

Table 1. Baseline Characteristics of Participants, by Hypothyroid Symptoms Score and Treatment Assignment

Characteristic	Hypothyroid Symptoms Score >30 (n = 132)		Hypothyroid Symptoms Score ≤30 (n = 506)	
	L-thyroxine (n = 66)	Placebo (n = 66)	L-thyroxine (n = 252)	Placebo (n = 254)
<b>Demographic characteristics</b>				
Mean age (SD) [range], y	73.9 (6.0) [66.4 to 91.2]	76.0 (6.2) [65.5 to 89.5]	73.8 (5.6) [65.2 to 93.0]	74.3 (6.9) [65.1 to 93.4]
Female, n (%)	43 (65.2)	42 (63.6)	123 (48.8)	128 (50.4)
White, n (%)	64 (97.0)	65 (98.5)	249 (98.8)	250 (98.4)
<b>Previous medical conditions/clinical descriptors</b>				
Ischemic heart disease, n (%)*	8 (12.1)	16 (24.2)	30 (11.9)	27 (10.7)
Hypertension, n (%)	36 (54.5)	35 (53.8)	127 (50.4)	121 (48.0)
Diabetes mellitus, n (%)	18 (27.3)	11 (16.7)	37 (14.7)	34 (13.4)
Current smokers, n (%)	7 (10.6)	3 (4.5)	18 (7.1)	21 (8.3)
Median number of medicines (IQR)	5 (3 to 8)	5 (3 to 8)	3 (2 to 5)	3 (2 to 6)
<b>Laboratory results</b>				
Thyroid-stimulating hormone level, mIU/L				
Mean (SD)	6.1 (1.4)	6.3 (1.9)	6.5 (2.2)	6.3 (1.7)
Median (IQR)	5.7 (5.1 to 6.3)	5.7 (5.3 to 6.6)	5.7 (5.1 to 6.9)	5.8 (5.1 to 7.0)
Mean free thyroxine level (SD)				
pmol/L	13.7 (2.1)	13.4 (1.8)	13.3 (2.0)	13.4 (1.9)
ng/dL	1.06 (0.16)	1.04 (0.14)	1.03 (0.16)	1.04 (0.15)
<b>ThyPRO Hypothyroid Symptoms and Tiredness scores†</b>				
Mean Hypothyroid Symptoms score (SD) [range]	45.1 (14.5) [31.2 to 87.5]	45.2 (14.7) [31.2 to 93.8]	8.9 (8.6) [0.0 to 25.0]	9.0 (8.5) [0.0 to 25.0]
Mean Tiredness score (SD) [range]	38.8 (22.8) [0.0 to 92.9]	35.9 (22.9) [0.0 to 96.4]	21.2 (16.9) [0.0 to 87.5]	22.6 (18.5) [0.0 to 85.7]
<b>Secondary outcomes</b>				
Mean EQ-5D Health Utility score (SD) [range]‡	0.752 (0.184) [-0.016 to 1.000]	0.777 (0.178) [-0.016 to 1.000]	0.892 (0.141) [0.055 to 1.000]	0.877 (0.153) [-0.077 to 1.000]
Mean handgrip strength (SD) [range], kg§	24.6 (8.8) [7.0 to 49.0]	25.0 (12.1) [4.0 to 63.0]	29.7 (10.4) [8.0 to 60.0]	28.7 (11.0) [4.0 to 60.0]

EQ-5D = EuroQol Group 5-Dimension Self-Report Questionnaire; IQR = interquartile range; ThyPRO = Thyroid-Related Quality-of-Life Patient-Reported Outcome.

\* Defined as previous myocardial infarction or history of angina pectoris.

† The Hypothyroid Symptoms score (4 items) and the Tiredness score (7 items) from the ThyPRO questionnaire range from 0 to 100. Higher scores indicate greater symptom burden (i.e., more hypothyroid symptoms or tiredness, respectively). Nine points is estimated to be the minimal clinically important difference for each score.

‡ The EQ-5D Health Utility score is based on the EQ-5D Self-Report Questionnaire and ranges from 0.00 to 1.000. Higher (more positive) scores indicate better quality of life. The minimal clinically important difference is approximately 0.050.

§ Handgrip strength was assessed using the Jamar isometric dynamometer, with the recorded score as the best of 3 measures in the dominant hand. Handgrip strength was available for 622 participants at baseline and follow-up (not 638 as for the other outcomes in this table).

Table 2. Baseline Characteristics of Participants, by Tiredness Score and Treatment Assignment

Characteristic	Tiredness Score >40 (n = 133)		Tiredness Score ≤40 (n = 505)	
	L-thyroxine (n = 65)	Placebo (n = 68)	L-thyroxine (n = 253)	Placebo (n = 252)
<b>Demographic characteristics</b>				
Mean age (SD) [range], y	74.2 (5.8) [65.2 to 93.0]	75.4 (6.5) [65.1 to 91.2]	73.7 (5.6) [65.3 to 91.2]	74.4 (6.8) [65.1 to 93.4]
Female, n (%)	43 (66.2)	44 (64.7)	123 (48.6)	126 (50.0)
White, n (%)	64 (98.5)	67 (98.5)	249 (98.4)	248 (98.4)
<b>Previous medical conditions/clinical descriptors</b>				
Ischemic heart disease, n (%)*	9 (13.8)	13 (19.1)	29 (11.5)	30 (12.0)
Hypertension, n (%)	42 (64.6)	37 (54.4)	121 (47.8)	119 (47.8)
Diabetes mellitus, n (%)	19 (29.2)	15 (22.4)	36 (14.2)	30 (11.9)
Current smokers, n (%)	6 (9.2)	5 (7.4)	19 (7.5)	19 (7.5)
Median number of medicines (IQR)	5 (3 to 8)	6 (3 to 8)	3 (2 to 5)	4 (2 to 5)
<b>Laboratory results</b>				
Thyroid-stimulating hormone level, mIU/L				
Mean (SD)	6.5 (2.2)	6.1 (1.6)	6.4 (2.0)	6.3 (1.8)
Median (IQR)	5.8 (5.1 to 6.9)	5.7 (5.1 to 6.7)	5.7 (5.1 to 6.8)	5.8 (5.2 to 6.9)
Mean free thyroxine level (SD)				
pmol/L	13.3 (2.2)	13.2 (1.9)	13.4 (2.0)	13.5 (1.8)
ng/dL	1.03 (0.17)	1.03 (0.15)	1.04 (0.16)	1.05 (0.14)
<b>ThyPRO Hypothyroid Symptoms and Tiredness scores†</b>				
Mean Tiredness score (SD) [range]	56.0 (12.8) [42.9 to 92.9]	57.4 (13.5) [42.9 to 96.4]	16.8 (11.3) [0.0 to 39.3]	16.7 (10.8) [0.0 to 39.3]
Mean Hypothyroid Symptoms score (SD) [range]	31.2 (22.1) [0.0 to 87.5]	24.4 (20.6) [0.0 to 93.8]	12.5 (14.3) [0.0 to 87.5]	14.3 (16.3) [0.0 to 81.2]
<b>Secondary outcomes</b>				
Mean EQ-5D Health Utility score (SD) [range]‡	0.742 (0.208) [-0.016 to 1.000]	0.750 (0.184) [-0.016 to 1.000]	0.894 (0.130) [0.055 to 1.000]	0.885 (0.145) [-0.077 to 1.000]
Mean handgrip strength (SD) [range], kg§	24.5 (9.0) [7.0 to 46.0]	24.1 (11.5) [4.0 to 62.0]	29.7 (10.3) [8.0 to 60.0]	29.0 (11.1) [4.0 to 63.0]

EQ-5D = EuroQol Group 5-Dimension Self-Report Questionnaire; IQR = interquartile range; ThyPRO = Thyroid-Related Quality-of-Life Patient-Reported Outcome.

\* Defined as previous myocardial infarction or history of angina pectoris.

† The Hypothyroid Symptoms score (4 items) and the Tiredness score (7 items) from the ThyPRO questionnaire range from 0 to 100. Higher scores indicate greater symptom burden (i.e., more hypothyroid symptoms or tiredness, respectively). Nine points is estimated to be the minimal clinically important difference for each score.

‡ The EQ-5D Health Utility score is based on the EQ-5D Self-Report Questionnaire and ranges from 0.00 to 1.000. Higher (more positive) scores indicate better quality of life. The minimal clinically important difference is approximately 0.050.

§ Handgrip strength was assessed using the Jamar isometric dynamometer, with the recorded score as the best of 3 measures in the dominant hand. Handgrip strength was available for 622 participants at baseline and follow-up (not 638 as for the other outcomes in this table).

Table 3. Effect of L-thyroxine Versus Placebo After 12 Months, With a Focus on High-Symptom Burden Groups

Measure	Mean Value (SD)		Treatment Effect (L-thyroxine vs. Placebo)	
	Baseline	12-mo Follow-up	Mean Within-Group Change (95% CI)	Adjusted Between-Group Difference (95% CI)* †
<b>ThyPRO Hypothyroid Symptoms score‡</b>				
Baseline score >30				
L-thyroxine (n = 66)	45.1 (14.5)	32.8 (20.9)	-12.3 (-16.6 to -8.0)	-2.0 (-5.5 to 1.5)
Placebo (n = 66)	45.2 (14.7)	34.8 (21.0)	-10.4 (-15.3 to -5.4)	
Baseline score ≤30				
L-thyroxine (n = 252)	8.9 (8.6)	12.4 (12.8)	3.5 (2.1 to 4.9)	0.6 (-1.6 to 2.7)
Placebo (n = 254)	9.0 (8.5)	12.0 (12.8)	2.9 (1.4 to 4.5)	
<b>ThyPRO Tiredness score‡</b>				
Baseline score >40				
L-thyroxine (n = 65)	56.0 (12.8)	47.1 (25.2)	-8.9 (-14.5 to -3.3)	0.0 (-4.1 to 4.0)
Placebo (n = 68)	57.4 (13.5)	46.5 (21.5)	-10.9 (-16.0 to -5.8)	
Baseline score ≤40				
L-thyroxine (n = 253)	16.8 (11.3)	23.9 (15.6)	7.1 (5.2 to 9.1)	0.5 (-2.0 to 3.0)
Placebo (n = 252)	16.7 (10.8)	23.7 (15.8)	7.0 (5.2 to 8.8)	
<b>EQ-5D Health Utility score§</b>				
Baseline score <0.75				
L-thyroxine (n = 76)	0.659 (0.161)	0.651 (0.262)	-0.008 (-0.063 to 0.047)	-0.093 (-0.129 to -0.057)
Placebo (n = 76)	0.648 (0.166)	0.716 (0.217)	0.068 (0.022 to 0.114)	
Baseline score ≥0.75				
L-thyroxine (n = 242)	0.927 (0.093)	0.890 (0.154)	-0.037 (-0.055 to -0.019)	-0.002 (-0.026 to 0.023)
Placebo (n = 244)	0.922 (0.093)	0.895 (0.160)	-0.027 (-0.045 to -0.008)	
<b>Handgrip strength  </b>				
Baseline strength <20 kg				
L-thyroxine (n = 57)	15.5 (3.1)	16.5 (5.9)	0.9 (-0.7 to 2.5)	0.7 (-0.8 to 2.2)
Placebo (n = 68)	14.1 (4.1)	15.3 (5.7)	1.1 (-0.1 to 2.4)	
Baseline strength ≥20 kg				
L-thyroxine (n = 255)	31.6 (9.0)	29.8 (9.7)	-1.8 (-2.5 to -1.2)	-0.3 (-1.1 to 0.6)
Placebo (n = 242)	31.8 (9.5)	30.2 (10.2)	-1.7 (-2.4 to -1.0)	

EQ-5D = EuroQol Group 5-Dimension Self-Report Questionnaire; ThyPRO = Thyroid-Related Quality-of-Life Patient-Reported Outcome.

\* Negative value indicates benefit of L-thyroxine therapy for ThyPRO Hypothyroid Symptoms and Tiredness scores, whereas positive values indicate benefit of L-thyroxine therapy for EQ-5D score and handgrip strength.

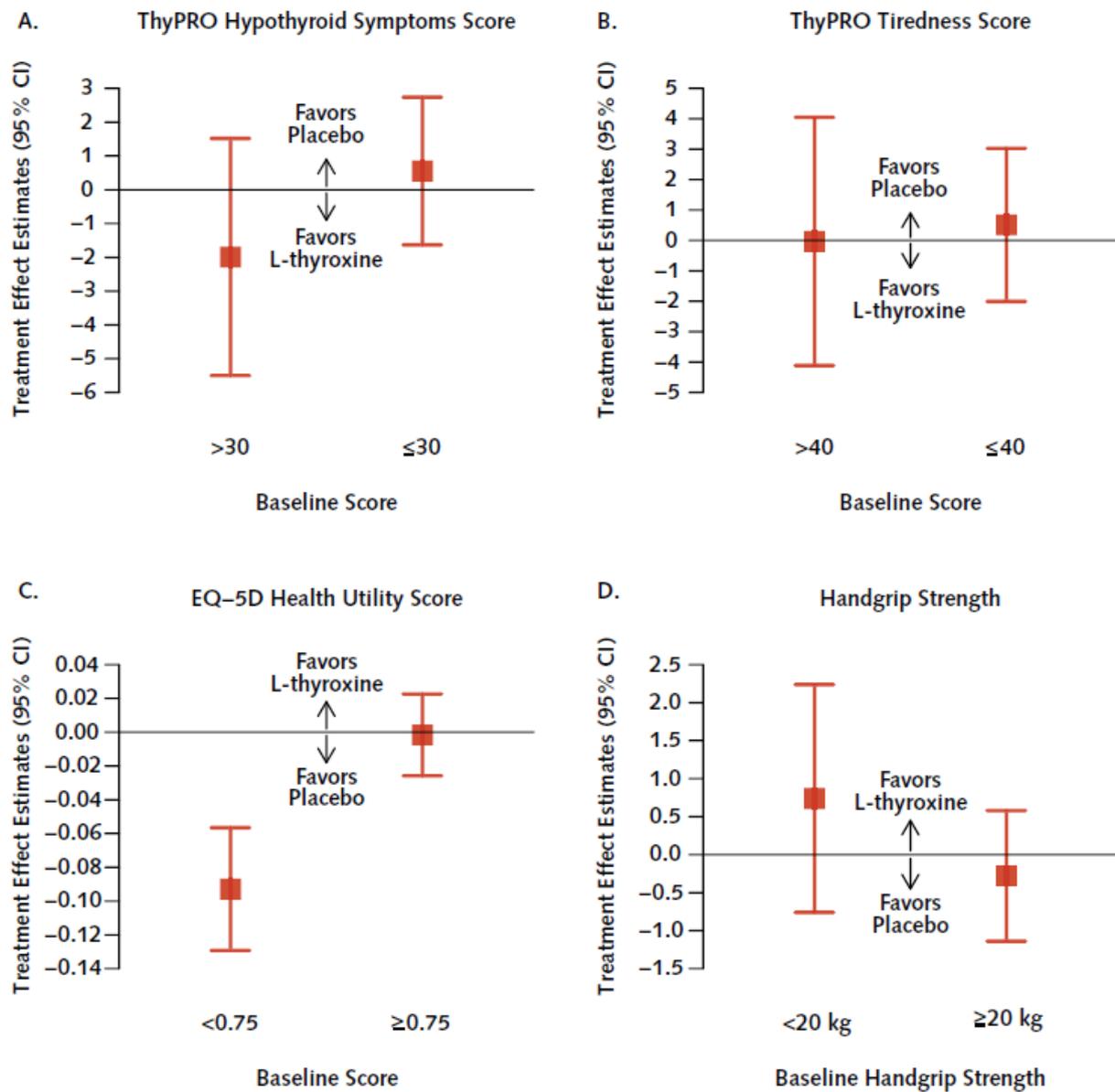
† Adjusted for sex, starting L-thyroxine dose at randomization, study center, and baseline symptom score. As an example, results read as follows: Among participants in the high-symptom burden group, Hypothyroid Symptoms score decreased by 2.0 points more on average in the L-thyroxine group than in the placebo group.

‡ The Hypothyroid Symptoms score (4 items) and the Tiredness score (7 items) from the ThyPRO questionnaire range from 0 to 100. Higher scores indicate greater symptom burden (i.e., more hypothyroid symptoms or tiredness, respectively). Nine points is estimated to be the minimal clinically important difference for each score.

§ The EQ-5D Health Utility score is based on the EQ-5D Self-Report Questionnaire and ranges from 0.00 to 1.000. Higher (more positive) scores indicate better quality of life. The minimal clinically important difference is approximately 0.050.

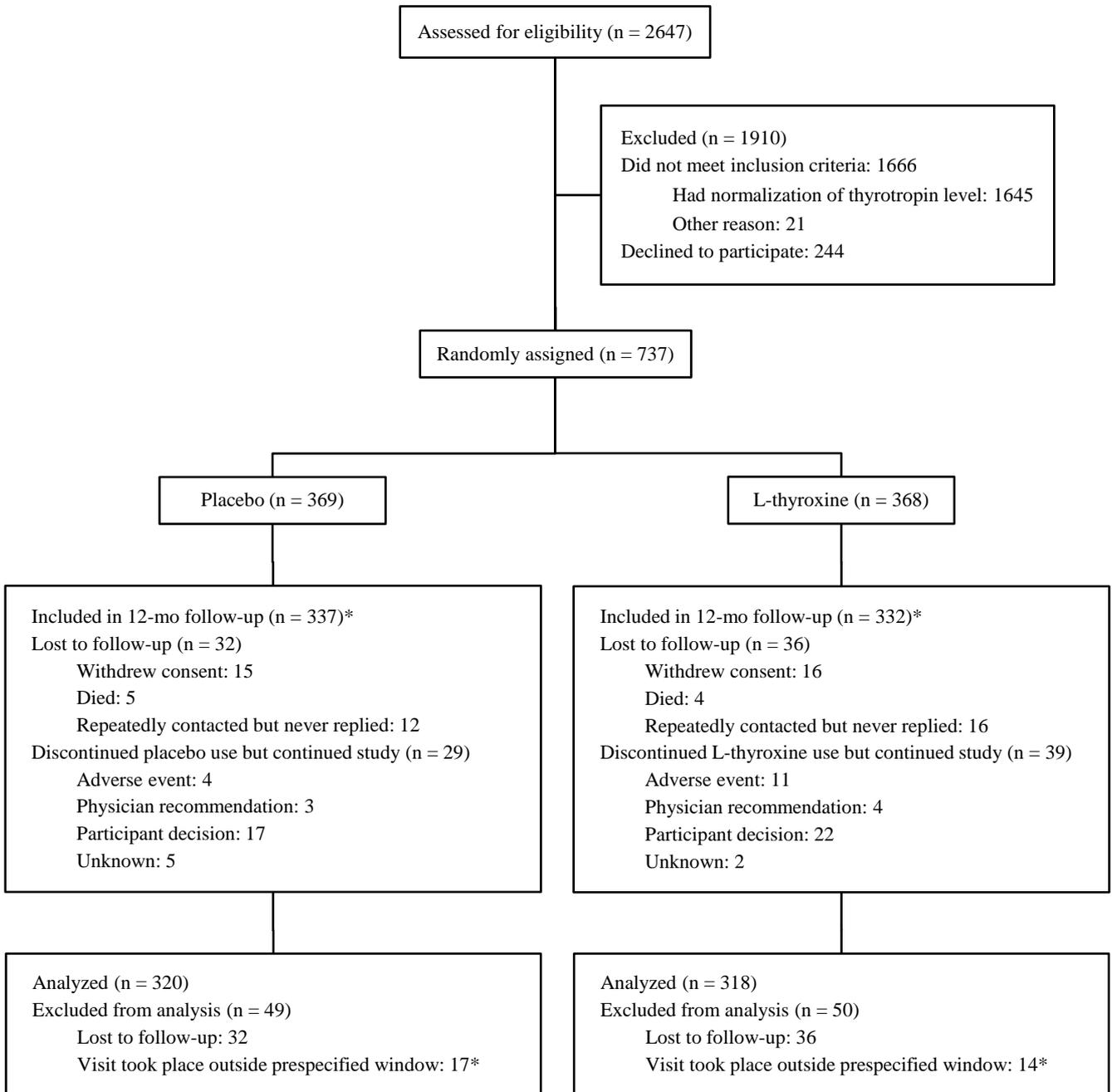
|| Handgrip strength was assessed using the Jamar isometric dynamometer, with the recorded score as the best of 3 measures in the dominant hand. Handgrip strength was available for 622 participants at baseline and follow-up (not 638 as for the other outcomes in this table).

Figure. Between-group differences in outcome measures at 1 y, stratified by baseline symptom burden.



The figure shows treatment effect estimates (L-thyroxine vs. placebo) with 95% CIs for study outcomes (ThyPRO Hypothyroid Symptoms score [A], ThyPRO Tiredness score [B], EQ-5D Health Utility score [C], and handgrip strength D)), stratified by baseline levels of the outcome variable. In each panel, the effect estimate on the left corresponds to the subgroup with higher symptom burden at baseline (i.e., Hypothyroid Symptoms score >30, Tiredness score >40, EQ-5D <0.75, and handgrip strength <20 kg). Treatment effect estimates were derived from repeated measures models of data collected at baseline, 6 to 8 wk, and 1 y. EQ-5D = EuroQol Group 5-Dimension; ThyPRO = Thyroid-Related Quality of-Life Patient-Reported Outcome.

Appendix Figure. Study flow diagram showing recruitment, randomization, and patient flow of the main trial on L-thyroxine therapy in older adults with subclinical hypothyroidism.



\* The prespecified range for the 12-mo follow-up visit was  $\pm 31$  d.

Appendix Table 1. Baseline and Follow-up Levels of Thyroid-Stimulating Hormone in High-Symptom Burden Groups

Variable	Baseline ThyPRO Hypothyroid Symptoms Score >30*		Baseline ThyPRO Tiredness Score >40*		Baseline EQ-5D Health Utility Score <0.75†		Baseline Handgrip Strength <20 kg‡	
	L-thyroxine	Placebo	L-thyroxine	Placebo	L-thyroxine	Placebo	L-thyroxine	Placebo
<b>Baseline</b>								
Thyroid-stimulating hormone, <i>mIU/L</i>								
Participants, <i>n</i>	66	66	65	68	76	76	57	68
Mean level (SD)	6.1 (1.4)	6.3 (1.9)	6.5 (2.2)	6.1 (1.6)	6.6 (2.3)	6.4 (1.8)	6.4 (2.1)	6.3 (1.7)
Median level (IQR)	5.7 (5.1 to 6.3)	5.7 (5.3 to 6.6)	5.8 (5.1 to 6.9)	5.7 (5.1 to 6.7)	5.7 (5.1 to 7.1)	5.9 (5.2 to 7.0)	6.0 (5.1 to 6.8)	5.7 (5.1 to 6.9)
<b>12-mo follow-up</b>								
Thyroid-stimulating hormone, <i>mIU/L</i>								
Participants, <i>n</i>	66	62	65	67	76	74	57	68
Mean level (SD)	3.6 (1.9)	5.4 (2.7)	3.6 (1.9)	5.3 (2.9)	3.6 (1.8)	5.3 (2.2)	3.3 (1.6)	5.0 (1.9)
Median level (IQR)	3.2 (2.5 to 4.4)	4.6 (3.8 to 5.9)	3.2 (2.6 to 4.2)	4.7 (3.6 to 6.2)	3.2 (2.5 to 4.2)	4.6 (3.8 to 6.1)	3.2 (2.3 to 4.0)	4.6 (3.8 to 6.0)
Participants with thyroid-stimulating hormone level <4.6 <i>mIU/L</i> , <i>n</i> (%)	52 (78.8)	30 (48.4)	52 (80.0)	31 (46.3)	62 (81.6)	37 (50.0)	49 (86.0)	24 (50.0)

EQ-5D = EuroQol Group 5-Dimension Self-Report Questionnaire; IQR = interquartile range; ThyPRO = Thyroid-Related Quality-of-Life Patient-Reported Outcome.

\* The Hypothyroid Symptoms score (4 items) and the Tiredness score (7 items) from the ThyPRO questionnaire ranges from 0 to 100. A higher score indicates greater symptom burden (i.e., more hypothyroid symptoms or tiredness). Nine points is estimated to be the minimal clinically important difference for each score.

† The EQ-5D Health Utility score is based on the EQ-5D Self-Report Questionnaire and ranges from 0.00 to 1.000. Higher (more positive) scores indicate better quality of life. The minimal clinically important difference is approximately 0.050.

‡ Handgrip strength was assessed by using the Jamar isometric dynamometer, with the recorded score as the best of 3 measures in the dominant hand.

Appendix Table 2. Sensitivity Analyses in High-Symptom Burden Groups

Measure	Treatment Effect After 12 Mo (L-thyroxine vs. Placebo)	
	Mean Within-Group Change in Scores (95% CI)	Adjusted Between-Group Difference (95% CI)*†
<b>Subgroup of participants who had both a Hypothyroid Symptoms score &gt;30 and a Tiredness score &gt;40 (n = 56)</b>		
ThyPRO Hypothyroid Symptoms score‡		
L-thyroxine	-15.4 (-22.1 to -8.8)	-6.8 (-17.4 to 3.9)
Placebo	-4.7 (-13.2 to 3.8)	
ThyPRO Tiredness score‡		
L-thyroxine	-9.7 (-16.8 to -2.5)	-1.6 (-13.3 to 10.0)
Placebo	-4.1 (-12.3 to 4.2)	
EQ-5D Health Utility score§		
L-thyroxine	-0.040 (-0.150 to 0.071)	-0.053 (-0.195 to 0.088)
Placebo	0.028 (-0.037 to 0.093)	
Handgrip strength		
L-thyroxine	-1.6 (-3.4 to 0.1)	-0.5 (-4.0 to 3.1)
Placebo	-3.8 (-8.1 to 0.5)	
<b>Adjusted additionally for ischemic heart disease, hypertension, diabetes mellitus, and smoking</b>		
ThyPRO Hypothyroid Symptoms score‡		
Baseline score >30 (n = 132)		
L-thyroxine	-12.3 (-16.6 to -8.0)	-2.0 (-5.5 to 1.5)
Placebo	-10.4 (-15.3 to -5.4)	
ThyPRO Tiredness score‡		
Baseline score >40 (n = 133)		
L-thyroxine	-8.9 (-14.5 to -3.3)	-0.5 (-4.6 to 3.6)
Placebo	-10.9 (-16.0 to -5.8)	
EQ-5D Health Utility score§		
Baseline score <0.75 (n = 152)		
L-thyroxine	-0.008 (-0.063 to 0.047)	-0.093 (-0.129 to -0.057)
Placebo	0.068 (0.022 to 0.114)	
Handgrip strength		
Baseline score <20 kg (n = 125)		
L-thyroxine	0.9 (-0.7 to 2.5)	0.8 (-0.7 to 2.3)
Placebo	1.1 (-0.1 to 2.4)	
<b>Analyses limited to those with baseline thyroid-stimulating hormone levels between 7.0 and 9.9 mIU/L</b>		
ThyPRO Hypothyroid Symptoms score‡		
Baseline score >30 (n = 20)		
L-thyroxine	-15.9 (-31.5 to -0.3)	0.9 (-10.1 to 11.8)
Placebo	-22.9 (-39.0 to -6.8)	
ThyPRO Tiredness score‡		
Baseline score >40 (n = 21)		
L-thyroxine	-15.4 (-36.5 to 5.7)	-2.1 (-13.3 to 9.1)
Placebo	-10.0 (-24.0 to 4.0)	
EQ-5D Health Utility score§		
Baseline score <0.75 (n = 29)		
L-thyroxine	-0.048 (-0.191 to 0.096)	-0.131 (-0.238 to -0.025)
Placebo	0.075 (-0.050 to 0.201)	
Handgrip strength		
Baseline score <20 kg (n = 24)		
L-thyroxine	6.0 (0.2 to 11.8)	4.4 (0.0 to 8.7)
Placebo	3.0 (-1.3 to 7.3)	

**Including data from 31 participants previously excluded because the 12-mo visit took place outside the prespecified window (12 mo ±31 d)**

ThyPRO Hypothyroid Symptoms score <sup>‡</sup>		
Baseline score >30 ( <i>n</i> = 144)		
L-thyroxine	-12.4 (-16.5 to -8.3)	-2.3 (-5.7 to 1.1)
Placebo	-10.6 (-15.3 to -5.9)	
ThyPRO Tiredness score <sup>‡</sup>		
Baseline score >40 ( <i>n</i> = 142)		
L-thyroxine	-8.8 (-14.4 to -3.1)	0.4 (-3.6 to 4.5)
Placebo	-11.2 (-16.3 to -6.1)	
EQ-5D Health Utility score <sup>§</sup>		
Baseline score <0.75 ( <i>n</i> = 166)		
L-thyroxine	-0.024 (-0.080 to 0.033)	-0.110 (-0.146 to -0.074)
Placebo	0.063 (0.018 to 0.108)	
Handgrip strength <sup>  </sup>		
Baseline score <20 kg ( <i>n</i> = 137)		
L-thyroxine	0.9 (-0.7 to 2.4)	0.7 (-0.7 to 2.1)
Placebo	1.0 (-0.3 to 2.2)	

EQ-5D = EuroQol Group 5-Dimension Self-Report Questionnaire; ThyPRO = Thyroid-Related Quality-of-Life Patient-Reported Outcome.

\* Negative values indicate benefit of L-thyroxine therapy for ThyPRO Hypothyroid Symptoms and Tiredness scores, whereas positive values indicate benefit of L-thyroxine therapy for EQ-5D score and handgrip strength.

† Adjusted for sex, starting L-thyroxine dose at randomization, study center, and baseline symptom score.

‡ The Hypothyroid Symptoms score (4 items) and the Tiredness score (7 items) from the ThyPRO questionnaire range from 0 to 100. Higher scores indicate greater symptom burden (i.e., more hypothyroid symptoms or tiredness, respectively). Nine points is estimated to be the minimal clinically important difference for each score.

§ The EQ-5D Health Utility score is based on the EQ-5D Self-Report Questionnaire and ranges from 0.00 to 1.000. Higher (more positive) scores indicate better quality of life. The minimal clinically important difference is approximately 0.050.

|| Handgrip strength was assessed using the Jamar isometric dynamometer, with the recorded score as the best of 3 measures in the dominant hand.