. The trail making test, part B: cognitive flexibility or ability to maintain set? *Applied neuropsychology.* 2002;9(2):106-109.
60. McMillan CV, Bradley C, Woodcock A, Razvi S, Weaver JU. Design of new questionnaires to measure quality of life and treatment satisfaction in hypothyroidism. *Thyroid : official journal of the American Thyroid Association.* 2004;14(11):916-925.
61. Huppert FA, Walters DE, Day NE, Elliott BJ. The factor structure of the General Health Questionnaire (GHQ-30). A reliability study on 6317 community residents. *The British journal of psychiatry : the journal of mental science.* 1989;155:178-185.
62. Billewicz WZ, Chapman RS, Crooks J, et al. Statistical methods applied to the diagnosis of hypothyroidism. *The Quarterly journal of medicine.* 1969;38(150):255-266.
63. Higgins J, Green S. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. 2011; Available from www.cochrane-handbook.org.

Table 1: Characteristics of 21 included randomized clinical trials on thyroid hormone therapy for subclinical hypothyroidism in adults

Author, y	Country	Funding source	Definition of SCH	n	Mean age in years (SD)	n women (%)	Intervention	Control	Planned follow-up duration [in months]	Outcome[a]	Hypothyroid symptoms at baseline [Intervention vs Control (SD)]
Stott,[12] 2017	Netherlands, Switzerland, UK, Ireland	Non-industry	TSH 4.6-19.99mU/l on 2 occasions & normal fT4	737	74 (6.3)	396 (54%)	Levothyroxine	Placebo	At least 12[b]	- ThyPRO questionnaire[48] - EQ-5D score[49] - Letter-digit coding test[50] - Handgrip strength - Blood pressure - BMI - Cardiovascular events - Mortality - Side effects[48]	ThyPRO hypothyroid symptom score 17.5 (±18.8) vs 16.9 (±17.9)
Zhao,[11] 2016	China	Non-industry	TSH 4.2 – 10.0mU/l, normal fT4 on 2 occasions	369	55 (7.6)	270 (73%)	Levothyroxine	No intervention	15	- Blood pressure - BMI	nr
Najafi,[24] 2015	Iran	Non-industry	TSH >4.5mU/l, normal fT4, positive TPO-Ab	60	34 (10.0)	51 (85%)	Levothyroxine	Placebo	3	- BDI[51]	Mean number of hypothyroid symptoms per participant (range 0 to 12): 4.8 (±nr) vs 5.1 (±nr)
Ersoy,[29] 2012	Turkey	Not declared	TSH 5.0 – 10.0mU/l, normal fT4	60	46 (13.1)	58 (97%)	Levothyroxine	No intervention	6	- Blood pressure - BMI	nr
Aghili,[25] 2012	Iran	Non-industry	TSH >4.5mU/l, normal fT4, positive TPO-Ab	60	34 (10.8)	51 (85%)	Levothyroxine	Placebo	3	- Cognitive function (Wechsler memory scale[52])	Mean number of hypothyroid symptoms per participant (range 0 to 7): 3.2 (±nr) vs 3.7 (±nr)
Reuters,[31] 2012	Brazil	Not declared	TSH >4.0mU/l, normal fT4 on 2 occasions	71	50 (10.9)	62 (87%)	Levothyroxine	Placebo	6	- Zulewski score[53] - SF-36 score[54] - BDI[51] - Quadriceps strength	Zulewski score nr (only change from baseline reported)
Cabral,[28] 2011	Brazil	Not declared	TSH >4mU/l + normal fT4 on 2 occasions	32	46 (9.0)	32 (100%)	Levothyroxine	No intervention	12	- BMI[c]	nr
Parle,[16] 2010	UK	Non-industry	TSH >5.5mU/l + normal fT4	94	74 (5.8)	57 (61%)	Thyroxine	Placebo	12	- HADS[55]	nr

										- Cognitive function (MMSE,[56] MEAMS,[57] SCOLP,[58] Trail making test[59])	
Nagasaki,[36] 2009	Japan	Non-industry	Increased TSH, normal fT3/4	95	65 (19.3)	95 (100%)	Levothyroxine	Placebo	5	- Blood pressure - BMI	nr
Teixeira,[30] 2008	Brazil	Industry supported	TSH >4mU/l + fT4 normal on ≥2 occasions	60	48 (10.5)	57 (95%)	Levothyroxine	Placebo	12	- BMI	nr
Razvi,[32] 2007	UK	Non-industry	TSH >4mU/l + normal fT4 on ≥2 occasions	100	54 (12.6)	82 (82%)	Levothyroxine	Placebo	3	- ThyDQoL[60] - Blood pressure - BMI[c]	ThyDQoL nr (only change from baseline reported)
Jorde,[26] 2006	Norway	Non-industry	TSH 3.5-10mU/l	69	62 (11.9)	32 (46%)	Thyroxine	Placebo	12	- GHQ-30[61] - BDI[51] - Composite cognitive score[26]	Mean number of hypothyroid symptoms per participant (range 0 to 19): 4.0 (±nr) vs 4.0 (±nr)
Iqbal,[37] 2006	Norway	Non-industry	TSH 3.5-10mU/l on 2 occasions, fT3/4 normal	64	64 (12.2)	31 (48%)	Thyroxine	Placebo	12	- BMI	nr
Caraccio,[38] 2005	Italy	Non-industry	TSH >3.6mU/l, normal fT3/4	23	32 (9.6)	21 (91%)	Levothyroxine	Placebo	6	- BMI	nr
Yazici,[35] 2004	Turkey	Not declared	Increased TSH, normal fT3/4	45	40 (7.9)	38 (84%)	Levothyroxine	Placebo	12	- Blood pressure - BMI	nr
Monzani,[34] 2004	Italy	Not declared	TSH >3.6mU/l	45	37 (11.0)	37 (82%)	Levothyroxine	Placebo	6	- Blood pressure - BMI	nr
Kong,[27] 2002	UK	Not declared	TSH 5-10mU/l, fT4 normal	40	50 (15.2)	40 (100%)	Thyroxine	Placebo	6	- GHQ-30[61] - HADS[55] - BMI	Overall, 33/40 (83%) reported fatigue, 32/40 (80% reported weight gain)
Caraccio,[39] 2002	Italy	Non-industry	TSH >3.6mU/l on 2 occasions, positive TPO-Ab	49	35 (9.1)	42 (86%)	Levothyroxine	Placebo	6	- BMI	nr
Monzani,[33] 2001	Italy	Not declared	TSH >3.6mU/l for >1 year, normal fT4	20	32 (12.1)	18 (90%)	Levothyroxine	Placebo	6	- Blood pressure - BMI	nr
Meier,[22] 2001	Switzerland	Non-industry & industry supported[d]	TSH >5mU/l on 2 consecutive blood tests, fT4 normal	66	57 (10.6)	66 (100%)	Levothyroxine	Placebo	12	- Billewicz score[62]	Billewicz score -25.7 (±15.2) vs -28.3 (±14.1)
Cooper,[23] 1984	USA	Non-industry	Increased TSH, normal fT3/4	33	54 (10.1)	32 (97%)	Levothyroxine	Placebo	12	- BMI	Mean number of hypothyroid symptoms

													per participant (range 0 to 6): 2.1 (±nr) vs 2.4 (±nr)

Abbreviations: y, Year; **SCH**, Subclinical hypothyroidism; **n**, Number of participants; **SD**, Standard deviation; **ThyPRO**, Thyroid-related quality-of-life patient-reported outcome measure (hypothyroid symptoms (4 items, range 0 to 100, higher scores indicate more hypothyroid symptoms) and tiredness score (7 items)); **EQ-5D**, Euro quality of life 5 dimensions questionnaire; **Letter-digit coding test** (assesses executive cognitive function); **BMI**, Body-mass index; **nr**, not reported; BDI, Becks Depression Inventory; **TPO-Ab**, Thyroid peroxidase antibody; **SF-36**, Short Form (36) Health Survey; **HADS**, Hospital anxiety and depression scale; **MMSE**, Mini mental state examination; **MEAMS**, Middlesex elderly assessment of mental state; **SCOLP**, Speed and capacity of language processing test; **ThyDQoL**, 18-item underactive thyroid-dependent quality of life; **GHQ-30**, General health questionnaire 30 items; **Billewicz score** (range -47 to 67, higher scores indicates worse hypothyroid symptoms);.

[a] Only relevant outcomes for this systematic review are listed, i.e. outcomes that were included in the study protocol and published on PROSPERO

[b] The letter digit coding test was available after 18 months of levothyroxine/placebo intervention, the other outcomes after 12 months.

[c] Data obtained through direct communication with author

[d] This work was supported by the Swiss Research Foundation and an unconditional research grants from Henning Berlin, Sandoz Research, and Roche Research Foundations

Titles and legends for figures

Figure 1 title: Forest plots of randomized clinical trials on levothyroxine therapy in subclinical hypothyroidism, showing quality of life and mood-related outcomes

Figure 1 legend: Mean values of the quality of life / mood-related outcome scales per study group are shown in appendix eTable 2. Fixed effect meta-analysis of standardized mean differences; weights are from a fixed effect analysis. All effect sizes are standardized. As a rule of thumb for the interpretation, a standardized mean difference of 0.2, 0.5 and 0.8 correspond to small, moderate and large clinical effects, respectively.[63] For references to the range of the original scales see Table 1.

Figure 1 footnote: There is a difference between the number of participants randomized, and the number of participants with available outcome data in the studies of Kong,[27] Jorde,[26] Reuters,[31] Stott,[12] and Meier[22] (see Table 1 & eTable 2). Of note, the study by Razvi et al is a cross-over study, it included 100 participants.[32]

Figure 2 title: Forest plot of randomized clinical trials on levothyroxine therapy in subclinical hypothyroidism, showing outcomes on cognitive function

Figure 2 legend: Mean values of the cognition scale per study group are shown in appendix eTable 2. Fixed effect meta-analysis of standardized mean differences; weights are from a fixed effect analysis; dashed line represents the overall mean effect. All effect sizes are standardized. As a rule of thumb for the interpretation, a standardized standardized mean difference of 0.2, 0.5 and 0.8 correspond to small, moderate and large clinical effects, respectively.[63] For references to the range of the original scales see Table 1.

Figure 2 footnote: There is a difference between the number of participants randomized, and the number of participants with available outcome data in the studies of Jorde,[26] and Stott[12] (see Table 1 & eTable 2).

Figure 3 title: Forest plots of randomized clinical trials on levothyroxine therapy in subclinical hypothyroidism, showing outcomes on systolic blood pressure

Figure 3 legend: Fixed effect meta-analysis of differences in blood pressure (mmHg); weights are from a fixed effect analysis; dashed line represents the overall mean effect.

Figure 3 footnote: There is a difference between the number of participants randomized, and the number of participants with available outcome data in the study of Stott[12] (see Table 1 & eTable 2). Of note, the study by Razvi et al is a cross-over study, it included 100 participants.[32]

Figure 4 title: Forest plots of randomized clinical trials on levothyroxine therapy in subclinical hypothyroidism, showing outcomes on body-mass index

Figure 4 legend: Random effects meta-analysis of differences in BMI (kg/m^2); weights are from a random effects analysis; dashed line represents the overall mean effect.

Figure 4 footnote: There is a difference between the number of participants randomized, and the number of participants with available outcome data in the studies of Kong,[27] Teixeira,[30] and Stott[12] (see Table 1 & eTable 2). Of note, the study by Razvi et al is a cross-over study, it included 100 participants.[32]