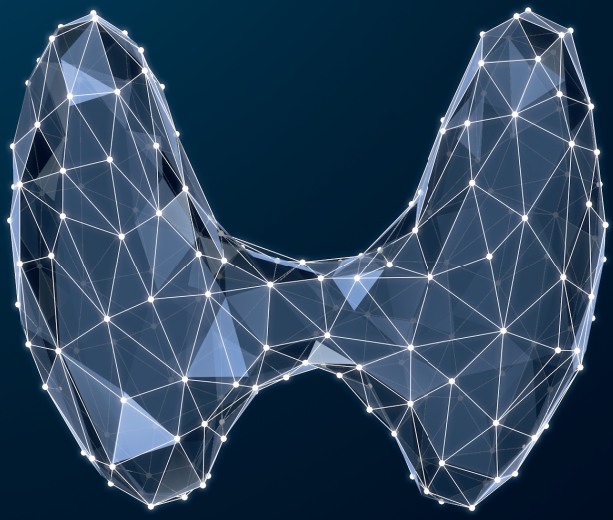
Subclinical hypothyroidism in community-dwelling older people: consequences and treatment outcomes

Du Puy, R.S.



Chapter 5

Thyroid hormone therapy for older adults with subclinical hypothyroidism

David J Stott
Nicolas Rodondi
Patricia M Kearney
Ian Ford
Rudi GJ Westendorp
Simon P Mooijaart
Naveed Sattar
Carole E Aubert
Drahomir A Aujesky
Doug C Bauer
Christine Baumgartner
Manuel R Blum
John P Browne
Stephen Byrne
Tinh-Hai Collet
Olaf M Dekkers
Wendy PJ den Elzen
Robert S Du Puy
Graham Ellis
Martin Feller
Carmen Floriani
Kirsty Hendry
Caroline Hurley
J Wouter Jukema
Sharon Kean
Maria Kelly
Danielle Krebs
Peter Langhorne
Gemma McCarthy
Vera McCarthy
Alex McConnachie
Mairi McDade
Martina Messow
Annemarie O'Flynn
David O'Riordan
Rosalinde KE Poortvliet
Terry J Quinn
Audrey Russell
Carol Sinnott
Jan WA Smit
H Anette Van Dorland
Kieran A Walsh
Elaine K Walsh
Torquil Watt
Robbie Wilson
Jacobijn Gussekloo for the
TRUST Study Group

New England Journal of Medicine, 2017;376(26):2534-2544

DOI: 10.1056/NEJMoa1603825

ABSTRACT

Background

The use of levothyroxine to treat subclinical hypothyroidism is controversial. We aimed to determine whether levothyroxine provided clinical benefits in older persons with this condition.

Methods

We conducted a double-blind, randomised, placebo-controlled, parallel-group trial involving 737 adults who were at least 65 years of age and who had persisting subclinical hypothyroidism (thyrotropin level, 4.60 to 19.99 mIU per litre; free thyroxine level within the reference range). A total of 368 patients were assigned to receive levothyroxine (at a starting dose of 50 µg daily, or 25 µg if the body weight was < 50 kg or the patient had coronary heart disease), with dose adjustment according to the thyrotropin level; 369 patients were assigned to receive placebo with mock dose adjustment. The two primary outcomes were the change in the Hypothyroid Symptoms score and Tiredness score on a thyroid-related quality-of-life questionnaire at 1 year (range of each scale is 0 to 100, with higher scores indicating more symptoms or tiredness, respectively; minimum clinically important difference, 9 points).

Results

The mean age of the patients was 74.4 years, and 396 patients (53.7%) were women. The mean (±SD) thyrotropin level was 6.40±2.01 mIU per litre at baseline; at 1 year, this level had decreased to 5.48 mIU per litre in the placebo group, as compared with 3.63 mIU per litre in the levothyroxine group ($p < 0.001$), at a median dose of 50 µg. We found no differences in the mean change at 1 year in the Hypothyroid Symptoms score (0.2±15.3 in the placebo group and 0.2±14.4 in the levothyroxine group; between-group difference, 0.0; 95% confidence interval [CI], −2.0 to 2.1) or the Tiredness score (3.2±17.7 and 3.8±18.4, respectively; between-group difference, 0.4; 95% CI, −2.1 to 2.9). No beneficial effects of levothyroxine were seen on secondary-outcome measures. There was no significant excess of serious adverse events prespecified as being of special interest.

Conclusions

Levothyroxine provided no apparent benefits in older persons with subclinical hypothyroidism. (Funded by European Union FP7 and others; TRUST ClinicalTrials.gov number, NCT01660126.)

INTRODUCTION

Subclinical hypothyroidism is defined as an elevated serum thyrotropin level and a serum free thyroxine level within the reference range.[1] Between 8% and 18% of adults 65 years of age or older have these biochemical features, and the prevalence is higher among women than among men.[2]

Subclinical hypothyroidism is a possible contributor to many problems in older persons. Thyroid hormones have multiple effects, since they act as an essential regulatory factor in numerous physiological systems, including the vascular tree and the heart,[3] the brain (including cognition),[4] skeletal muscle, and bone.[5] Tiredness is the most important symptom of overt hypothyroidism,[6] but most patients with subclinical hypothyroidism have no symptoms or have nonspecific symptoms.[7] There is a convincing epidemiologic association with subsequent coronary heart disease.[8]

Randomised, controlled trials of levothyroxine replacement for the treatment of subclinical hypothyroidism have been small [9,10] and have yielded only limited evidence regarding the possible benefits and risks of treatment.[1] We aimed to determine whether there are clinical benefits from levothyroxine replacement in older persons with subclinical hypothyroidism.

METHODS

Trial overview

The trial protocol, which was published previously [11] and is available with the full text of this article at NEJM.org, was approved by the relevant ethics committees and regulatory authorities in all the countries involved in the trial. Participants provided written informed consent.

The trial was conducted in accordance with the principles of the Declaration of Helsinki [12] and Good Clinical Practice guidelines.[13] The Robertson Centre for Biostatistics at the University of Glasgow was the trial data and biostatistics centre.

The European Union FP7 provided primary financial support for the conduct of the trial. Supplies of levothyroxine and matching placebo were provided free of charge by Merck (Darmstadt, Germany). The funder, the trial sponsors (NHS Greater Glasgow and Clyde Health Board and University of Glasgow, United Kingdom; University College Cork, Ireland; Leiden University Medical Center, the Netherlands; and University of Bern and Bern University Hospital, Switzerland), and Merck played no role in the design, analysis, or reporting of the trial.

The main sponsor (NHS Greater Glasgow and Clyde Health Board) contributed to the writing of the protocol. None of the sponsors had any involvement in the analysis or the reporting of the results. The authors vouch for the accuracy and completeness of the data and analyses reported and for the fidelity of the trial to the protocol.

Participants

Participants were identified from clinical laboratory and general practice databases and records. The inclusion criteria were an age of 65 years or more and persistent subclinical hypothyroidism, defined as an elevated thyrotropin level (4.60 to 19.99 mIU per litre) that was measured on at least two occasions that were 3 months to 3 years apart, with the free thyroxine level within the reference range. The main exclusion criteria for the trial were a current prescription for levothyroxine, antithyroid drugs, amiodarone, or lithium; thyroid surgery or receipt of radioactive iodine within the previous 12 months; dementia; hospitalization for a major illness or an elective surgery within the previous 4 weeks; an acute coronary syndrome (including myocardial infarction or unstable angina) within the previous 4 weeks; and terminal illness.[11]

Trial design and regimen

We conducted a randomised, double-blind, parallel-group trial of levothyroxine versus placebo. Patients underwent randomization in a 1:1 ratio, with stratification according to country, sex, and starting dose, with the use of randomly permuted blocks.

The active intervention started with levothyroxine at a dose of 50 µg daily (or 25 µg in patients with a body weight of < 50 kg or with known coronary heart disease [previous myocardial infarction or symptoms of angina pectoris]) or matching placebo. Dose adjustment in the levothyroxine group was aimed to result in a thyrotropin level within the reference range (0.40 to 4.59 mIU per litre). Details regarding how the dose was adjusted and the mock adjustment in the placebo group are provided in the Supplementary Appendix, available at NEJM.org. All dose adjustments were generated and executed by means of computer without the intervention of a physician. The participants, investigators, and treating physicians were unaware of the results of thyrotropin measurements throughout the course of the trial.

Procedures and outcomes

The two primary outcomes for the trial were the change from baseline to 12 months in the Thyroid-Related Quality-of-Life Patient-Reported Outcome measure (ThyPRO) Hypothyroid Symptoms score (4 items) and Tiredness score (7 items); each scale ranges from 0 to 100, with higher scores indicating more symptoms and tiredness, respectively.[14] A recent systematic review recommended ThyPRO as the preferred measurement tool for the assessment of health-related quality of life in patients with benign thyroid disease.[15] The ThyPRO and

other instruments were administered in English, French, German, or Dutch as appropriate. We had initially planned for cardiovascular events and thyroid-specific quality of life to be the two primary outcomes. However, this plan was modified during the trial to thyroid-specific quality-of-life scores as the two primary outcomes and cardiovascular events as a secondary outcome when it became apparent that the trial would be underpowered for cardiovascular events owing to delays and difficulties in recruitment.[11]

The secondary outcomes included changes from baseline in generic health-related quality of life (as assessed by the EuroQoL [EQ] Group 5-Dimension Self-Report Questionnaire [EQ-5D]; scores on the EQ-5D descriptive index range from −0.59 to 1.00, and scores on the EQ visual-analogue scale range from 0 to 100, with higher scores indicating better quality of life),[16] comprehensive thyroid-related quality of life (as assessed by the ThyPRO-39 score, a shorter version of the ThyPRO measure,[17] at final follow-up only), hand-grip strength (as assessed by means of the Jamar isometric dynamometer, with the recorded score as the best of three measures in the dominant hand),[18] executive cognitive function (as assessed with the letter–digit coding test, which indicates the speed of processing according to the number of correct responses in matching nine letters with nine digits in 90 seconds; minimum score, 0, with higher scores indicating better executive cognitive function; there is no maximum score),[19] blood pressure (systolic and diastolic), weight, body-mass index, waist circumference, activities of daily living (as assessed by the Barthel Index of functional levels in activities of daily living, on a scale ranging from 0 to 20, with higher scores indicating better performance),[20] the Instrumental Activities of Daily Living score (on a scale from 0 to 14, with higher scores indicating better performance in activities of daily living),[21] and fatal and nonfatal cardiovascular events. The minimum follow-up was 1 year, and the maximum follow-up was 3 years.

Safety and recording of adverse events

Adverse events were assessed, managed, recorded, reported, and analysed in accordance with the Medicines for Human Use (Clinical Trials) Regulations 2004 (as amended). Adverse events of special interest included new atrial fibrillation, heart failure, fracture, and new diagnosis of osteoporosis. The score on the ThyPRO Hyperthyroid Symptoms scale was recorded as a measure of possible adverse effects (on a scale from 0 to 100, with higher scores indicating more symptoms; minimum clinically important difference has been estimated as 9 points).[14]

Statistical analysis

The Hypothyroid Symptoms and Tiredness scores from the ThyPRO14 were the two primary outcomes, with the required p value for statistical significance split equally to each test (0.05/2=0.025 for each test). We assumed standard deviations for data at 1 year of 13.3 and

18.3 on the 100-point scales, respectively, after adjustment for baseline values. These calculations provided the trial with 80% power to detect a change with levothyroxine treatment (vs. placebo) of 3.0 points on the Hypothyroid Symptoms score and 4.1 points on the Tiredness score with our revised maximum expected number of recruited participants of 750, and with changes of 3.5 points and 4.9 points, respectively, with our minimum expected number of 540 participants. Justification for these power calculations is provided in the trial protocol.[11]

The methods of analysis of the continuous efficacy outcomes involving measurements at baseline and follow-up were analysed at each time point for the comparison of the two trial groups, with adjustment for stratification variables (country, sex, and starting dose of levothyroxine) and baseline levels of the same variable with the use of multivariate linear regression (see the Supplementary Appendix). The efficacy and safety analyses were carried out in a modified intention-to-treat population, which included participants with data on the outcome of interest. Patients who discontinued the trial regimen continued to be followed for the modified intention-to-treat analysis. These analyses were supported with sensitivity analyses that used mixed-effects models and multiple imputations for missing data. The primary and secondary outcomes at 12 months were also analysed in prespecified subgroups according to sex and baseline thyrotropin level.[11] Analyses were repeated in the per-protocol population, which included participants who continued to take the trial regimen per the trial protocol.

RESULTS

We screened 2,647 community-dwelling persons who were at least 65 years of age and who were identified as having biochemical subclinical hypothyroidism. A total of 737 participants underwent randomization, 369 of whom were assigned to receive placebo and 368 to receive levothyroxine (Figure 1). The characteristics at baseline were similar in the two groups (Table 1, and Supplemental table 1). The mean age of the patients was 74.4 years, and 396 patients (53.7%) were women. A score of 0 (indicating no symptoms) at baseline was observed in 199 of 737 participants (27.0%) on the Hypothyroid Symptoms scale and in 64 (8.7%) on the Tiredness scale; 36 participants (4.9%) had a score of 0 in both domains.

A total of 337 participants (91.3%) who were randomly assigned to the placebo group completed 12-month follow-up, as did 332 (90.2%) in the levothyroxine group. The median follow-up for all the participants who underwent randomization (including participants who discontinued the trial regimen) was 17.3 months (interquartile range, 12.0 to 24.4) in the placebo group and 18.0 months (interquartile range, 11.0 to 25.4) in the levothyroxine group. The median dose of levothyroxine at 1 year was 50 µg. The numbers of patients who were included in the analyses are presented in Figure 1.

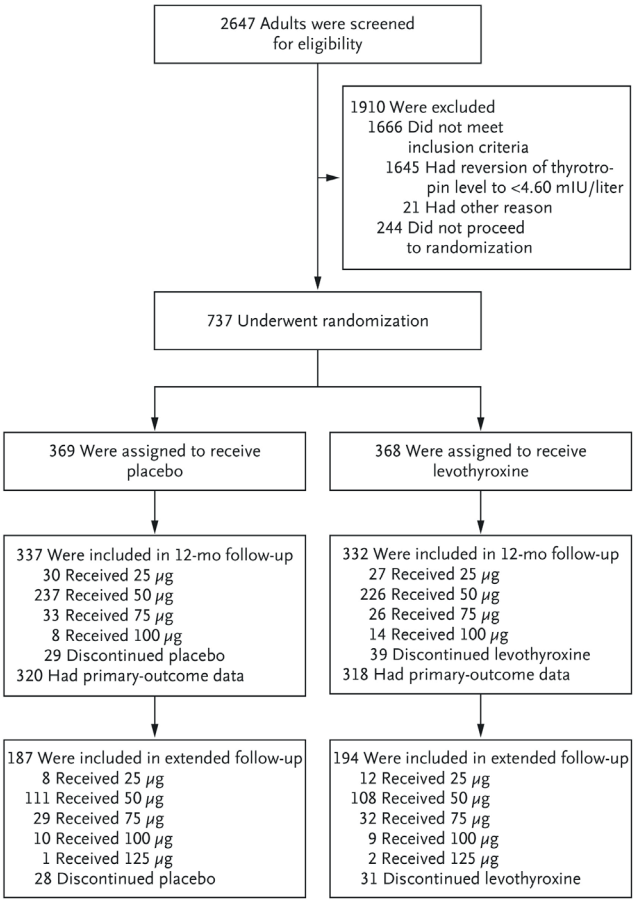


Figure 1. Randomization, Follow-up, and Dose Levels. Exclusions for other reasons included use of antithyroid medication (in 17 persons), recent thyroid surgery (in 1), recent acute coronary syndrome (in 1), current participation in another trial (in 1), and adrenal insufficiency (in 1). Two patients who were excluded because the thyrotropin level reverted to less than 4.60 mIU per litre also had an additional exclusion of galactose intolerance. Extended follow-up beyond 12 months was conducted in a subgroup of patients, with a median duration of follow-up from baseline of 24.2 months (interquartile range, 18.4 to 30.3) in the placebo group and 24.5 months (interquartile range, 18.4 to 30.5) in the levothyroxine group.

Table 1. Characteristics of the participants at baseline.

Characteristic[a]	Placebo Group (n = 369)	Levothyroxine Group (n = 368)
Age — yr		
Mean	74.8±6.8	74.0±5.8
Range	65.1–93.4	65.2–93.0
Female sex — no. (%)	198 (53.7)	198 (53.8)
White race — no. (%)[b]	362 (98.1)	362 (98.4)
Standard housing — no. (%)[c]	356 (96.5)	358 (97.3)
Previous medical conditions and clinical descriptors — no./total no.(%)		
Ischemic heart disease[d]	50/369 (13.6)	50/368 (13.6)
Atrial fibrillation	44/368 (12.0)	45/364 (12.4)
Hypertension	183/366 (50.0)	192/368 (52.2)
Diabetes mellitus	54/368 (14.7)	63/368 (17.1)
Osteoporosis	47/367 (12.8)	41/364 (11.3)
Current smoking	33/369 (8.9)	29/368 (7.9)
Median no. of concomitant medicines (IQR)	4 (2–6)	4 (2–6)
Median Mini–Mental State Examination score (IQR)[e]	29 (28–30)	29 (27–30)
Weight < 50 kg — no. (%)	5 (1.4)	5 (1.4)
Laboratory results		
Thyrotropin — mIU/litre	6.38±2.01	6.41±2.01
Median (IQR)	5.76 (5.10–6.94)	5.73 (5.12–6.83)
Range	4.60–17.60	4.60–17.60
Free thyroxine — pmol/litre[f]	13.3±1.9	13.4±2.1
Outcome measures[g]		
Hypothyroid Symptoms score	16.9±17.9	17.5±18.8
Tiredness score	25.5±20.3	25.9±20.6
EQ-5D descriptive index	0.847±0.171	0.846±0.187
EQ visual-analogue scale score	76.5±16.3	78.4±15.3
Hand-grip strength — kg	27.5±11.3	28.0±10.2
Letter–digit coding test score	25.2±8.3	24.9±7.4
Blood pressure — mm Hg		
Systolic	140.4±18.9	141.2±18.7
Diastolic	74.8±11.7	74.1±11.6
Body-mass index	27.7±4.6	28.1±5.3
Waist circumference — cm	97.5±12.8	98.5±13.6
Median Barthel Index (IQR)	20 (14–20)	20 (13–20)
Median Instrumental Activities of Daily Living score (IQR)	14 (7–14)	14 (7–14)

[a] Plus–minus values are means ±SD. There were no significant between-group differences in the baseline characteristics. IQR denotes interquartile range.

[b] Race was reported by the patient.

[c] Standard housing was defined as nonsheltered community accommodation. By contrast, sheltered housing is purpose-built grouped housing for older persons, often with an on-site manager or warden.

[d] Ischemic heart disease was defined as a history of angina pectoris or previous myocardial infarction.

[e] The Mini–Mental State Examination score is on a scale from 0 to 30, with higher scores indicating better cognitive function.

[f] To convert the values for free thyroxine to nanograms per decilitre, divide by 12.87.

[g] The Hypothyroid Symptoms score and the Tiredness score from the Thyroid-Related Quality of Life Patient-Reported Outcome (ThyPRO) questionnaire are each assessed on a scale from 0 to 100, with higher scores indicating more symptoms and tiredness, respectively. The minimum clinically important difference for each score has been estimated as 9 points. The EuroQoL [EQ] Group 5-Dimension Self-Report Questionnaire (EQ-5D) scores included both the EQ- 5D descriptive index (on a scale from −0.59 to 1.00) and the score on the EQ visual-analogue scale (on a scale from 0 to 100); higher scores on each scale indicate better quality of life. The score on the letter–digit coding test (a test of executive cognitive function) indicates the speed of processing according to the number of correct responses in matching nine letters with nine digits in 90 seconds (minimum score is 0, with higher scores indicating better executive cognitive function; there is no maximum score). The body-mass index is the weight in kilograms divided by the square of the height in meters. The Barthel Index uses a scale from 0 to 20 points, with higher numbers indicating better performance on activities of daily living. The Instrumental Activities of Daily Living scale has a maximum score of 14 (range, 0 to 14), with higher scores indicating better performance in activities of daily living.

Thyroid-function tests

The mean (±SD) thyrotropin level at baseline was 6.40±2.01 mIU per litre. The thyrotropin levels were reduced from baseline to a greater extent in the levothyroxine group than in the placebo group at all time points of review, with a mean between-group difference of 2.29 mIU per litre at 6 to 8 weeks after randomization ($p < 0.001$) (Supplemental table 2). At 12 months, the mean thyrotropin level was 5.48±2.48 mIU per litre in the placebo group, as compared with 3.63±2.11 mIU per litre in the levothyroxine group, resulting in a between-group difference of 1.92 mIU per litre ($p < 0.001$) (Table 2 and Figure 2). There was a significant interaction between the trial group and the office visit ($p=0.03$), with a reduction in the thyrotropin level being the greatest at 6 to 8 weeks.

Free thyroxine levels were not routinely measured, although the data were available in a subgroup of patients. The mean free thyroxine level was 2.3 pmol per litre (0.2 ng per decilitre) higher in the levothyroxine group than in the placebo group both at 6 to 8 weeks and at 12 months ($p < 0.001$ for both comparisons) (Supplemental table 3).

Table 2. Outcomes at 12 months and extended follow-up.[a]

Variable	Baseline		At 12 Mo				At Extended Follow-up Visit[b]			
	Placebo (N = 369)	Levothyroxine (N = 368)	Placebo (N = 320)	Levothyroxine (N = 318)	Difference (95% CI)	P Value	Placebo (N = 187)	Levothyroxine (N = 194)	Difference (95% CI)	P Value
Thyrotropin — mIU/litre	6.38±2.01	6.41±2.01	5.48±2.48	3.63±2.11	−1.92 (−2.24 to −1.59)	<0.001	5.28±2.50	3.47±2.08	−1.88 (−2.32 to −1.45)	<0.001
Median (IQR)	5.76 (5.10 to 6.94)	5.70 (5.12 to 6.83)	4.90 (3.91 to 6.46)	3.16 (2.45 to 4.22)	—	—	4.94 (3.78 to 6.26)	3.00 (2.26 to 4.16)	—	—
Primary outcomes[c]										
Hypothyroid Symptoms score	16.9±17.9	17.5±18.8	16.7±17.5	16.6±16.9	0.0 (−2.0 to 2.1)	0.99	15.2±15.9	17.9±9.1	1.0 (−1.9 to 3.9)	0.50
Tiredness score	25.5±20.3	25.9±20.6	28.6±19.5	28.7±20.2	0.4 (−2.1 to 2.9)	0.77	31.9±22.1	30.2±20.5	−3.5 (−7.0 to 0.0)	0.05
Secondary outcomes										
EQ-5D descriptive score	0.847±0.171	0.846±0.187	0.853±0.191	0.833±0.212	−0.025 (−0.050 to 0.000)	0.05	0.829±0.209	0.864±0.188	0.040 (0.005 to 0.075)	0.03
EQ VAS score	76.5±16.3	78.4±15.3	77.4±13.7	77.3±15.6	−1.3 (−3.2 to 0.6)	0.18	77.2±13.5	76.8±14.2	−0.8 (−3.2 to 1.7)	0.56
Hand-grip strength — kg	27.5±11.3	28.0±10.2	27.1±11.2	27.5±10.5	−0.1 (−0.9 to 0.7)	0.84	24.9±10.6	24.4±10.1	−0.6 (−1.7 to 0.6)	0.34
Blood pressure — mm Hg										
Systolic	140.4±18.9	141.2±18.7	138.4±17.8	138.3±18.7	0.1 (−2.1 to 2.4)	0.90	137.5±19.2	136.8±17.6	1.1 (−4.1 to 2.1)	0.51
Diastolic	74.8±11.7	74.1±11.6	73.5±11.1	72.8±11.4	−0.1 (−1.5 to 1.3)	0.93	72.3±11.4	72.0±11.5	0.5 (−1.4 to 2.4)	0.59
Body-mass index	27.7±4.6	28.1±5.3	27.7±4.6	27.9±5.1	0.0 (−0.2 to 0.2)	0.89	27.2±4.5	27.9±4.9	0.2 (−0.1 to 0.5)	0.30

Table 2. Outcomes at 12 months and extended follow-up.[a] (*continued*)

Variable	Baseline		At 12 Mo				At Extended Follow-up Visit[b]			
	Placebo (N = 369)	Levothyroxine (N = 368)	Placebo (N = 320)	Levothyroxine (N = 318)	Difference (95% CI)	P Value	Placebo (N = 187)	Levothyroxine (N = 194)	Difference (95% CI)	P Value
Waist circumference — cm	97.5±12.8	98.5±13.6	96.8±13.1	98.0±13.2	0.4 (−0.4 to 1.3)	0.34	96.0±13.8	97.6±13.4	0.3 (−0.9 to 1.5)	0.66
Adverse symptom assessment										
Hyperthyroid Symptoms score[d]	10.5±11.2	10.5±11.2	10.3±11.3	10.5±10.8	0.6 (−0.7 to 1.9)	0.35	9.8±11.0	11.1±11.7	0.7 (−1.2 to 2.5)	0.46

[a] All these analyses were conducted in the modified intention-to-treat population, which included all the participants who had undergone randomization (excluding those who had under- gone randomization in error) for whom data was available on the outcome of interest. For analyses at the 12-month visit to be valid, they must have been conducted at 12 months (within a ±31-day window) after randomization. Results at 12 months, at the extended follow-up visit, and between-group differences are adjusted for stratification variables (country, sex, and starting dose of levothyroxine) and baseline levels of the same variable with the use of linear regression. Between-group differences are the value in the levothyroxine group minus the value in the placebo group. Data for the extended follow-up visit were additionally adjusted for time to visit. Data were missing for the following outcomes: for the thyrotropin level at 12 months for 7 patients in the placebo group and 1 in the levothyroxine group and at extended follow-up for 7 in the placebo group and 6 in the levothyroxine group; for the EQ-5D score at extended follow-up for 1 in the levothyroxine group; for the EQ visual-analogue scale (VAS) score at 12 months for 1 in the placebo group and at extended follow-up for 1 in the levothyroxine group; for hand-grip strength at baseline for 11 in the placebo group and 10 in the levothyroxine group, at 12 months for 22 in the placebo group and 16 in the levothyroxine group, and at extended follow-up for 14 in the placebo group and 8 in the levothyroxine group; for blood pressure at baseline and at 12 months for 1 in the placebo group and at extended follow-up for 5 in each group; on body-mass index at baseline for 1 in each group, at 12 months for 2 in the placebo group and 1 in the levothyroxine group, and at extended follow-up for 2 in the placebo group and 4 in the levothyroxine group; and for waist circumference at baseline for 1 in each group, at 12 months for 1 in the placebo group and 2 in the levothyroxine group, and at extended follow-up for 2 in the placebo group and 4 in the levothyroxine group. CI denotes confidence interval.

[b] Extended follow-up beyond 12 months was performed in a subgroup of patients. The median duration of follow-up from baseline was 24.2 months (interquartile range, 18.4 to 30.3) in the placebo group and 24.5 months (interquartile range, 18.4 to 30.5) in the levothyroxine group.

[c] The two primary outcomes were the Hypothyroid Symptoms score and the Tiredness score from the ThyPRO questionnaire at 12 months (adjusted as stated above). The range of each scale is 0 to 100, with higher scores indicating more symptoms. The minimum clinically important difference for each score has been estimated as 9 points.

[d] The score on the Hyperthyroid Symptoms scale was recorded as a measure of possible adverse effects (on a scale from 0 to 100, with higher scores indicating more symptoms; minimum clinically important difference has been estimated as 9 points).

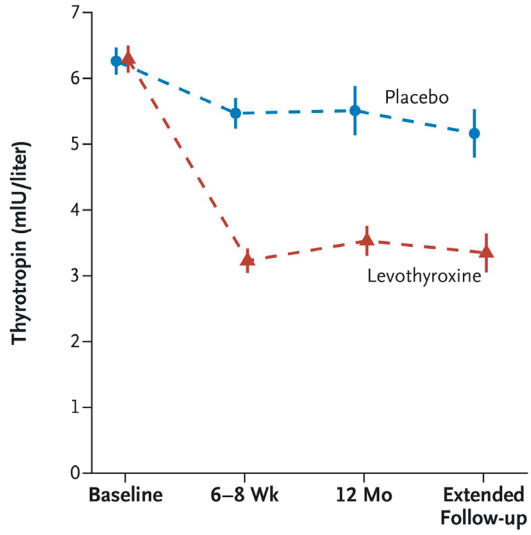


Figure 2. Thyrotropin levels in the Placebo Group and Levothyroxine Group. Shown are the results of a modified intention-to-treat analysis. Data are means, and error bars indicate 95% confidence intervals. Extended follow-up beyond 12 months was conducted in a subgroup of patients, with a median duration of follow-up from baseline of 24.2 months (interquartile range, 18.4 to 30.3) in the placebo group and 24.5 months (interquartile range, 18.4 to 30.5) in the levothyroxine group. $p < 0.001$ for between-group differences in the thyrotropin level at 6 to 8 weeks, 12 months, and extended follow-up. Analyses were adjusted for stratification variables (country, sex, and starting dose of levothyroxine) and baseline thyrotropin level with the use of linear regression; data for the extended follow-up visit were additionally adjusted for time to visit.

Thyroid-specific quality of life

The mean Hypothyroid Symptoms score at 12 months (with adjustment for baseline score) was 16.7±17.5 in the placebo group and 16.6±16.9 in the levothyroxine group (p=0.99). The mean Tiredness score was 28.6±19.5 in the placebo group and 28.7±20.2 in the levothyroxine group (p=0.77). We found no differences in the mean change at 1 year in the Hypothyroid Symptoms score (0.2±15.3 in the placebo group and 0.2±14.4 in the levothyroxine group) or the Tiredness score (3.2±17.7 and 3.8±18.4, respectively) (Table 2). There were no significant between-group differences in either of these measures at 6 to 8 weeks (Supplemental table 4). There was a small-magnitude between-group difference in the Tiredness score, with a lower value in the levothyroxine group than in the placebo group (difference, −3.49; p=0.05) at the extended follow-up review (Table 2). Prespecified analyses according to sex and baseline thyrotropin level did not reveal any subgroups of patients who benefited from treatment with levothyroxine. Per-protocol analyses and sensitivity analyses with the use of multiple imputation of missing values showed no significant differences between the levothyroxine group and the placebo group (Supplemental tables 4 and 5).

Other outcome measures

The EQ-5D descriptive index showed a small deterioration at 12 months (mean difference between the levothyroxine group and the placebo group, −0.025; p=0.05) but a minor improvement at extended follow-up (mean difference, 0.040; p=0.03); there were no significant between-group differences at 6 to 8 weeks. There were no significant between-group differences in the score on the EQ visual-analogue scale (Table 2, and Supplemental table 2).

No significant effects were seen in any of the other secondary-outcome measures, either in the modified intention-to-treat or per-protocol analyses or in the prespecified subgroups (Table 2, and Supplemental tables 4, 6, 7, and 8). Results regarding cardiovascular events and total and cardiovascular mortality are provided in Table 3 and in Supplemental figures 1 and 2.

Table 3. Clinical outcomes and adverse events.[a]

Variable	All Patients (N = 737)	Placebo Group (N = 369)	Levothyroxine Group (N = 368)	Hazard Ratio (95% CI)
Clinical outcome				
Fatal or nonfatal cardiovascular event — no. (%)	38 (5.2)	20 (5.4)	18 (4.9)	0.89 (0.47–1.69)
Cardiovascular death — no. (%)	3 (0.4)	1 (0.3)	2 (0.5)	—
Death from any cause — no. (%)	15 (2.0)	5 (1.4)	10 (2.7)	1.91 (0.65–5.60)
Serious adverse event				
No. of patients with ≥ 1 serious adverse event	181 (24.6)	103 (27.9)	78 (21.2)	0.94 (0.88–1.00)[b]
No. of events	343	201	142	—
Adverse event of special interest				
New-onset atrial fibrillation — no. (%)	24 (3.3)	13 (3.5)	11 (3.0)	0.80 (0.35–1.80)
Heart failure — no. (%)	9 (1.2)	6 (1.6)	3 (0.8)	—
Fracture — no. (%)	17 (2.3)	8 (2.2)	9 (2.4)	1.06 (0.41–2.76)
New diagnosis of osteoporosis — no. (%)	7 (0.9)	4 (1.1)	3 (0.8)	—
Withdrawal				
Permanent discontinuation of trial regimen — no.(%)	160 (21.7)	79 (21.4)	81 (22.0)	1.06 (0.78–1.44)
Withdrawal from follow-up — no. (%)	41 (5.6)	22 (6.0)	19 (5.2)	0.84 (0.46–1.56)

[a] This table includes serious adverse events and adverse events of special interest in the modified intention-to-treat population and data on withdrawals from trial regimen and follow-up. Hazard ratios were not calculated for cardiovascular death, heart failure, or new diagnosis of osteoporosis owing to the small number of events.

[b] p = 0.05. Hazard ratios for treatment were obtained from a Cox proportional-hazards regression model predicting survival from randomised trial group and stratification variables (country, sex, and dose at randomization).

Adverse effects and events

We found no significant difference in the Hyperthyroid Symptoms score (according to the ThyPRO assessment) with levothyroxine, as compared with placebo, at any time point (Table 2, and Supplemental table 2). The incidence of serious adverse events of special interest (atrial fibrillation, heart failure, fracture, or new diagnosis of osteoporosis) was similar in the two groups (Table 3). The number of patients with at least one serious adverse event was slightly higher in the placebo group than in the levothyroxine group (p=0.049), as was the total number of serious adverse events. However, we observed no pattern of event type that contributed to this difference. The proportions of patients who discontinued the trial regimen or who withdrew from follow-up were similar in the two groups (Table 3).

DISCUSSION

In this multicentre, double-blind, randomised, placebo-controlled, parallel-group trial involving older participants with subclinical hypothyroidism, treatment with levothyroxine was associated with a persistently lower serum thyrotropin level than was placebo (between-group difference, approximately 2 mIU per litre), with the maximum effects seen at time of first review (6 to 8 weeks). We found that levothyroxine had no consistent beneficial effect on thyroid-related symptoms. This finding was true in both older men and older women and for different thyrotropin levels at baseline. Our trial had good statistical power to detect a clinically meaningful effect on thyroid-related quality of life, with 95% confidence intervals that excluded a beneficial effect greater than 2.1 points (on a scale from 0 to 100) in either of the two primary outcomes. If a symptom benefit was to have occurred, it would have been expected to be seen at 12 months.

The subsequent small-magnitude between-group difference in tiredness with levothyroxine versus placebo in the subgroup of patients who had extended follow-up is likely to be a chance finding. In contrast, an observational study of the treatment of autoimmune hypothyroidism in middle-age participants (median baseline thyrotropin level, 8.1 mIU per litre) showed that the Tiredness score improved markedly (reduction of 12 points at 6 months) and that the Hypothyroid Symptoms score also was reduced (by 2 points).[22] A small reduction in tiredness has previously been shown in a short-term trial of levothyroxine for the treatment of subclinical hypothyroidism in 120 middle-age participants.[23] There are limited data from high-quality, randomised, controlled trials regarding the effects of levothyroxine replacement in older persons with subclinical hypothyroidism.[1] Studies have generally been small (≤ 120 participants) and underpowered, often focusing on younger participants and with a short duration of follow-up.[9,10]

Levothyroxine treatment yielded no significant beneficial effects on a range of secondary-outcome measures. We found a slight deterioration (of borderline statistical significance) in the EQ-5D descriptive index with levothyroxine versus placebo at 12 months but an improvement versus placebo in the subgroup of patients who completed extended follow-up (median, 24.5 months). The effects we observed were in opposite directions at these different time points and were of very small magnitude (−0.025 at 12 months and 0.040 at extended follow-up), and therefore these are likely to be random chance findings. The estimated minimally important difference in the EQ-5D descriptive index that has been reported for other conditions is summarised in a recent review as being between 0.037 and 0.069.[24] No effect of treatment was seen with regard to the EQ visual-analogue scale scores. Therefore, it appears that levothyroxine had no clinically significant effects on generic health-related quality of life.

Muscle function has been described as being adversely affected by underactive thyroid.[25] However, we found that hand-grip strength did not change from baseline significantly more with levothyroxine treatment than with placebo. Similarly, it has been suggested that the speed of information processing is slowed in persons with subclinical hypothyroidism.[4] However, we found no benefit with levothyroxine with regard to executive cognitive function as measured by the letter–digit coding test. There also was no effect of treatment on blood pressure, weight, waist circumference, body-mass index, or the Barthel Index or Instrumental Activities of Daily Living scores.

Participants were monitored closely for adverse effects from levothyroxine treatment. We found no increase in hyperthyroid symptoms after the initiation of treatment, and there was no significant excess of serious adverse events of special interest, including atrial fibrillation, heart failure, fracture, or new diagnosis of osteoporosis. We believe that the slight excess of patients who had serious adverse events in the placebo group is a chance finding; the events were spread among a range of body systems, and no particular pattern was observed. Observational studies also have not shown any association of treatment of subclinical hypothyroidism with an increased risk of adverse events.[26]

Many older persons with biochemical results that are consistent with subclinical hypothyroidism will have reversion to a euthyroid state if they are followed up without treatment. In total, approximately three out of five persons that we screened for entry into the trial on the basis of previously elevated thyrotropin levels had reversion to normal thyroid biochemical results and were therefore excluded from the trial. These data are consistent with several other observational and trial cohorts that showed a high proportion of participants with an elevated thyrotropin level having reversion to biochemical euthyroidism during follow-up.[4,27,28]

Our trial has certain strengths. The trial included a sufficient number of participants to provide good statistical power to show no benefits regarding symptoms. We used validated measures of thyroid-specific quality of life that have been shown to be sensitive to change,[14,17] as well as a range of secondary outcomes of clinical relevance. However, the trial also had certain limitations. First, we chose to set a thyrotropin target of 0.40 to 4.60 mIU per litre with levothyroxine treatment, which is an approach that reflects recent guidelines, particularly for older persons.[7] However, some authorities have recommended a lower thyrotropin target (e.g., 0.40 to 2.50 mIU per litre).[29] We cannot exclude the possibility that this more aggressive treatment approach might be beneficial. Second, since few participants had a baseline thyrotropin level of more than 10 mIU per litre, we cannot address whether there are benefits from treatment in this subgroup. Third, the symptom levels at trial entry were low, so we cannot exclude the possibility of benefit in persons with more marked symptoms. Fourth, we did not measure thyroid antibody levels. Antibody-positive patients are more likely than antibody-negative patients to have progressive hypothyroidism and therefore may be more likely to have a benefit from long-term levothyroxine treatment.[7] Finally, our trial was underpowered to detect any effect of levothyroxine on the incidence of cardiovascular events or mortality. Therefore, we cannot exclude the possibility that treatment with levothyroxine may provide cardiovascular protection or cause harm.

In conclusion, this trial indicated that treatment with levothyroxine in older persons with subclinical hypothyroidism provided no symptomatic benefits.

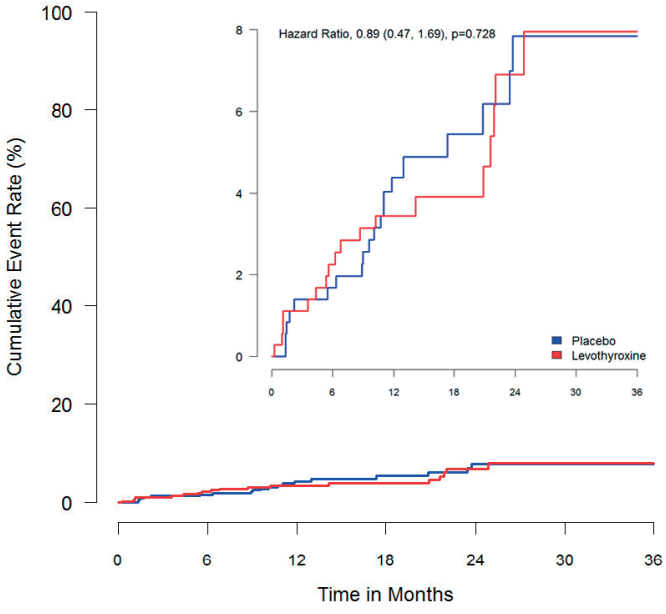
Online supplemental material

https://www.nejm.org/doi/full/10.1056/NEJMoa1603825

REFERENCES

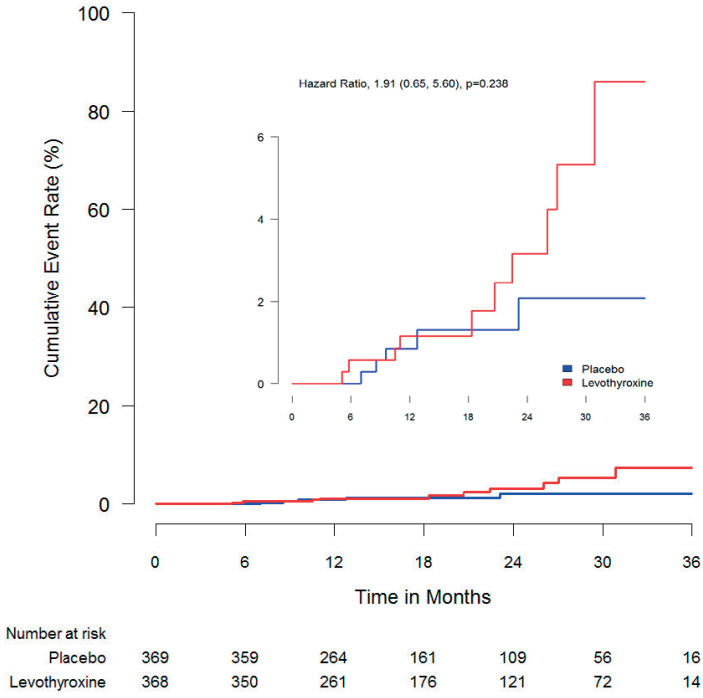
1. Rugge JB, Bougatsos C, Chou R. Screening and treatment of thyroid dysfunction: an evidence review for the U.S. Preventive Services Task Force. *Ann Intern Med.* 2015;162(1):35-45.
2. Canaris GJ, Manowitz NR, Mayor G, Ridgway EC. The Colorado thyroid disease prevalence study. *Arch Intern Med.* 2000;160(4):526-534.
3. 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. *J Clin Endocrinol Metab.* 2001;86(3):1110-1115.
4. 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. *J Clin Endocrinol Metab.* 2010;95(8):3623-3632.
5. Baumgartner C, den Elzen WP, Blum MR, et al. Variation in treatment strategies of Swiss general practitioners for subclinical hypothyroidism in older adults. *Swiss Med Wkly.* 2015;145:w14156.
6. Carle A, Pedersen IB, Knudsen N, Perrild H, Ovesen L, Laurberg P. Hypothyroid symptoms and the likelihood of overt thyroid failure: a population-based case-control study. *European journal of endocrinology / European Federation of Endocrine Societies.* 2014;171(5):593-602.
7. Javed Z, Sathyapalan T. Levothyroxine treatment of mild subclinical hypothyroidism: a review of potential risks and benefits. *Ther Adv Endocrinol Metab.* 2016;7(1):12-23.
8. Rodondi N, den Elzen WP, Bauer DC, et al. Subclinical hypothyroidism and the risk of coronary heart disease and mortality. *JAMA.* 2010;304(12):1365-1374.
9. Mikhail GS, Alshammari SM, Alenezi MY, Mansour M, Khalil NA. Increased atherogenic low-density lipoprotein cholesterol in untreated subclinical hypothyroidism. *Endocrine practice : official journal of the American College of Endocrinology and the American Association of Clinical Endocrinologists.* 2008;14(5):570-575.
10. Villar HC, Saconato H, Valente O, Atallah AN. Thyroid hormone replacement for subclinical hypothyroidism. *Cochrane Database Syst Rev.* 2007;2007(3):CD003419.
11. Stott DJ, Gussekloo J, Kearney PM, et al. Study protocol; Thyroid hormone Replacement for Untreated older adults with Subclinical hypothyroidism - a randomised placebo controlled Trial (TRUST). *BMC Endocr Disord.* 2017;17(1):6.
12. World Medical A. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. *JAMA.* 2013;310(20):2191-2194.
13. Medical Research Council. MRC guidelines for good clinical practice in clinical trials. In. London1998.
14. 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.
15. Wong CK, Lang BH, Lam CL. A systematic review of quality of thyroid-specific health-related quality-of-life instruments recommends ThyPRO for patients with benign thyroid diseases. *J Clin Epidemiol.* 2016;78:63-72.
16. EuroQol G. EuroQol--a new facility for the measurement of health-related quality of life. *Health Policy.* 1990;16(3):199-208.
17. Watt T, Bjorner JB, Groenvold M, et al. Development of a Short Version of the Thyroid-Related Patient-Reported Outcome ThyPRO. *Thyroid.* 2015;25(10):1069-1079.
18. Gunther CM, Burger A, Rickert M, Crispin A, Schulz CU. Grip strength in healthy caucasian adults: reference values. *J Hand Surg Am.* 2008;33(4):558-565.

19. 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. *J Neurol Neurosurg Psychiatry.* 2002;73(4):385-389.
20. Mahoney FI, Barthel DW. Functional Evaluation: The Barthel Index. *Maryland state medical journal.* 1965;14:61-65.
21. Katz S, Ford AB, Moskowitz RW, Jackson BA, Jaffe MW. Studies of Illness in the Aged. The Index of Adl: A Standardized Measure of Biological and Psychosocial Function. *JAMA.* 1963;185:914-919.
22. Winther KH, Cramon P, Watt T, et al. Disease-Specific as Well as Generic Quality of Life Is Widely Impacted in Autoimmune Hypothyroidism and Improves during the First Six Months of Levothyroxine Therapy. *PLoS One.* 2016;11(6):e0156925.
23. 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. *J Clin Endocrinol Metab.* 2007;92(5):1715-1723.
24. McClure NS, Sayah FA, Xie F, Luo N, Johnson JA. Instrument-Defined Estimates of the Minimally Important Difference for EQ-5D-5L Index Scores. *Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research.* 2017;20(4):644-650.
25. Reuters VS, Teixeira Pde F, Vigario PS, et al. Functional capacity and muscular abnormalities in subclinical hypothyroidism. *Am J Med Sci.* 2009;338(4):259-263.
26. Razvi S, Weaver JU, Butler TJ, Pearce SH. Levothyroxine treatment of subclinical hypothyroidism, fatal and nonfatal cardiovascular events, and mortality. *Arch Intern Med.* 2012;172(10):811-817.
27. Diez JJ, Iglesias P, Burman KD. Spontaneous normalization of thyrotropin concentrations in patients with subclinical hypothyroidism. *J Clin Endocrinol Metab.* 2005;90(7):4124-4127.
28. Parle JV, Franklyn JA, Cross KW, Jones SC, Sheppard MC. Prevalence and follow-up of abnormal thyrotrophin (TSH) concentrations in the elderly in the United Kingdom. *Clin Endocrinol (Oxf).* 1991;34(1):77-83.
29. Pearce SH, Brabant G, Duntas LH, et al. 2013 ETA Guideline: Management of Subclinical Hypothyroidism. *Eur Thyroid J.* 2013;2(4):215-228.



Number at risk							
Placebo	369	344	243	151	101	52	14
Levothyroxine	368	342	249	165	111	66	14

Supplemental figure 1. Time course of incident fatal plus nonfatal cardiovascular events in placebo and levothyroxine groups (modified intention to treat population).



Supplemental figure 2. Time course of all-cause mortality in placebo and levothyroxine groups (intention to treat population).

Supplemental table 1. Baseline characteristics of study participants who provided the co- primary outcome measures (ThyPRO Hypothyroid Symptom and Fatigue scores at 12 months).

	Placebo (n=320)	Levothyroxine (n=318)
Demographics		
Age (years)	74.6 (6.8)	73.8 (5.7)
Mean, SD and range	[65.1-93.4]	[65.2-93.0]
Female sex	170 (53.1%)	166 (52.2%)
White race	315 (98.4%)	313 (98.4%)
Standard housing	307 (95.9%)	311 (97.8%)
Previous medical conditions / clinical descriptors		
Ischemic heart disease	43 (13.5%)	38 (11.9%)
Atrial fibrillation	38 (11.9%)	35 (11.1%)
Hypertension	156 (49.2%)	163 (51.3%)
Diabetes mellitus	45 (14.1%)	55 (17.3%)
Osteoporosis	44 (13.8%)	36 (11.5%)
Current smokers	24 (7.5%)	25 (7.9%)
Number of concomitant medicines [median and lower quartile, upper quartile]	4 (2, 6)	4 (2, 6)
Mini-mental state examination [median (lower quartile, upper quartile)]	29 (28, 30)	29 (27, 29)
Weight < 50Kg	3 (0.9%)	4 (1.3%)
Laboratory results		
TSH (mIU/L)		
Mean (SD)	6.3 (1.7)	6.4 (2.1)
Median (lower quartile, upper quartile)	5.8 (5.1, 6.8)	5.7 (5.1, 6.8)
fT4 (pmol/L)	13.4 (1.8)	13.4 (2.0)
Outcome measures		
ThyPRO Hypothyroid Symptoms (0-100)	16.5 (17.8)	16.4 (17.8)
ThyPRO Tiredness (0-100)	25.4 (20.2)	24.8 (19.6)
EuroQol-5D	0.857 (0.163)	0.863 (0.161)
EuroQol visual analogue scale	77.0 (16.2)	79.1 (15.4)
Handgrip strength (Kg)	27.9 (11.3)	28.6 (10.3)
Letter Digit Coding Test	25.3 (8.1)	25.2 (7.5)
Systolic Blood Pressure (mmHg)	141.4 (18.4)	141.2 (18.4)
Diastolic Blood Pressure (mmHg)	75.4 (11.7)	74.3 (11.6)
Body Mass Index (kg/m^2)	27.7 (4.5)	28.1 (5.2)
Waist circumference (cm)	97.5 (12.9)	98.4 (13.2)
Barthel index [median and range]	20 (14, 20)	20 (13, 20)
Instrumental Activities of Daily Living [median and range]	14 (7, 14)	14 (7, 14)

Results for continuous variables are expressed as mean (SD), categorical variables as number (percent), unless otherwise stated.

Standard housing was defined as non-sheltered community accommodation. Sheltered housing is purpose-built grouped housing for older people, often with an on-site scheme manager or warden. Ischemic heart disease was defined as history of angina pectoris and / or previous myocardial infarction. The Letter Digit Coding Test score (test of executive cognitive function) is number of correct responses in 90 seconds. The body mass index is the weight in kilograms divided by the square of the height in meters. The Barthel index uses the 20-point scale; the Instrumental Activities of Daily Living scale has a maximum score of 14; higher scores are associated with better performance.

No statistically significant between-group differences were seen in baseline characteristics of study participants who provided the co-primary outcome measures.

Abbreviations: fT4, free thyroxine; SD, Standard Deviation; TSH, thyroid-stimulating hormone

Supplemental table 2. Study outcomes recorded at 6-8 weeks (modified intention to treat analysis).

	Baseline		6-8 weeks		Between group difference at 6-8 weeks (95% CI);
	Placebo	**Levothyroxine**	**Placebo**	**Levothyroxine**	**Levothyroxine-placebo**
TSH (mIU/L)					
Mean (SD)	6.38 (2.01)	6.41 (2.01)	5.59 (2.24)	3.35(1.78)	-2.29 (-2.54, -2.03)[a]
Median (lower quartile, upper quartile)	5.76(5.10,6.94)	5.72 (5.10,6.83)	5.10 (4.08,6.48)	3.04 (2.20,4.02)	
	n=369	n=368	n=355	n=350	
Primary outcomes					
ThyPRO Hypothyroid Symptoms (0-100)	16.9 (17.9)	17.5 (18.8)	14.2 (17.3)	15.2 (17.2)	0.7 (-1.3, 2.7)
	n=369	n=368	n=357	n=351	
ThyPRO Tiredness (0-100)	25.5 (20.3)	25.9 (20.6)	24.1 (19.9)	25.8 (20.2)	1.6 (-0.5, 3.7)
	n=369	n=368	n=357	n=351	
Secondary outcomes					
EuroQol 5D	0.847 (0.171)	0.846 (0.187)	0.857(0.174)	0.848 (0.200)	-0.010 (-0.031, 0.010)
	n=369	n=368	n=356	n=351	
EuroQol visual anologue scale	76.5 (16.3)	78.40 (15.3)	77.9 (15.8)	77.9 (15.6)	-1.1 (-2.9, 0.6)
	n=369	n=368	n=356	n=352	
Adverse symptom assessment					
ThyPRO Hyperthyroid Symptoms (0-100)	10.5 (11.2)	10.5 (11.2)	8.8 (10.2)	8.9 (9.7)	0.3 (-0.8, 1.4)
	n=369	n=368	n=357	n=351	

Results are expressed as mean (SD) unless otherwise stated.

Results at 6-8 weeks and between-group differences are adjusted for stratification variables (country, sex and starting dose of levothyroxine) and baseline levels of the same variable using linear regression.

[a] $p < 0.001$

Abbreviations: CI, confidence interval; TSH, thyroid-stimulating hormone; ThyPRO, Thyroid-Related quality of life Patient-Reported Outcome questionnaire.

Supplemental table 3. Serum free thyroxine at study baseline, 6-8 weeks and 12 months (modified intention to treat analysis).

	Baseline		6-8 weeks		Between group difference at 6-8 weeks (95% CI)	12 months		Between group difference at 12months (95% CI)
	Placebo	Levothyroxine	Placebo	Levothyroxine	Levothyroxine-placebo difference	Placebo	Levothyroxine	Levothyroxine-placebo difference
Free thyroxine (pmol/L)	13.3 (1.9)	13.4 (2.1)	12.6 (2.4)	15.1 (2.9)	2.3 (1.8, 2.8)[a]	12.4 (1.8)	14.7 (2.8)	2.3 (1.8, 2.7)[a]
	n=369	n=368	n=110	n=108		n=89	n=98	

Free thyroxine was routinely measured at study baseline, but in trial it was not routinely measured; the data presented at 6-8 weeks and 12 months are from a convenience sample.

[a] $p < 0.001$

Supplemental table 4. Study outcomes at 12 months and extended follow-up (per protocol analysis).

	Baseline		12 months		Between group difference at 12 months (95% CI)	Extended follow-up visit		Between group difference at extended follow- up visit (95% CI)
	Placebo	Levothyroxine	Placebo	Levothyroxine	Levothyroxine-placebo difference	Placebo	Levothyroxine	Levothyroxine-placebo difference
TSH (mIU/L)								
Mean (SD)	6.38 (2.01)	6.41 (2.01)	5.54 (2.49)	3.37 (1.86)	-2.22 (-2.55,-1.90) **	5.39 (2.53)	3.14 (1.69)	-2.38 (-2.80, -1.95) **
Median (lower quartile, upper quartile)	5.76 (5.10,6.94)	5.72 (5.10,6.83)	4.91 (3.96,6.55)	3.01 (2.38,3.93)		5.00 (3.79,6.24)	2.90 (2.17,3.72)	
	n=369	n=368	n=291	n=277		n=154	n=161	
Primary outcomes								
ThyPRO Hypothyroid Symptoms (0-100)	16.9 (17.9)	17.5 (18.8)	16.2 (17.1)	15.7 (16.6)	0.0 (-2.1, 2.2)	14.9 (16.9)	18.3 (18.9)	2.6 (-0.5, 5.7)
	n=369	n=368	n=295	n=278		n=158	n=162	
ThyPRO Tiredness (0-100)	25.5 (20.3)	25.9 (20.6)	28.1 (18.9)	27.9 (19.7)	0.4 (-2.2,3.0)	30.9 (22.3)	29.7 (20.2)	-2.1 (-5.7,1.6)
	n=369	n=368	n=295	n=278		n=158	n=162	
Secondary outcomes								
EuroQol 5D	0.847 (0.171)	0.846 (0.187)	0.860 (0.178)	0.842 (0.198)	-0.025 (-0.050,0.000)*	0.842 (0.188)	0.867 (0.189)	-0.027 (-0.008,0.063)
	n=369	n=368	n=295	n=278		n=158	n=162	
EuroQol visual analogue scale	76.5 (16.3)	78.4 (15.3)	77.5 (13.5)	78.4 (14.75)	-0.4 (-2.3, 1.5)	78.4 (12.9)	76.9 (14.3)	-1.5 (-4.1, 1.2)
	n=369	n=368	n=294	n=278		n=158	n=162	

Supplemental table 4. Study outcomes at 12 months and extended follow-up (per protocol analysis). (*continued*)

	Baseline		12 months		Between group difference at 12 months (95% CI)	Extended follow-up visit		Between group difference at extended follow- up visit (95% CI)
Handgrip strength (kg)	27.5 (11.3)	28.0 (10.2)	27.2 (11.1)	27.8 (10.5)	0.0 (-0.9, 0.8)	24.9 (10.4)	24.8 (10.1)	-0.6 (-1.9, 0.6)
	n=358	n=358	n=274	n=264		n=149	n=158	
Systolic blood pressure (mmHg)	140.4 (18.9)	141.2 (18.7)	138.4 (17.4)	138.0 (18.7)	0.1 (-2.2, 2.5)	135.8 (17.9)	136.4 (17.9)	1.0 (-2.2, 4.2)
	n=368	n=368	n=294	n=278		n=155	n=160	
Diastolic blood pressure (mmHg)	74.8 (11.7)	74.1 (11.6)	73.6 (10.9)	72.8 (11.4)	0.3 (-1.2, 1.8)	72.1 (10.4)	71.8 (11.8)	1.3 (-0.7, 3.3)
	n=368	n=368	n=294	n=278		n=155	n=160	
Body mass index (kg/m^2)	27.7 (4.6)	28.1 (5.3)	27.8 (4.7)	28.0 (5.0)	0.0 (-0.2, 0.2)	27.3 (4.6)	28.1 (4.8)	0.1 (-0.2, 0.4)
	n=368	n=367	n=293	n=277		n=158	n=162	
Waist circumference (cm)	97.5 (12.8)	98.5 (13.6)	97.0 (13.2)	98.0 (12.6)	0.3 (-0.6,1.2)	95.7 (14.3)	97.9 (13.0)	0.2 (-1.1, 1.4)
	n=368	n=367	n=294	n=276		n=158	n=161	
Adverse symptom assessment								
ThyPRO Hyperthyroid Symptoms (0-100)	10.5 (11.2)	10.5 (11.2)	10.1 (11.2)	10.1 (10.6)	0.8 (-0.5,2.1)	9.8 (11.3)	10.7 (11.5)	0.5 (-1.5, 2.5)
	n=369	n=368	n=295	n=278		n=158	n=162	

* p=0.05,** p< 0.001. Results are expressed as mean (SD) unless otherwise stated.

Results at 12 months, at the extended follow-up visit and between-group differences are adjusted for stratification variables (country, sex and starting dose of levothyroxine) and baseline levels of the same variable using linear regression; data for the extended follow-up visit are additionally adjusted for time to visit. The body mass index is the weight in kilograms divided by the square of the height in metres.

Abbreviations: CI, confidence interval; IQR, interquartile range; TSH, thyroid-stimulating hormone; ThyPRO, Thyroid-Related quality of life Patient-Reported Outcome questionnaire.

Supplemental table 5. Sensitivity analyses for missing data for study outcomes at 12 months: treatment effect estimated using multiple imputation of missing values.

	Treatment Effect (Levothyroxine - Placebo)	95% confidence interval
Primary Outcomes		
ThyPRO Hypothyroid Symptoms (0-100)	-0.3	-2.2, 1.5
ThyPRO Tiredness (0-100)	0.8	-1.5, 3.0
Secondary Outcomes		
EuroQol 5D	-0.26[a]	-0.05, -0.00
EuroQol visual analogue scale	-1.1	-2.8, 0.6
Handgrip strength (kg)	-0.1	-0.8, 0.6
Systolic blood pressure (mmHg)	0.5	-1.5, 2.5
Diastolic blood pressure (mmHg)	0.2	-1.1, 1.4
Body mass index (kg/m^2)	0.1	-0.1, 0.2
Waist circumference (cm)	0.5	-0.3, 1.2
Adverse symptom assessment		
ThyPRO Hyperthyroid Symptoms (0-100)	0.3	-0.9, 1.4

Missing values are predicted from age, gender, baseline value and value at 6-8 week visit if available, in 10 imputations. The treatment effect is estimated in a linear mixed effects regression model predicting change from baseline to follow-up visit with the following covariates: randomised treatment, baseline value of outcome variable and stratification variables (site, gender, dose at randomisation).

[a] p = 0.03

Supplemental table 6. Executive cognitive function (letter digit coding test), activities of daily living, and comprehensive thyroid-related quality of life at final review in placebo and levothyroxine groups (modified intention to treat analysis).

	Baseline		Final follow-up visit		Between group difference at final follow-up visit (95% CI)
	Placebo	**Levothyroxine**	**Placebo**	**Levothyroxine**	**Levothyroxine-placebo difference**
Letter-digit coding test	25.2 (8.3)	28.0 (10.2)	27.1 (11.2)	27.5 (10.5)	-0.1 (-0.9, 0.7)
	n=366	n=358	n=298	n=302	
Barthel index [median and range]	20 (14, 20)	20 (13, 20)	20 (12, 20)	20 (9, 20)	-0.1 (-0.3, 0.1)
	n=369	n=367	n=325	n=321	
Instrumental Activities of Daily Living [median and range]	14 (7, 14)	14 (7, 14)	14 (5, 14)	14 (3, 14)	-0.1 (-0.3, 0.1)
	n=368	n=368	n=325	n=322	
ThyPRO-39 (0-100)	-	-	15.4 (11.3)	15.0 (12.0)	-0.5 (-2.2, 1.3)
			n=325	n=323	

Final follow-up visit was the last review visit; this included participants who stopped the study at 12 months, and participants with extended follow-up (beyond 12 months); final review in the placebo group was at a median of 17.7 months (lower quartile 12.0, upper quartile 24.6) and in the levothyroxine group at a median of 17.9 months (12.0, 17.0).

The Letter Digit Coding Test score (test of executive cognitive function; number of correct responses in 90 seconds), Barthel Index (20-point scale) and Instrumental Activities of Daily Living (14-point scale) were recorded at baseline and at the final visit. ThyPRO-39 was recorded only at the final visit.

There were no significant between-group differences in these outcome measures at final follow-up visit.

Abbreviations: ThyPRO-39, Thyroid-Related quality of life Patient-Reported Outcome 39-point questionnaire (comprehensive assessment of thyroid-related quality of life).

Supplemental table 7. Primary outcomes for subgroups of sex at 12-month follow-up (modified intention to treat analysis).

	Baseline		12 months		Between group difference at 12 months (95% CI)
	Placebo	Levothyroxine	Placebo	Levothyroxine	Levothyroxine-placebo difference
Male					
TSH (mIU/L)					
Mean (SD)	6.34 (2.02)	6.37 (1.87)	5.58 (2.55)	3.76 (1.97)	-1.83 (-2.41,-1.25)[a]
Median (lower quartile, upper quartile)	5.72 (5.12,6.75)	5.70 (5.10,6.91)	4.95 (3.96,6.63)	3.38 (2.60,4.21)	
	n=155	n=157	n=155	n=157	
ThyPRO Hypothyroid Symptoms (0-100)	14.3 (15.5)	12.6 (14.6)	13.5 (15.8)	14.1 (15.8)	1.6 (-1.4, 4.6)
	n=159	n=157	n=159	n=157	
ThyPRO Tiredness (0-100)	21.9 (19.2)	20.9 (16.5)	25.8 (18.8)	26.8 (19.0)	1.6 (-2.0, 5.2)
	n=159	n=157	n=159	n=157	
Female					
TSH (mIU/L)					
Mean (SD)	6.35 (1.81)	6.41 (2.16)	5.67 (4.06)	3.55 (2.22)	-2.14 (-2.69, -1.58)[a]
Median (lower quartile, upper quartile)	5.80 (5.10,6.99)	5.74 (5.11,6.82)	4.82 (3.90,6.42)	3.06 (2.23,4.25)	
	n=173	n=174	n=173	n=174	
ThyPRO Hypothyroid Symptoms (0-100)	18.7 (19.5)	20.9 (20.5)	20.3 (18.6)	19.6 (18.5)	-2.0 (-4.8, 0.8)
	n=178	n=175	n=178	n=175	
ThyPRO Tiredness (0-100)	29.2 (20.7)	29.0 (22.3)	31.6 (20.2)	31.0 (21.6)	-0.4 (-3.8, 3.0)
	n=178	n=175	n=178	n=175	

Results at 12 months are adjusted for stratification variables (country, sex and starting dose of levothyroxine) and baseline levels of the same variable using linear regression.

[a] $p < 0.001$

Supplemental table 8. Primary outcomes for subgroups of baseline TSH at 12-month follow-up (modified intention to treat analysis).

	Baseline		12 months		Between group difference at 12 months (95% CI)
	Placebo	**Levothyroxine**	**Placebo**	**Levothyroxine**	**Levothyroxine-placebo difference**
< 7 mIU/L					
TSH (mIU/L)					
Mean (SD)	5.52 (0.64)	5.55 (0.66)	5.0 (3.3)	3.4 (1.7)	-1.61 (-2.07, -1.16)[a]
Median (lower quartile, upper quartile)	5.40 (4.98,5.99)	5.40 (5.00,6.02)	4.58 (3.69,5.67)	3.12 (2.40,3.99)	
	n=250	n=256	n=250	n=256	
ThyPRO Hypothyroid Symptoms (0-100)	17.0 (17.9)	17.1 (17.7)	18.1 (18.0)	17.4 (17.2)	-0.8 (-3.1, 1.6)
	n=255	n=257	n=255	n=257	
ThyPRO Tiredness (0-100)	25.9 (20.4)	25.2 (20.1)	29.9 (19.8)	30.3 (20.9)	0.9 (-1.9, 3.7)
	n=255	n=257	n=255	n=257	
7-9.99 mIU/L					
TSH (mIU/L)					
Mean (SD)	8.08 (0.84)	8.18 (0.89)	6.95 (2.18)	3.99 (2.26)	-3.01 (-3.95, -2.07)[a]
Median (lower quartile, upper quartile)	8.02 (7.38,8.62)	7.9 (7.46,8.70)	6.76 (5.42,8.00)	3.56 (2.50,4.53)	
	n=62	n=57	n=62	n=57	
ThyPRO Hypothyroid Symptoms (0-100)	14.4 (16.2)	18.1 (21.5)	12.9 (13.7)	16.4 (18.8)	1.2 (-3.6, 6.0)
	n=65	n=57	n=65	n=57	
ThyPRO Tiredness (0-100)	25.5 (21.4)	25.1 (19.3)	24.8 (19.3)	24.4 (16.3)	-0.8 (-6.5, 5.0)
	n=65	n=57	n=65	n=57	
≥10 mIU/L					
TSH (mIU/L)					
Mean (SD)	12.48 (2.35)	12.78 (2.43)	10.06 (5.15)	6.10 (3.94)	-4.19 (-5.95, -2.43)[a]
Median (lower quartile, upper quartile)	11.60 (10.74,13.31)	11.95 (10.95,14.10)	10.90 (7.73,13.21)	4.62 (3.72,7.42)	

Supplemental table 8. Primary outcomes for subgroups of baseline TSH at 12-month follow-up (modified intention to treat analysis). (*continued*)

	Baseline		12 months		Between group difference at 12 months (95% CI)
	n=16	n=18	n=16	n=18	
ThyPRO Hypothyroid Symptoms (0-100)	20.2 (22.5)	12.2 (18.2)	19.1 (23.4)	14.2 (16.4)	0.6 (-8.3, 9.6)
	n=17	n=18	n=17	n=18	
ThyPRO Tiredness (0-100)	25.0 (15.8)	25.7 (24.8)	28.8 (20.3)	25.4 (24.3)	-2.8 (-13.6, 8.0)
	n=17	n=18	n=17	n=18	

Results at 12 months are adjusted for stratification variables (country, sex and starting dose of levothyroxine) and baseline levels of the same variable using linear regression.

[a] $p < 0.001$