



Universiteit
Leiden
The Netherlands

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

Du Puy, R.S.

Citation

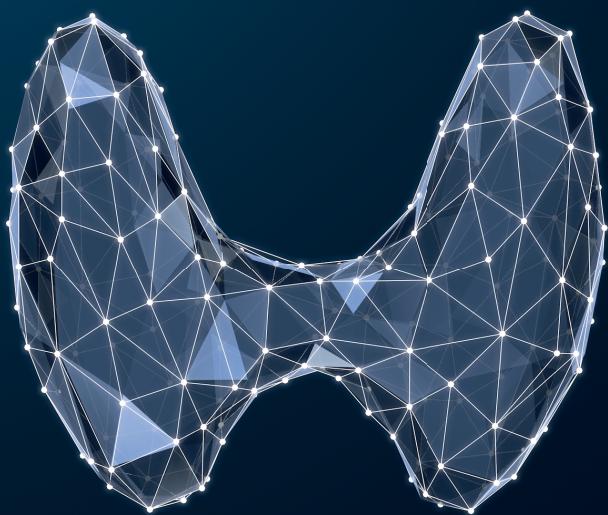
Du Puy, R. S. (2021, September 23). *Subclinical hypothyroidism in community-dwelling older people: consequences and treatment outcomes*. Retrieved from <https://hdl.handle.net/1887/3213499>

Version: Publisher's Version

[Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/3213499>

Note: To cite this publication please use the final published version (if applicable).



Chapter 2

Outcomes of thyroid dysfunction in people aged 80 years and older: an individual patient data meta-analysis of four prospective studies (TULIPS consortium)

Robert S Du Puy
Rosalinde KE Poortvliet
Simon P Mooijaart
Wendy PJ den Elzen
Carol Jagger
Simon HS Pearce
Yasumichi Arai
Nobuyoshi Hirose
Ruth Teh
Oliver Menzies
Anna Rolleston
Ngaire Kerse
Jacobijn Gussekloo for the TULIPS Consortium.

Thyroid, 2020; ahead of print

DOI: 10.1089/thy.2020.0567

ABSTRACT

Background

Subclinical and overt thyroid dysfunction is easily detectable, often modifiable and, in younger age groups, has been associated with clinically relevant outcomes. Robust associations in very old persons however are currently lacking. This study aimed to investigate the associations between (sub-)clinical thyroid dysfunction and disability in daily living, cognitive function, depressive symptoms, physical function and mortality in people aged 80 years and older.

Methods

Four prospective cohorts participating in the Towards Understanding Longitudinal International older People Studies (TULIPS) consortium were included. We performed a two-step Individual Participant Data meta-analysis on source data from community-dwelling participants aged 80 years and older from the Netherlands, New Zealand, United Kingdom and Japan. Outcome measures included disability in daily living (disability in activities of daily living questionnaires), cognitive function (MMSE), depressive symptoms (GDS), physical function (grip strength) at baseline and after 5 years of follow-up, and all- cause five-year mortality.

Results

Of the total 2,116 participants at baseline (mean age 87 years, range 80-109 years), 105 participants (5.0%) were overtly hypothyroid, 136 (6.4%) subclinically hypothyroid, 1,811 (85.6%) euthyroid, 60 (2.8%) subclinically hyperthyroid and 4 (0.2%) overtly hyperthyroid. Participants with thyroid dysfunction at baseline had non-significantly different activities of daily living scores compared to euthyroid participants at baseline and had similar MMSE scores, GDS scores and grip strength. There was no difference in the change of any of these functional measures in participants with thyroid dysfunction during five years of follow-up. Compared to the euthyroid participants, no 5-year survival differences were identified in participants with overt hypothyroidism (Hazard Ratio [HR] 1.0, 95% Confidence Interval [95%CI] 0.6 to 1.6), subclinical hypothyroidism (HR 0.9, 95%CI 0.7 to 1.2), subclinical hyperthyroidism (HR 1.1, 95%CI 0.8 to 1.7) and overt hyperthyroidism (HR 1.5, 95%CI 0.4 to 5.9). Results did not differ after excluding participants using thyroid-influencing medication.

Conclusions

In community-dwelling people aged 80 years and older, (sub-)clinical thyroid dysfunction was not associated with functional outcomes or mortality and may therefore be of limited clinical significance.

INTRODUCTION

Circulating levels of thyroid hormones influence cell metabolism of nearly all types of tissue in the human body.[1] Consequently, in young and middle-aged individuals both hypo- and hyperthyroidism have been associated with several adverse clinical outcomes.[2-5]

However, evidence is starting to accumulate that adverse health associations of thyroid dysfunction, found in younger age groups, cannot be extrapolated to persons aged 80 years and over.[6] Possibly due to changes in thyroid metabolism associated with aging. For example, in a recent meta-analysis of 13 studies investigating the association between subclinical hypothyroidism and cognitive decline or the risk of dementia, a significant risk was identified in participants younger than 75 years of age but not in participants older than 75 years.[7] In addition, clinical outcomes such as cardiovascular risk [8-10] and activities of daily living,[11] associated with thyroid dysfunction in younger age groups, have not been replicated in older adults.[12]

Researchers are currently unable to reach robust conclusions for the people aged 80 years and over due to methodological or logistic difficulties, such as low numbers of older participants included. In this study, we combine the individual participant data from four large, international prospective cohorts of community-dwelling very old persons, to investigate associations between (sub-)clinical thyroid dysfunction and disabilities in activities of daily living, functioning (including cognitive function, depressive symptoms and physical function) and mortality in people aged 80 years and older.

MATERIALS AND METHODS

Study Population

The Towards Understanding Longitudinal International older People Studies (TULIPS) Consortium was established in 2014 and set out to investigate determinants of successful aging and health outcomes in older persons using the data from large-scale, prospective, population-based cohort studies. The data for the analyses in this manuscript were provided by four international cohort studies: The Leiden 85-plus Study, The Life and Living in Advanced Age: a cohort study in New Zealand (LiLACS NZ) Study, Newcastle 85+ Study and Tokyo Centenarian Study (TCS).

The Leiden 85-plus Study

All inhabitants of Leiden, The Netherlands, reaching the age of 85 between September 1997 and September 1999 (N=705) were eligible for participation of which 599 participated.[13]

During a 5-year follow-up, participants were visited for interviews, performing of functional tests and the collection of venous blood samples. Thyroid stimulating hormone (TSH) and free thyroxine (fT4) at baseline were available for 553 participants.

LiLACS NZ Study

LiLACS NZ consists of two cohorts running in parallel, one with exclusively Māori (the indigenous people of New Zealand) and one with non-Māori participants.[14] All potential participants living in the Lakes or Bay of Plenty District Health Board areas in 2010 (aged 85 in 2010 for non-Māori or aged 80-90 in 2010 for Māori) were eligible (N=1,636) of which 937 were enrolled.[15] Participants were visited annually for a structured questionnaire, a physical health assessment and venous blood samples. TSH and fT4 at baseline were available for 195 participants in the Māori cohort and 352 participants in the non-Māori cohort.

Newcastle 85+ Study

All citizens born in 1921 registered with a participating general practitioner in Newcastle upon Tyne or North Tyneside primary care trusts, the United Kingdom, were approached in 2006 for participation (N=1,470), of which 851 were enrolled with complete health assessment and record review.[16] Participants received visits at baseline, at 18 months, at 36 months and at 60 months of follow-up, for the questionnaire collection, measurements and function tests and blood tests.[17] TSH and fT4 at baseline were available for 763 participants.

Tokyo Centenarian Study

Out of an estimated 1,735 centenarians living in Tokyo, Japan, 1,194 centenarians (68.8%) were randomly selected and approached with a survey between 2000 and 2002, of which 304 agreed to participate.[18] All participants were visited for blood sample collection, cognitive, mental and physical functioning assessments. TSH and fT4 at baseline were available for 251 participants.

Data collected

For all cohorts, baseline data per participant on socio-demographic characteristics (age, sex, ethnicity, current tobacco and alcohol use and education level) were collected. Follow-up data for thyroid-influencing medication use (levothyroxine, anti-thyroid medication, amiodarone, lithium, glucocorticosteroids or interferon- α) and the outcome variables described below were collected at every visit.

Thyroid Status

In the Leiden 85-plus study non-fasting blood samples were drawn between 8 and 11 am. and plasma TSH and fT4 levels were analysed using a Elecsys 2010 automated system (Hitachi, Tokyo, Japan).[19] For the LiLACS NZ cohort fasting blood samples were either drawn at time

of interview or in a local public laboratory. Thyroid assays were analysed using a Cobas immunoassay system (Roche, Switzerland). In the Newcastle 85+ Study, blood was drawn after an overnight fast before 10.30 am. and analysed using a Centaur chemiluminescent immunoassay system (Siemens, UK).[20] In the TCS study, non-fasting blood samples were obtained at time of interview and were stored at -80°C until the subsequent assay was performed. Serum TSH and fT4 concentrations were measured using a chemiluminescent enzyme immunoassay with a Lumipulse Forte fully automated system (Fuji Rebio, Tokyo, Japan).

For all studies, the reference ranges for TSH were defined as 0.3 mIU/L to 4.8 mIU/L for serum TSH and as 13 pmol/L to 23 pmol/L (1.01 to 1.79 ng/dL) for fT4. Five clinical strata of thyroid function at baseline were used to classify participants; overt hypothyroidism (TSH > 4.8 and fT4 < 13), subclinical hypothyroidism (TSH > 4.8 and fT4 within reference range), euthyroidism (0.3 ≤ TSH ≤ 4.8), subclinical hyperthyroidism (TSH < 0.3 and fT4 within reference range) and overt hyperthyroidism (TSH < 0.3 and fT4 > 23).

Outcome measures

Disability in activities of daily living

Disability in activities of daily living was measured in the Leiden 85-plus Study using the Groningen Activity Restriction Scale (GARS [21], range 18-72, higher summed scores more disability) consisting of 18 items of daily function and independent living. In the Newcastle 85+ study a summed disability score derived from 17 activities of daily living was used (range 0-17, higher summed scores more disability).[20] For the LiLACS NZ study the Nottingham Extended Activities of Daily Living Index (NEADL, range 0-22, higher summed scores less disability) [22] was used. For the TCS study, activities of daily living was assessed at baseline using the Barthel Index (BI, range 0-100, higher summed index scores less disability).[23] All disability in activities of daily living questionnaires used have been shown to be valid and reliable indicators of physical disability in older persons.[24-27]

Cognitive function

In all studies, cognitive function was assessed with the Mini-Mental State Examination (MMSE) questionnaire at every visit (maximum score of 30). The MMSE questionnaire has previously been validated in this age group.[28] Higher scores indicate better cognitive functioning. MMSE scores in TCS were available at baseline.

Depressive symptoms

The Geriatric Depression Scale (GDS-15) was used to evaluate a presence of depressive symptoms in the Leiden 85-plus, LiLACS NZ, and Newcastle 85+ cohort.[29] This questionnaire contains 15 items adding up to maximum summed score of 15, with higher scores indicating more depressive symptoms. The GDS-15 questionnaire has previously been validated in this

age group.[29] In the Leiden 85-plus and Newcastle 85+ cohort the GDS was not undertaken in individuals with established cognitive impairment. Depressive symptoms were not assessed in TCS.

Physical function

Hand grip strength measurements in kilogram were used as a proxy for overall muscle strength [30] and hand-held dynamometry has been validated in this age group for measuring muscle strength.[31] A Jamar hand dynamometer (Sammons Preston INC., Illinois, USA) was used in the Leiden 85-plus study. A Takei hand dynamometer Grip-D (Takei Scientific Instruments Co., Niigata-City, Japan) was used in the LiLACS NZ and Newcastle 85+ cohort. Grip strength was not assessed in TCS.

Mortality

Date of death was collected from the records of Statistics Netherlands (Centraal Bureau voor de Statistiek, The Hague) for the Leiden 85-plus Study. General practitioner records and mortality administrative data (held by the Ministry of Health) were consulted for the LiLACS NZ study. Information on mortality for the Newcastle 85+ Study was provided by NHS Digital, UK. For TCS, all-cause mortality was ascertained by telephone contact or mail survey conducted every 12 months.[32] Survival time was calculated as the time between the baseline visit and date of death or censored at 5-years if death had not yet occurred.

Statistical analyses

All measurements are analysed at yearly intervals from baseline (i.e. after 1, 2, 3, 4 and 5 years). Measurements for Phase 2 of the Newcastle 85+ Study, however, were taken 18 months post baseline and, to allow linear mixed model analyses, were assumed constant at the 2-year analysis. Results were reported separately for the Māori and non-Māori cohorts to do justice to potential differences in Māori and non-Māori preferences and practices in research.[33]

A two-stage Individual Participant Data (IPD) meta-analysis approach was employed for all analyses.[34,35] All variables for the analyses were coded similarly for each study, analysed on a study level using the same syntax and finally pooled for meta-analysis.

Study level analyses

Continuous variables were described as mean with standard deviation or as median with interquartile range (IQR), where appropriate, and categorical variables were presented as frequency with percentage of the total.

Five-year survival analysis was carried out using Cox proportional-hazards regression models, corrected for age and sex, and presented as a hazard ratio (HR) with 95% confidence intervals

(95%CI). The proportionality of hazards assumption was checked. Repeated measurements linear mixed-effect models using restricted maximum likelihood estimation and an unstructured covariance matrix, corrected for age and sex, were used to assess the associations between thyroid function and MMSE, GDS, grip strength and disability in activities of daily living.

Disability in activities of daily living scores were inverted for TCS (BI), and for LiLACS NZ, the mean was subtracted from the scores (NEADL), to standardise the direction of the scales, a prerequisite for pooling.[36] Considering that for disability in daily living the same outcome was measured using different scales in the cohorts, standardised mean differences (SMDs) were calculated for each study using independent-groups pretest- posttest calculations. [37] By dividing the mean outcome difference between a thyroid dysfunction group and the euthyroid group with the standard deviation among the participants, the size of the between group effect in each study is normalised relative to the variability observed in the study (See supplementary material).

All analyses used the euthyroid participants as the reference group or used TSH levels as a continuous independent variable.

Pooled analyses

All cohort effect estimates were subsequently pooled using random-effects models with inverse-variance weighting and summarised in forest plots. Inconsistency between cohorts due to heterogeneity was quantified using the I^2 -statistic and found not to be important for all summarised values ($I^2 < 40\%$).

Sensitivity analysis

All participants (n=236, 11%) using any form or combination of thyroid influencing medication (levothyroxine, anti-thyroid medication, amiodarone, lithium, glucocorticosteroids or interferon-a) were excluded for additional sensitivity analyses. The study-level analyses were performed using IBM SPSS Statistics version 22.0 (IBM, Armond, NY, USA). All pooled analyses were performed using Review Manager 5.3 (The Cochrane Collaboration, Copenhagen, Denmark).[38]

RESULTS

Table 1 displays the baseline characteristics of the combined study population and for the cohorts separately. The combined cohort consisted of 2,116 participants. Mean participant age was 87 (range 79 to 109) years and 1,311 (62%) were female. The median follow-up time over all cohorts was 5.3 (IQR 2.6 to 6.9) years.

Table 1. Baseline characteristics of study participants.

Study	Total	Leiden 85-plus Study		LiLACS NZ Study		Newcastle 85+ Study		Tokyo Centenarian Study
		Māori	Non-Māori	352 (64.3%)	763	251		
N=	2,116	555	195 (35.7%)					
Study Population:								
		Longitudinal cohort of adults aged 85 (born between 1912 and 1914) in Leiden, the Netherlands	Longitudinal cohort of Māori aged 80-90 (born between 1920 and 1930) in New Zealand	Longitudinal cohort of Non-Māori aged 85 (born in 1925) in New Zealand	Longitudinal cohort in over-85s (born in 1921) in Newcastle upon Tyne, UK	Longitudinal cohort in over-85s (born in 1921) in Newcastle upon Tyne, UK	Dynamic cohort study of centenarians (born before 1902) living in Tokyo, Japan	
Socio-demographic characteristics:								
Age, mean (range)	87 (80-109)	85	82 (80-84)	85 (84-86)	85 (85-86)	85 (85-86)	102 (100-109)	
Female, no. (%)	1,311 (62.0)	366 (65.9)	108 (55.4)	177 (50.3)	465 (60.3)	195 (77.7)		
Follow-up time, median (IQR), years	5.3 (2.6-6.9)	5.4 (2.7-8.6)	6.0 (4.1-6.2)	6.0 (4.6-6.2)	5.4 (2.6-7.7)	2.1 (1.0-3.8)		
Current smoker, no. (%)	169 (8.0)	88 (15.9)	20 (10.3)	16 (4.5)	42 (5.5)	3 (1.2)		
Current alcohol user, no. (%)	1,156 (54.6)	279 (50.3)	106 (54.4)	257 (73.0)	464 (60.8)	50 (19.9)		
Laboratory measurements, median (IQR):								
TSH, mIU/L	2.1 (1.3-3.2)	1.8 (1.2-2.9)	2.1 (1.3-3.2)	2.6 (1.7-3.6)	2.0 (1.2-3.1)	2.3 (1.6-3.9)		
fT4, pmol/L	14.9 (13.2-16.5)	14.3 (12.8-15.8)	15.2 (14.0-16.6)	15.1 (13.9-16.9)	15.0 (14.0-17.0)	13.4 (11.7-14.8)		
Thyroid function, no. (%):								
Overt hypothyroidism	105 (5.0)	37 (6.7)	4 (2.1)	10 (2.8)	19 (2.5)	35 (13.9)		
Subclinical hypothyroidism	136 (6.4)	29 (5.2)	12 (6.2)	30 (8.5)	55 (7.2)	10 (4.0)		
Euthyroidism	1,811 (85.6)	471 (84.9)	172 (88.2)	307 (87.2)	656 (86.0)	205 (81.7)		
Subclinical hyperthyroidism	60 (2.8)	16 (2.9)	7 (3.6)	5 (1.4)	31 (4.1)	1 (0.4)		
Overt hyperthyroidism	4 (0.2)	2 (0.4)	-	-	2 (0.3)	-		

Functional parameters, median (IQR):

Disability in activities of daily living	28 (22-40)	19 (16-20)	19 (17-20)	3.0 (1-7)	4.5 (15-80)
Disability in activities of daily living questionnaire	GARS	NEADL	NEADL	ADL sum scores	BI
Mini Mental State Examination score	27 (23-29)	26 (22-28)	28 (26-29)	28 (27-29)	28 (25-29)
Geriatric Depression Scale score	1 (1-3)	2 (1-3)	2 (1-3)	2 (1-3)	1 (1-2)
Grip Strength (Kg)	21 (16-28)	21 (17-28)	25 (19-30)	23 (18-31)	19 (15-26)

Abbreviations: IQR, interquartile range; GARS, Groningen Activity Restriction Scale; NEADL, Nottingham Extended Activities of Daily Living; BI, Barthel Index.

The median serum TSH level in the combined cohort was 2.1 (IQR 1.3 to 3.2) mIU/L and median serum fT4 was 14.9 (IQR 13.2 to 16.5) pmol/L, both within the normal reference ranges. One thousand eight hundred and eleven (85.6%) participants were euthyroid. Thyroid dysfunction was found in 305 (14.4%) participants; 136 (6.4%) participants with subclinical hypothyroidism, 105 (5.0%) with overt hypothyroidism, 60 (2.8%) with subclinical hyperthyroidism and 4 (0.2%) with overt hyperthyroidism. In the hypothyroidism categories 18 (17.1%) participants with overt hypothyroidism and 41 (30.1%) of participants with subclinical hypothyroidism used a preparation of thyroid hormone at any point during follow-up.

The median MMSE score was 27 points (IQR 23 to 29) and median GDS score 1 point (IQR 1 to 2). Median disability in activities of daily living scores were 28 (IQR 22 to 40) on the GARS questionnaire (Leiden 85-plus Study); 19 (IQR 16 to 20) for the Māori subgroup and 19 (IQR 17 to 20) for the non-Māori subgroup on the NEADL questionnaire (LILACS NZ); 3.0 (IQR 1 to 7) on the ADL sum scores (Newcastle 85+ Study) and 45 (IQR 15 to 80) on the Barthel Index questionnaire (TCS).

Figure 1 shows activities of daily living scores according to clinical thyroid state at baseline and after 5 years of follow-up. At baseline no significant standardised mean differences in activity of daily living scores were identified in the clinical strata with thyroid dysfunction compared to the euthyroid participants; mean difference 0.01 (95%CI -0.19 to 0.21) points for overt hypothyroidism, -0.00 (95%CI -0.18 to 0.17) points for subclinical hypothyroidism, 0.00 (95%CI -0.25 to 0.26) points for subclinical hyperthyroidism and 0.01 (95%CI -0.97 to 1.00) points for overt hypothyroidism. Although all groups showed decline in activities of daily living, over time there were no significant standardised mean differences of decline in activities of daily living scores (SMD between -0.01 and 0.00, IQR between -0.99 and 0.97) between euthyroid participants and participants with thyroid dysfunction. In this study population a calculated SMD of -0.01 corresponds with a 3.5-point decrease on the GARS scale.

At baseline, estimated marginal means of MMSE scores did not differ significantly between participants with thyroid dysfunction (means between 23.6 points [95%CI 18.2 to 29.0] and 26.3 points [95%CI 25.5 to 27.4]) and participants with euthyroidism (mean 23.7 points [95%CI 21.0 to 26.5]). Similarly, no differences were found for depressive symptom scores (means between 1.9 points [95%CI 1.5 to 2.3] and 5.1 points [95%CI 2.7 to 7.5] for participants with thyroid dysfunction, and 2.1 points [95%CI 1.4 to 2.7] for euthyroid participants) and grip strength (means between 18.1 kg [95%CI 9.8 to 26.4] and 26.0 kg [95%CI 22.6 to 29.5] for participants with thyroid dysfunction, and 24.1 kg [95%CI 23.1 to 25.1] for euthyroid participants) (Table 2).

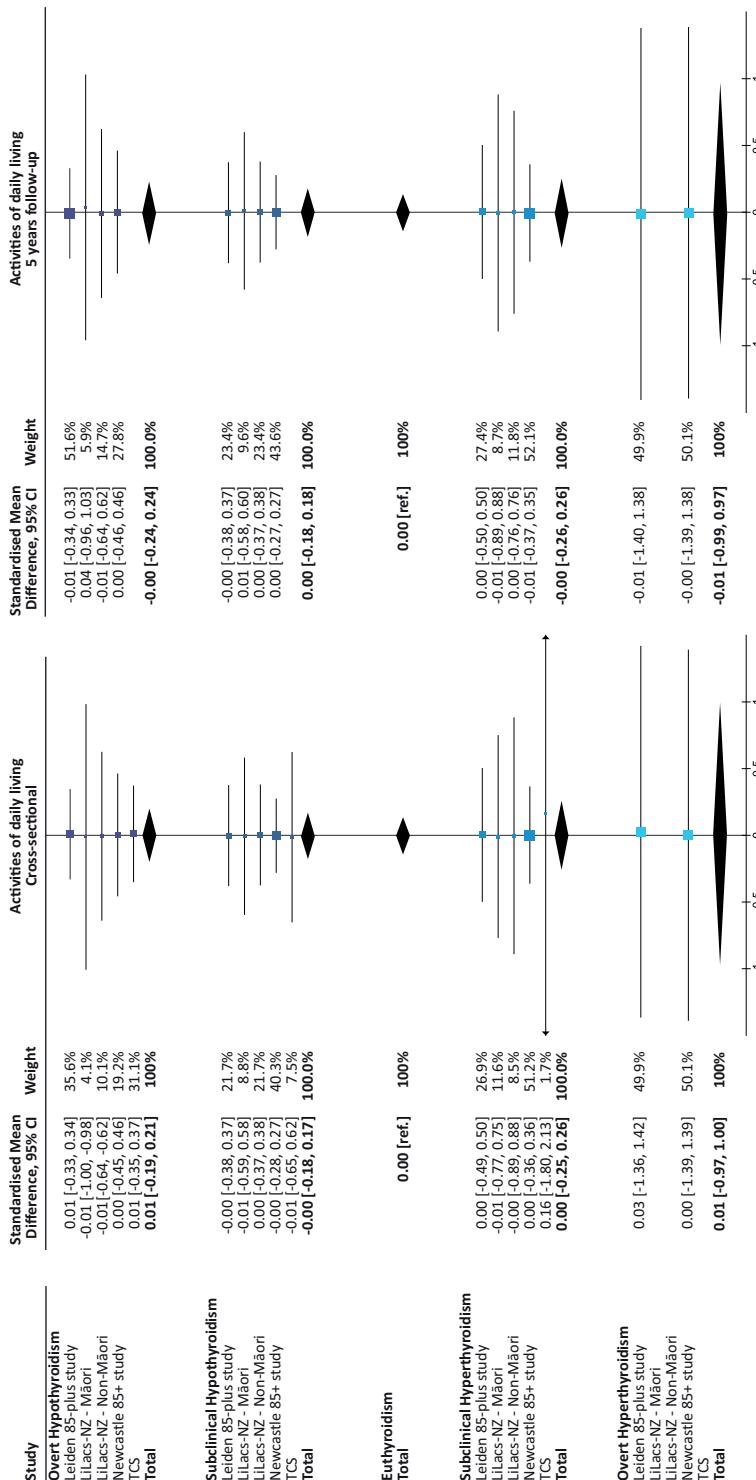


Figure 1. Activities of daily living per clinical thyroid stratum at baseline and after five years of follow-up. Standardized mean differences calculated per clinical thyroid stratum and per cohort using independent-groups pretest-post-test calculations compared with the euthyroid participants. Higher scores represent worse outcomes. Abbreviations: IQR, interquartile range; GARS, Groningen Activity Restriction Scale; NEADL, Nottingham Extended Activities of Daily Living; BI, Barthel Index.

Table 2. Functional parameters per clinical thyroid stratum at baseline.

	Overt Hypothyroidism (N=105)	Subclinical hypothyroidism (N=136)	Euthyroidism (N=1181)	Subclinical hyperthyroidism (N=60)	Overt Hyperthyroidism (N=4)
Baseline^a					
Cognitive function (MMSE)	23.7 (20.4, 27.0)	25.5 (23.6, 27.5)	23.7 (21.0, 26.5)	26.3 (25.2, 27.4)	23.6 (18.2, 29.0)
Depressive symptoms (GDS)	2.1 (1.6, 2.7)	1.9 (1.5, 2.3)	2.1 (1.4, 2.7)	2.4 (1.9, 3.0)	5.1 (2.7, 7.5)
Physical function (grip strength)	26.0 (22.6, 29.5)	25.3 (24.0, 25.3)	24.1 (23.1, 25.1)	23.3 (21.3, 25.3)	18.1 (9.8, 26.4)
5-year follow-up^b					
Cognitive function (MMSE)	-0.4 (-1.4, 0.5)	0.4 (-0.3, 1.0)	0 (ref.)	0.1 (-0.9, 1.1)	0.6 (-0.9, 2.1)
Depressive symptoms (GDS)	-0.1 (-0.6, 0.3)	0.0 (-0.4, 0.3)	0 (ref.)	0.3 (-0.2, 0.7)	1.7 (-0.2, 3.5)
Physical function (grip strength)	1.7 (-1.9, 5.2)	1.4 (-0.4, 2.1)	0 (ref.)	-0.7 (-2.5, 1.0)	-3.3 (-10.0, 11.3)

^a Estimated Margin Means (95% Confidence Intervals) estimated using Linear Mixed models per cohort, adjusted for baseline age and sex, and pooled using random-effects models with inverse-variance weighting.

^b Beta's (95% Confidence Intervals) estimated using repeated measures Linear Mixed models per cohort, adjusting for baseline age and sex, and pooled using random-effects models with inverse-variance weighting, representing change in outcome measure over five years of follow-up. Euthyroidism is the reference group.

Over time, no differences were identified in decline in MMSE scores for participants with thyroid dysfunction (means between -0.4 points less decline [95%CI -1.4 to 0.5] and 0.4 points more decline [95%CI -0.9 to 2.1]) compared to euthyroid patients. Nor were differences identified for depressive symptoms scores (means between -0.1 points less decline [95%CI -0.6 to 0.3] and 1.7 points more decline [95%CI -0.2 to 3.5] for thyroid dysfunction compared to euthyroid participants) and grip strength (means between -3.3 [95%CI -10.0 to 11.3] and 1.7 [95%CI -1.9 to 5.2] for thyroid dysfunction compared to euthyroid participants) (Table 2). Forest plots for all thyroid categories stratified per cohort are presented in the supplementary material.

Pooled HRs were calculated to estimate five-year mortality (Figure 2). No difference in survival risks were found for overt hypothyroidism (HR 0.98, 95%CI 0.63 to 1.51), subclinical hypothyroidism (HR 0.89, 95%CI 0.67 to 1.17), subclinical hyperthyroidism (HR 1.11, 95%CI 0.75 to 1.65) or overt hyperthyroidism (HR 1.47, 95%CI 0.37 to 5.92). Similarly, in a cox-regression survival analysis across the clinical thyroid strata, TSH as a continuous variable did not predict 5-year mortality (HR 0.99, 95%CI 0.96 to 1.02).

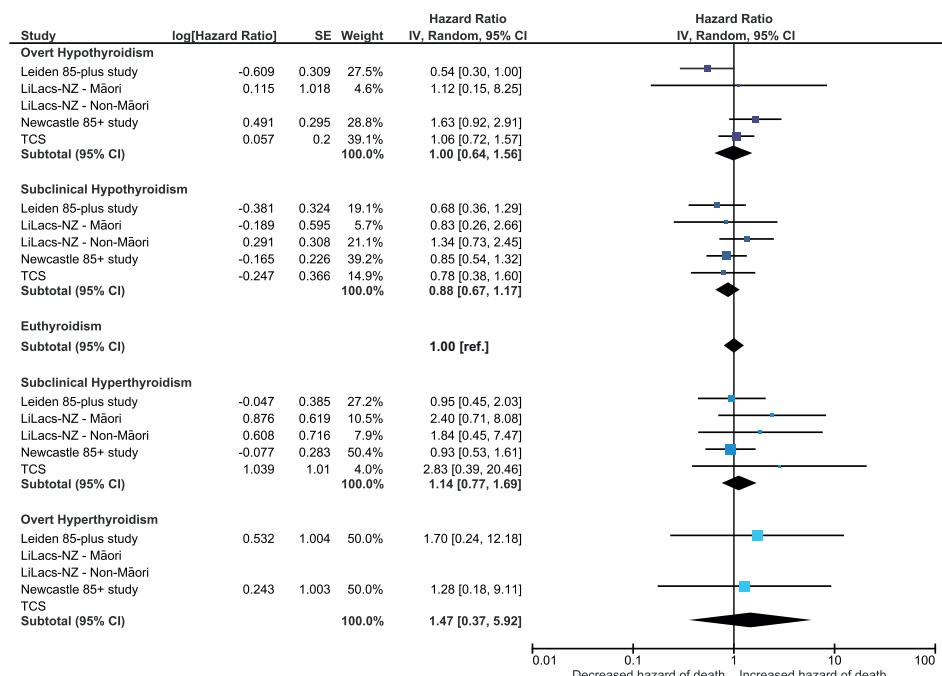


Figure 2. All-cause five-year mortality per clinical thyroid stratum at baseline. Pooled hazard ratios [95% CI] per clinical thyroid stratum estimated using Cox proportional hazards regression models adjusting for sex and age at baseline compared with participants with euthyroidism. When mortality rates were too low hazard ratios could not be estimated. Abbreviations: CI, confidence interval.

In sensitivity analyses, excluding all participants (n=236, 11%) using any or a combination of thyroid influencing medications at any point during follow-up, did not alter the association estimates substantially in either direction of effect or significance (data not shown).

DISCUSSION

The findings of this large IPD meta-analysis of 4 prospective observational cohorts among community-dwelling participants aged 80 years and older are two-fold. First, there was no association of thyroid dysfunction with disability in activities of daily living, nor with cognition, physical function or depressive symptoms, at baseline or during follow-up. Second, thyroid dysfunction was not associated with increased mortality during an average 5 years of follow-up.

Earlier research investigating the outcomes of thyroid dysfunction in old age yielded greatly conflicting data. For example, some studies demonstrated decreased cognitive function,[39-41] increased depressive symptoms,[42,43] decreased physical function [44] and increased cardiovascular or all-cause mortality [45-47] in participants with subclinical or overt thyroid dysfunction. Others, however, were unable to replicate these findings.[10,48-58] Due to differences in, for example, the thyroid reference and age-ranges used, follow-up time chosen, the participants selected, study design employed, and outcome measures used, the aforementioned studies could not be pooled without suffering from significant methodological heterogeneity. This study was optimised to address these specific limitations, by pooling the results of four of the most comparable cohort studies of community-dwelling oldest old.

Our findings are in line with most previous publications from the individual datasets used in this study, such as an absence of association between thyroid function and disabilities in daily living, cognitive function and depressive symptoms.[19] Additionally no association was found between fT4 or TSH and all-cause mortality in earlier studies.[20,59] However, some nonunanimous results have been described using these datasets as well. For instance, in the Leiden 85-plus study increasing levels of TSH were associated with lower mortality rates.[19] It should be noted that the aggregate negative findings in our pooled analyses do not invalidate any earlier identified associations in any of the individual cohorts or other publications, as these may be a representation of associations in that specific population being sampled, at that time, using those specific methods, study designs and analyses. Nevertheless, the findings from this IPD meta-analysis, do more accurately reflect an estimation of effects for a more universal population of older people, with uniform definitions and follow-up times, by pooling the results and allowing statistically for sampling error and biases. These findings

suggest that the influence of thyroid dysfunction on disabilities in daily living, cognitive function, depressive symptoms, physical function and mortality in persons aged 80 years and older is limited, particularly in the subclinical thyroid disease subgroups.

For older age groups with subclinical thyroid dysfunction, evidence supporting beneficial effects from treatment has generally been of lower grade and contradictory, as acknowledged in international guidelines. Currently, these recommend routine or trial treatment in patients older than 65 with subclinical thyroid dysfunction, albeit individualised, gradual and closely monitored.[60,61] However, they simultaneously acknowledge the lack of evidence in scientific literature and call for large prospective studies with a long-term follow-up in older persons. Recently two international multicentre trials demonstrated no beneficial effects of levothyroxine treatment for subclinical hypothyroidism in persons aged 65 years and over, and in persons aged 80 years and over.[62,63]

The lack of associations between (mild) thyroid dysfunction and relevant outcomes described in this manuscript, and the absence of beneficial effects of levothyroxine treatment for subclinical hypothyroidism in older persons, reinforces the notion that for older persons with subclinical hypothyroidism, routine treatment does not seem indicated and that the thresholds for treatment initiation (i.e. currently the normal reference ranges) ought to be redefined. This is not necessarily the same as the definition of the reference ranges. Whether all thyroid reference ranges in aging need to be re-evaluated to distinguish physiologic states from pathophysiologic ones is still a much-debated issue and will require further evaluation. Future research, focusing on how thyroid-related (patho-) physiological processes change with aging and how these affect outcomes, are needed, including the influence of other established thyroid markers and treatment modalities. This study used data from some of the largest, unselected, population-based observational cohorts of community-dwelling persons aged 80 years and older to date, with considerable follow-up time and a substantial international representation of countries, cultures and persons. The IPD meta-analysis design allowed for harmonization of thyroid function categories, standardization of definitions and variables, and consistency in analyses. The two-step design allowed the presentation of overall results on an aggregate level, while maintaining maximum transparency on any potential inter-cohort variation, for example between Māori and non-Māori populations.

Some limitations do need to be acknowledged. First, TSH and fT4 were only available for the baseline measurements and it could be possible that thyroid function changed during the observational periods. However, a recent study in almost 3,000 community dwelling older persons has demonstrated high stability of thyroid function over 5 years of follow-up and therefore we estimate the odds of erroneous misclassification as small.[64] Second, the activities of daily living questionnaires were too heterogeneous in scale and structure to pool in any

traditional way. Standardization with the Standardised Mean Difference- method allowed for methodologically sound pooling but may have introduced additional uncertainty to the analyses leading to underestimation of true associations.

In participants aged 80 years and older, (sub-)clinical thyroid dysfunction is not associated with functional outcomes or mortality and may therefore be of limited clinical significance.

REFERENCES

1. Mullur R, Liu YY, Brent GA. Thyroid hormone regulation of metabolism. *Physiol Rev*. 2014;94(2):355-382.
2. Beydoun MA, Beydoun HA, Kitner-Triolo MH, Kaufman JS, Evans MK, Zonderman AB. Thyroid hormones are associated with cognitive function: moderation by sex, race, and depressive symptoms. *J Clin Endocrinol Metab*. 2013;98(8):3470-3481.
3. Hage MP, Azar ST. The Link between Thyroid Function and Depression. *J Thyroid Res*. 2012;2012:590648.
4. Laulund AS, Nybo M, Brix TH, Abrahamsen B, Jorgensen HL, Hegedus L. Duration of thyroid dysfunction correlates with all-cause mortality. the OPENTHYRO Register Cohort. *PLoS One*. 2014;9(10):e110437.
5. Mainenti MR, Vigario PS, Teixeira PF, Maia MD, Oliveira FP, Vaisman M. Effect of levothyroxine replacement on exercise performance in subclinical hypothyroidism. *J Endocrinol Invest*. 2009;32(5):470-473.
6. Ruggeri RM, Trimarchi F, Biondi B. MANAGEMENT OF ENDOCRINE DISEASE: I-Thyroxine replacement therapy in the frail elderly: a challenge in clinical practice. *European journal of endocrinology / European Federation of Endocrine Societies*. 2017;177(4):R199-R217.
7. Pasqualetti G, Pagano G, Rengo G, Ferrara N, Monzani F. Subclinical Hypothyroidism and Cognitive Impairment: Systematic Review and Meta-Analysis. *J Clin Endocrinol Metab*. 2015;100(11):4240-4248.
8. Chaker L, Baumgartner C, den Elzen WP, et al. Subclinical Hypothyroidism and the Risk of Stroke Events and Fatal Stroke: An Individual Participant Data Analysis. *J Clin Endocrinol Metab*. 2015;100(6):2181-2191.
9. Pasqualetti G, Tognini S, Polini A, Caraccio N, Monzani F. Is subclinical hypothyroidism a cardiovascular risk factor in the elderly? *J Clin Endocrinol Metab*. 2013;98(6):2256-2266.
10. Razvi S, Shakoor A, Vanderpump M, Weaver JU, Pearce SH. The influence of age on the relationship between subclinical hypothyroidism and ischemic heart disease: a metaanalysis. *J Clin Endocrinol Metab*. 2008;93(8):2998-3007.
11. SO ES, Chan IT, Lobo Santos MA, et al. Impact of thyroid status and age on comprehensive geriatric assessment. *Endocrine*. 2014;47(1):255-265.
12. Mariotti S. Thyroid function and aging: do serum 3,5,3'-triiodothyronine and thyroid-stimulating hormone concentrations give the Janus response? *J Clin Endocrinol Metab*. 2005;90(12):6735-6737.
13. der Wiel AB, van Exel E, de Craen AJ, et al. A high response is not essential to prevent selection bias: results from the Leiden 85-plus study. *J Clin Epidemiol*. 2002;55(11):1119-1125.
14. Hayman KJ, Kerse N, Dyall L, et al. Life and living in advanced age: a cohort study in New Zealand—e Puawaitanga o Nga Tapuwae Kia Ora Tonu, LiLACS NZ: study protocol. *BMC Geriatr*. 2012;12:33.
15. Kerse N, Teh R, Moyes SA, et al. Cohort Profile: Te Puawaitanga o Nga Tapuwae Kia Ora Tonu, Life and Living in Advanced Age: a Cohort Study in New Zealand (LiLACS NZ). *Int J Epidemiol*. 2015;44(6):1823-1832.
16. Collerton J, Davies K, Jagger C, et al. Health and disease in 85 year olds: baseline findings from the Newcastle 85+ cohort study. *BMJ*. 2009;339:b4904.
17. Collerton J, Barratt K, Bond J, et al. The Newcastle 85+ study: biological, clinical and psychosocial factors associated with healthy ageing: study protocol. *BMC Geriatr*. 2007;7:14.

18. Gondo Y, Hirose N, Arai Y, et al. Functional status of centenarians in Tokyo, Japan: developing better phenotypes of exceptional longevity. *J Gerontol A Biol Sci Med Sci*. 2006;61(3):305-310.
19. Gussekloo J, van Exel E, de Craen AJ, Meinders AE, Frolich M, Westendorp RG. Thyroid status, disability and cognitive function, and survival in old age. *JAMA*. 2004;292(21):2591-2599.
20. Pearce SH, Razvi S, Yadegarfar ME, et al. Serum Thyroid Function, Mortality and Disability in Advanced Old Age: The Newcastle 85+ Study. *J Clin Endocrinol Metab*. 2016;101(11):4385-4394.
21. Kempen GI, Miedema I, Ormel J, Molenaar W. The assessment of disability with the Groningen Activity Restriction Scale. Conceptual framework and psychometric properties. *Social science & medicine (1982)*. 1996;43(11):1601-1610.
22. Essink-Bot ML, Krabbe PF, Bonsel GJ, Aaronson NK. An empirical comparison of four generic health status measures. The Nottingham Health Profile, the Medical Outcomes Study 36-item Short-Form Health Survey, the COOP/WONCA charts, and the EuroQol instrument. *Med Care*. 1997;35(5):522-537.
23. Mahoney FI, Barthel DW. Functional Evaluation: The Barthel Index. *Maryland state medical journal*. 1965;14:61-65.
24. Hopman-Rock M, van Hirtum H, de Vreede P, Freiberger E. Activities of daily living in older community-dwelling persons: a systematic review of psychometric properties of instruments. *Aging Clin Exp Res*. 2019;31(7):917-925.
25. Kingston A, Collerton J, Davies K, Bond J, Robinson L, Jagger C. Losing the ability in activities of daily living in the oldest old: a hierarchic disability scale from the Newcastle 85+ study. *PLoS One*. 2012;7(2):e31665.
26. Lin SY, Kerse N, McLean C, Moyes SA. Validation of quality of life and functional measures for older people for telephone administration. *J Prim Health Care*. 2010;2(1):35-42.
27. van Houwelingen AH, Cameron ID, Gussekloo J, et al. Disability transitions in the oldest old in the general population. The Leiden 85-plus study. *Age (Dordr)*. 2014;36(1):483-493.
28. Heeren TJ, Lagaay AM, von Beek WC, Rooymans HG, Hijmans W. Reference values for the Mini-Mental State Examination (MMSE) in octo- and nonagenarians. *J Am Geriatr Soc*. 1990;38(10):1093-1096.
29. de Craen AJ, Heeren TJ, Gussekloo J. Accuracy of the 15-item geriatric depression scale (GDS-15) in a community sample of the oldest old. *Int J Geriatr Psychiatry*. 2003;18(1):63-66.
30. Bohannon RW. Dynamometer measurements of hand-grip strength predict multiple outcomes. *Percept Mot Skills*. 2001;93(2):323-328.
31. Abizanda P, Navarro JL, Garcia-Tomas MI, Lopez-Jimenez E, Martinez-Sanchez E, Paterna G. Validity and usefulness of hand-held dynamometry for measuring muscle strength in community-dwelling older persons. *Arch Gerontol Geriatr*. 2012;54(1):21-27.
32. Arai Y, Inagaki H, Takayama M, et al. Physical independence and mortality at the extreme limit of life span: supercentenarians study in Japan. *J Gerontol A Biol Sci Med Sci*. 2014;69(4):486-494.
33. Bishop R. Kaupapa Māori research: An indigenous approach to creating knowledge. *Robertson, N (Ed)*. 1999;Maori and psychology : research and practice(The proceedings of a symposium sponsored by the Maori and Psychology Research Unit):1-6.
34. Riley RD, Lambert PC, Abo-Zaid G. Meta-analysis of individual participant data: rationale, conduct, and reporting. *BMJ*. 2010;340:c221.
35. Stewart LA, Tierney JF. To IPD or not to IPD? Advantages and disadvantages of systematic reviews using individual patient data. *Eval Health Prof*. 2002;25(1):76-97.
36. Higgins JPT, Green S. Cochrane Handbook for Systematic Reviews of Interventions. In: The Cochrane Collaboration; 2011: www.handbook.cochrane.org. Accessed updated March 2011.

37. Morris SB, DeShon RP. Combining effect size estimates in meta-analysis with repeated measures and independent-groups designs. *Psychol Methods*. 2002;7(1):105-125.

38. *Review Manager (RevMan)* [computer program]. Version 5.3. Copenhagen: The Nordic Cochrane Centre; 2014.

39. Kalmijn S, Mehta KM, Pols HA, Hofman A, Drexhage HA, Breteler MM. Subclinical hyperthyroidism and the risk of dementia. The Rotterdam study. *Clin Endocrinol (Oxf)*. 2000;153(6):733-737.

40. Tan ZS, Beiser A, Vasan RS, et al. Thyroid function and the risk of Alzheimer disease: the Framingham Study. *Arch Intern Med*. 2008;168(14):1514-1520.

41. van Osch LA, Hogervorst E, Combrinck M, Smith AD. Low thyroid-stimulating hormone as an independent risk factor for Alzheimer disease. *Neurology*. 2004;62(11):1967-1971.

42. Blum MR, Wijsman LW, Virgini VS, et al. Subclinical Thyroid Dysfunction and Depressive Symptoms among the Elderly: A Prospective Cohort Study. *Neuroendocrinology*. 2016;103(3-4):291-299.

43. Medici M, Direk N, Visser WE, et al. Thyroid function within the normal range and the risk of depression: a population-based cohort study. *J Clin Endocrinol Metab*. 2014;99(4):1213-1219.

44. Bano A, Chaker L, Darweesh SK, et al. Gait patterns associated with thyroid function: The Rotterdam Study. *Sci Rep*. 2016;6:38912.

45. Haentjens P, Van Meerhaeghe A, Poppe K, Velkeniers B. Subclinical thyroid dysfunction and mortality: an estimate of relative and absolute excess all-cause mortality based on time-to-event data from cohort studies. *European journal of endocrinology / European Federation of Endocrine Societies*. 2008;159(3):329-341.

46. Iervasi G, Molinaro S, Landi P, et al. Association between increased mortality and mild thyroid dysfunction in cardiac patients. *Arch Intern Med*. 2007;167(14):1526-1532.

47. Sgarbi JA, Matsumura LK, Kasamatsu TS, Ferreira SR, Maciel RM. Subclinical thyroid dysfunctions are independent risk factors for mortality in a 7.5-year follow-up: the Japanese-Brazilian thyroid study. *European journal of endocrinology / European Federation of Endocrine Societies*. 2010;162(3):569-577.

48. de Jong FJ, den Heijer T, Visser TJ, et al. Thyroid hormones, dementia, and atrophy of the medial temporal lobe. *J Clin Endocrinol Metab*. 2006;91(7):2569-2573.

49. de Jongh RT, Lips P, van Schoor NM, et al. Endogenous subclinical thyroid disorders, physical and cognitive function, depression, and mortality in older individuals. *European journal of endocrinology / European Federation of Endocrine Societies*. 2011;165(4):545-554.

50. Grabe HJ, Volzke H, Ludemann J, et al. Mental and physical complaints in thyroid disorders in the general population. *Acta Psychiatr Scand*. 2005;112(4):286-293.

51. Ochs N, Auer R, Bauer DC, et al. Meta-analysis: subclinical thyroid dysfunction and the risk for coronary heart disease and mortality. *Ann Intern Med*. 2008;148(11):832-845.

52. Roberts LM, Pattison H, Roalfe A, et al. Is subclinical thyroid dysfunction in the elderly associated with depression or cognitive dysfunction? *Ann Intern Med*. 2006;145(8):573-581.

53. Simonsick EM, Chia CW, Mammen JS, Egan JM, Ferrucci L. Free Thyroxine and Functional Mobility, Fitness, and Fatigue in Euthyroid Older Men and Women in the Baltimore Longitudinal Study of Aging. *J Gerontol A Biol Sci Med Sci*. 2016;71(7):961-967.

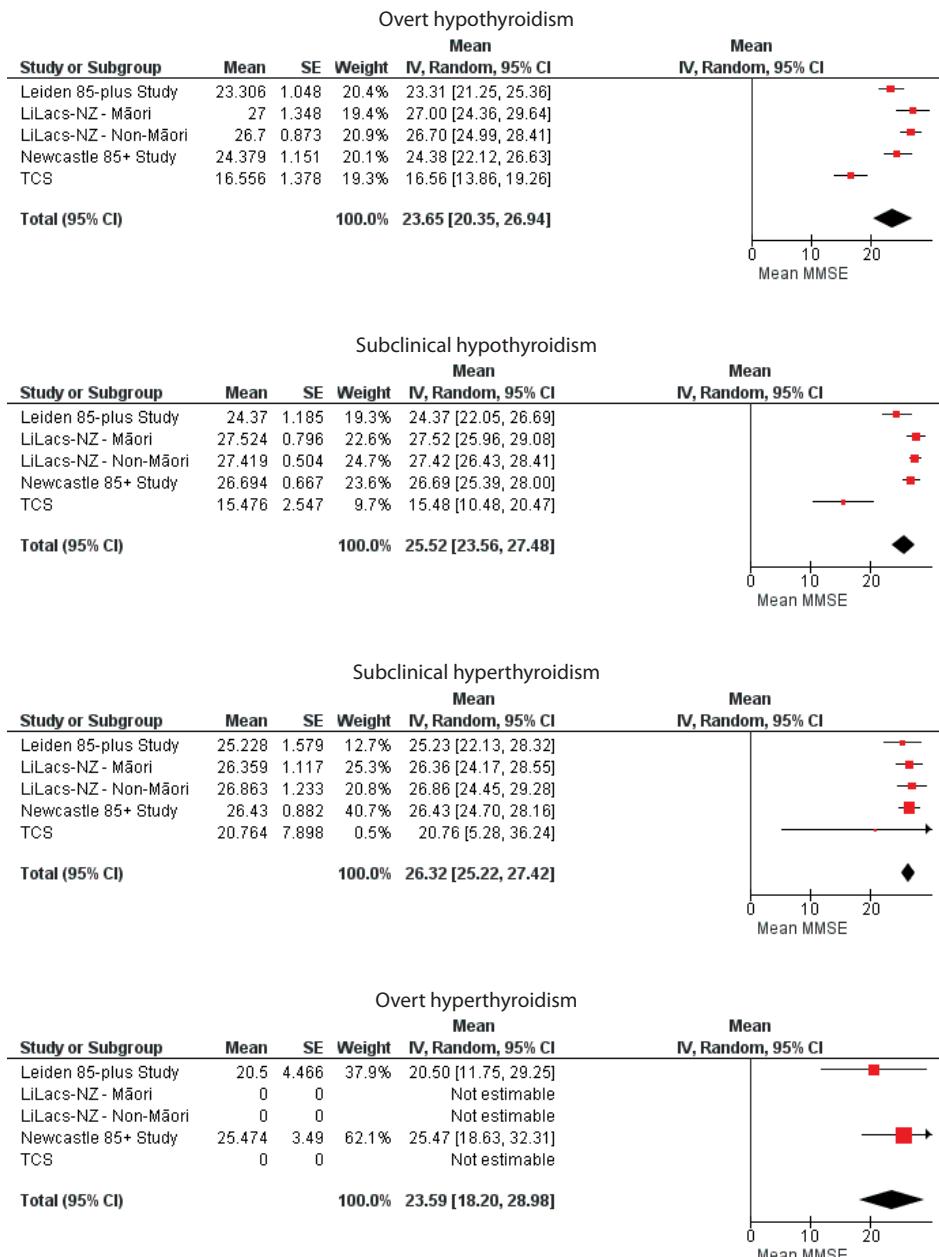
54. Simonsick EM, Newman AB, Ferrucci L, et al. Subclinical hypothyroidism and functional mobility in older adults. *Arch Intern Med*. 2009;169(21):2011-2017.

55. Singh S, Duggal J, Molnar J, Maldonado F, Barsano CP, Arora R. Impact of subclinical thyroid disorders on coronary heart disease, cardiovascular and all-cause mortality: a meta-analysis. *Int J Cardiol*. 2008;125(1):41-48.

56. van der Cammen TJ, Mattace-Raso F, van Harskamp F, de Jager MC. Lack of association between thyroid disorders and Alzheimer's disease in older persons: a cross-sectional observational study in a geriatric outpatient population. *J Am Geriatr Soc.* 2003;51(6):884.
57. Virgini VS, Wijsman LW, Rodondi N, et al. Subclinical thyroid dysfunction and functional capacity among elderly. *Thyroid.* 2014;24(2):208-214.
58. Volzke H, Schwahn C, Wallaschofski H, Dorr M. Review: The association of thyroid dysfunction with all-cause and circulatory mortality: is there a causal relationship? *J Clin Endocrinol Metab.* 2007;92(7):2421-2429.
59. van Vliet NA, van der Spoel E, Beekman M, et al. Thyroid status and mortality in nonagenarians from long-lived families and the general population. *Aging (Albany NY).* 2017;9(10):2223-2234.
60. Biondi B, Bartalena L, Cooper DS, Hegedus L, Laurberg P, Kahaly GJ. The 2015 European Thyroid Association Guidelines on Diagnosis and Treatment of Endogenous Subclinical Hyperthyroidism. *Eur Thyroid J.* 2015;4(3):149-163.
61. Pearce SH, Brabant G, Duntas LH, et al. 2013 ETA Guideline: Management of Subclinical Hypothyroidism. *Eur Thyroid J.* 2013;2(4):215-228.
62. Mooijaart SP, Du Puy RS, Stott DJ, et al. Association Between Levothyroxine Treatment and Thyroid-Related Symptoms Among Adults Aged 80 Years and Older With Subclinical Hypothyroidism. *JAMA.* 2019;1-11.
63. 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.
64. Roberts L, McCahon D, Johnson O, Haque MS, Parle J, Hobbs FR. Stability of thyroid function in older adults: the Birmingham Elderly Thyroid Study. *Br J Gen Pract.* 2018;68(675):e718-e726.

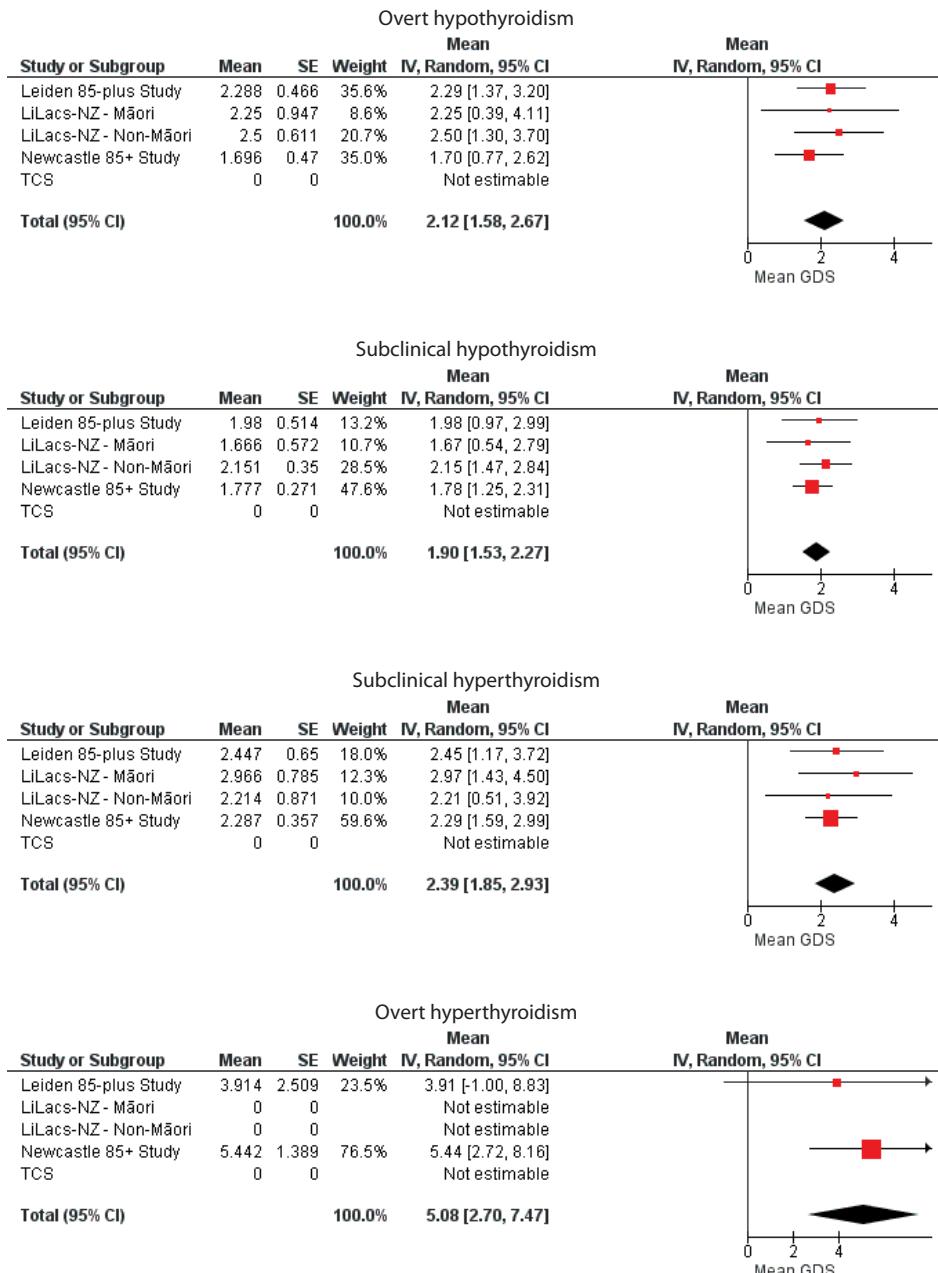
Supplementary data. Cross-sectional forest plots: MMSE

Estimated Marginal Means (95% Confidence Intervals) estimated using Linear Mixed models per cohort, adjusted for baseline age and sex, and pooled using random-effects models with inverse-variance weighting.



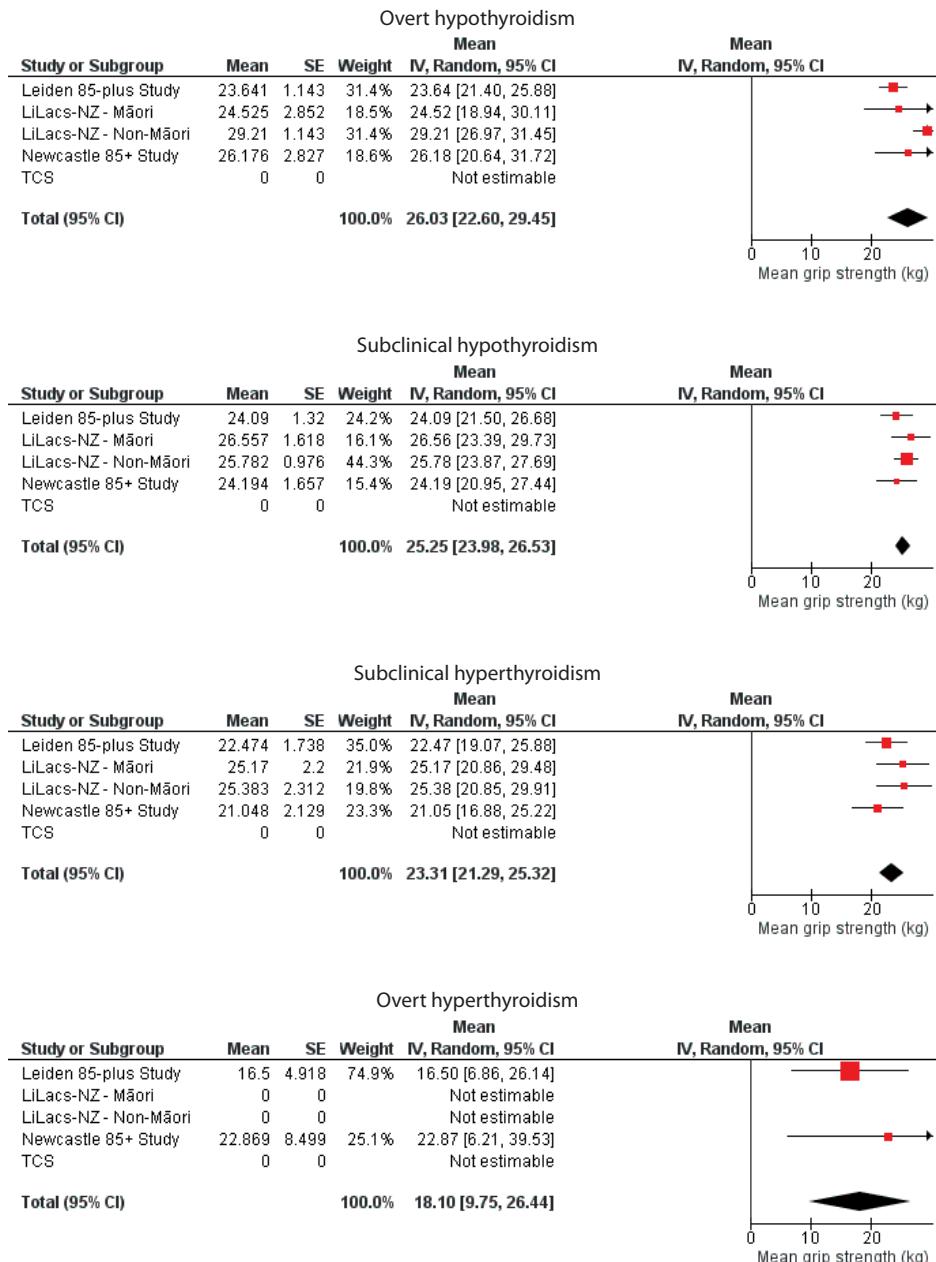
Supplementary data. Cross-sectional forest plots: GDS

Estimated Marginal Means (95% Confidence Intervals) estimated using Linear Mixed models per cohort, adjusted for baseline age and sex, and pooled using random-effects models with inverse-variance weighting.



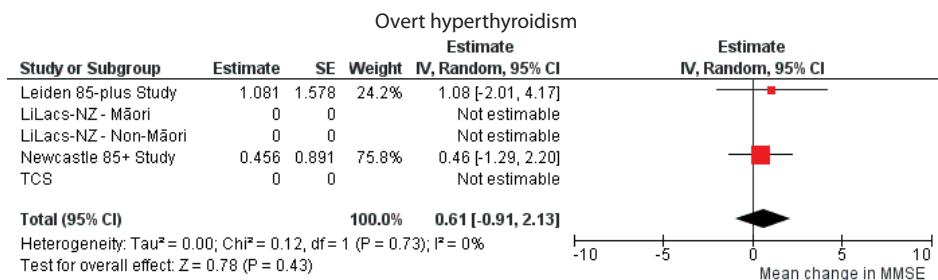
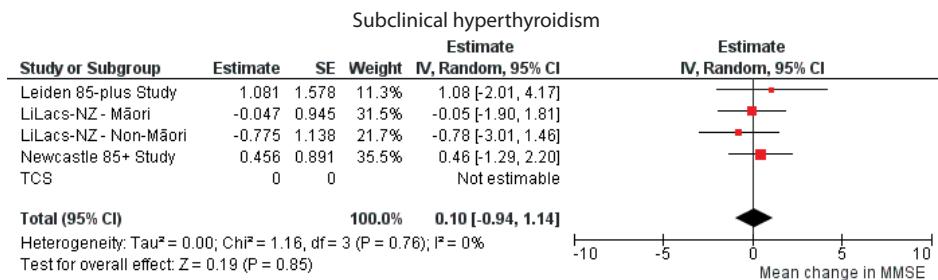
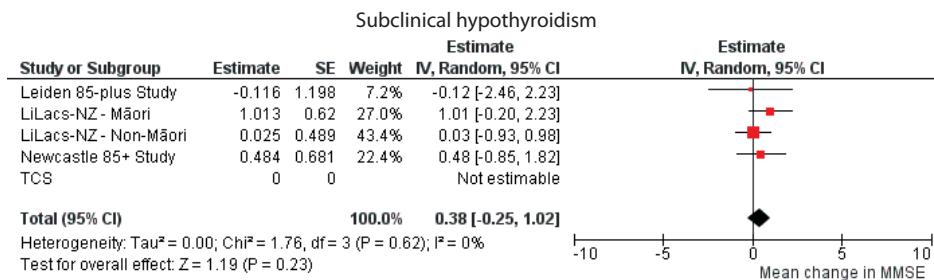
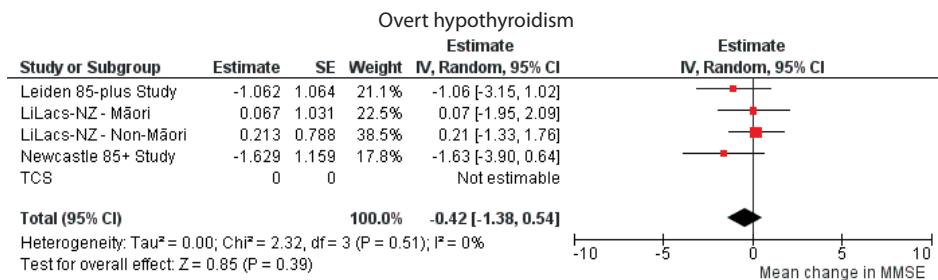
Supplementary data. Cross-sectional forest plots: Grip strength

Estimated Marginal Means (95% Confidence Intervals) estimated using Linear Mixed models per cohort, adjusted for baseline age and sex, and pooled using random-effects models with inverse-variance weighting.



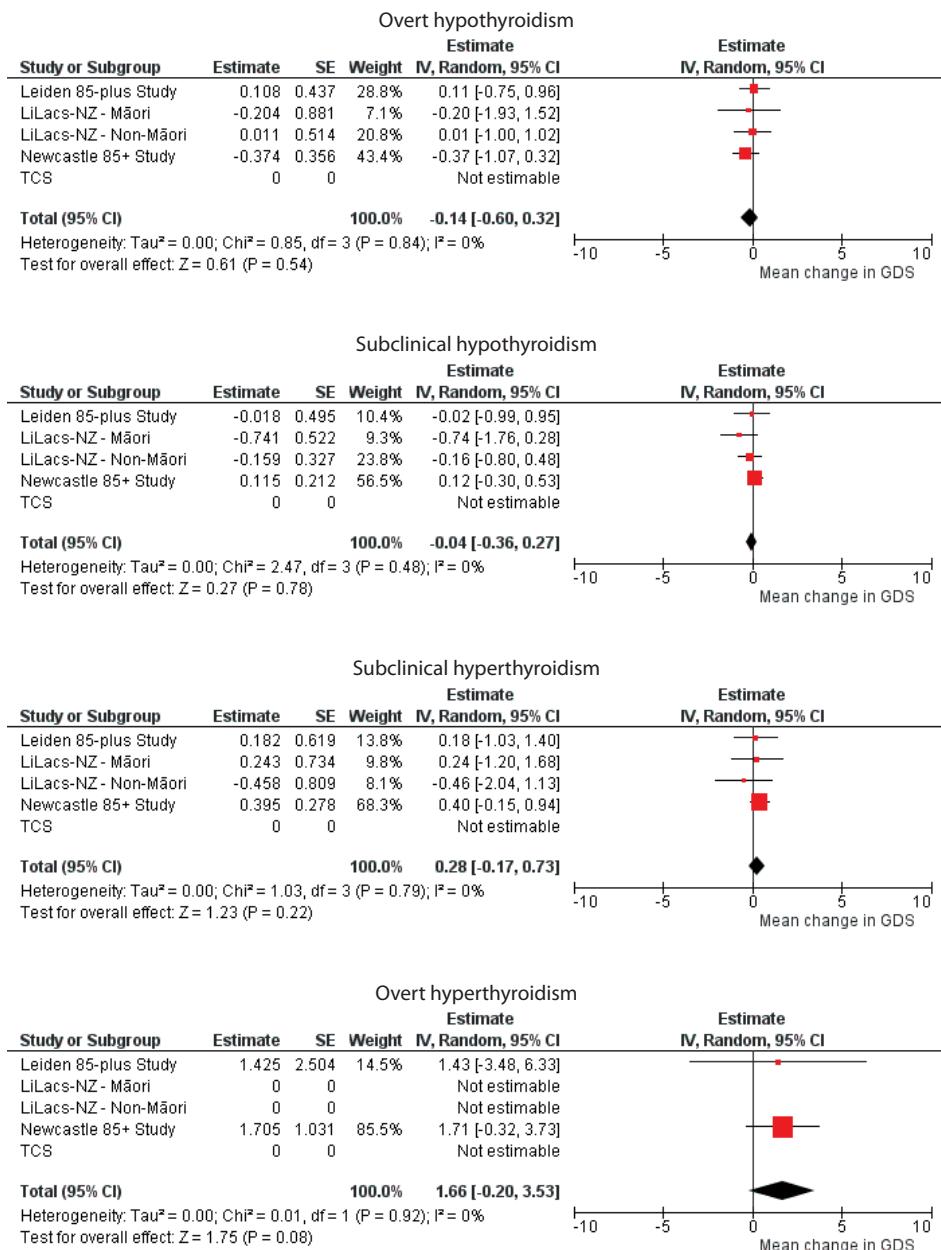
Supplementary data. Longitudinal forest plots: MMSE

Beta's (95% Confidence Intervals) estimated using repeated measures Linear Mixed models per cohort, adjusting for baseline age and sex, and pooled using random effects models with inverse-variance weighting, representing change in outcome measure over five years of follow-up. Euthyroidism is the reference group.



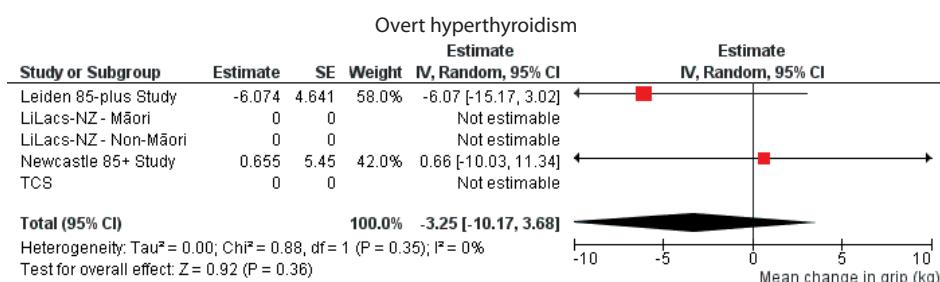
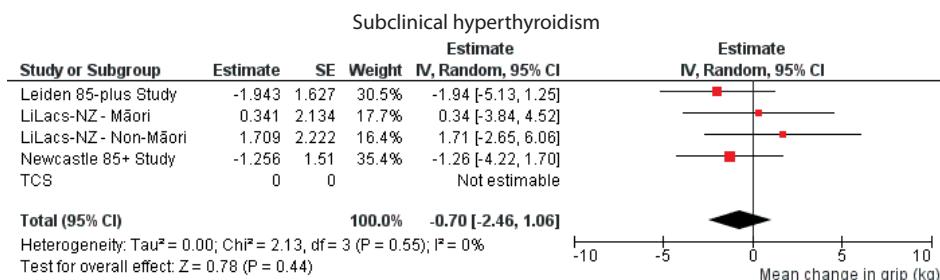
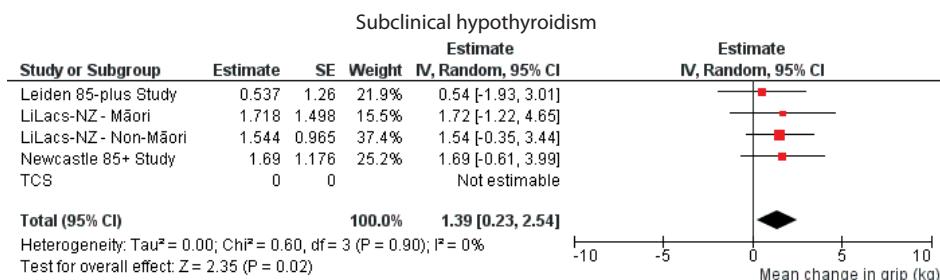
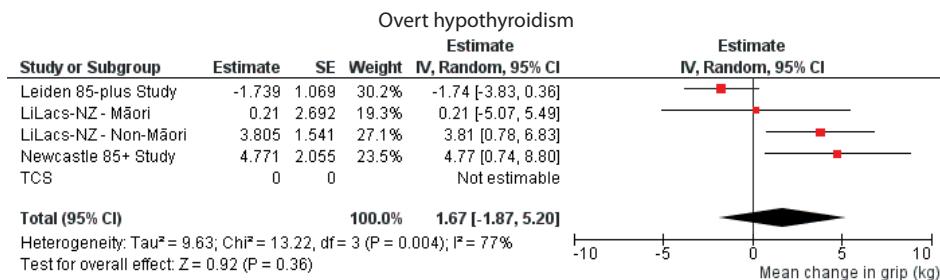
Supplementary data. Longitudinal forest plots: GDS

Beta's (95% Confidence Intervals) estimated using repeated measures Linear Mixed models per cohort, adjusting for baseline age and sex, and pooled using random effects models with inverse-variance weighting, representing change in outcome measure over five years of follow-up. Euthyroidism is the reference group.



Supplementary data. Longitudinal forest plots: Grip strength

Beta's (95% Confidence Intervals) estimated using repeated measures Linear Mixed models per cohort, adjusting for baseline age and sex, and pooled using random effects models with inverse-variance weighting, representing change in outcome measure over five years of follow-up. Euthyroidism is the reference group.



Supplementary method. SMD calculation

Standardised Mean Difference (SMD) and corresponding standard errors (SE) were calculated using an independent-groups pretest-posttest formula:

$$SMD = \frac{M_{D,E} - M_{D,C}}{SD_{pooled}}$$

with $SD_{pooled} = \sqrt{\frac{(N_E-1)SD_E^2 + (N_C-1)SD_C^2}{N_{tot}}}$

$$\text{and } SE = \sqrt{\frac{N_{tot}}{N_e * N_c} + \frac{SMD^2}{2(N_{tot} - 2)}}$$

SMD = standardised mean difference, $M_{D,E}$ = mean difference post - pre-test in thyroid dysfunction group, $M_{D,C}$ = mean difference post - pre-test in euthyroid group, N_E = participants in thyroid dysfunction category, SD_E = standard deviation in thyroid dysfunction group, N_C = participants in euthyroid category, SD_C = standard deviation in euthyroid category, N_{tot} = total number of participants.