scientific reports

OPEN

An individual participant data analysis of prospective cohort studies on the association between subclinical thyroid dysfunction and depressive symptoms

Lea Wildisen[1,22✉], Cinzia Del Giovane[1], Elisavet Moutzouri[1,2], Shanthi Beglinger[1,2], Lamprini Syrogiannouli[1], Tinh-Hai Collet[3], Anne R. Cappola[4], Bjørn O. Åsvold[5,6], Stephan J. L. Bakker[7], Bu B. Yeap[8], Osvaldo P. Almeida[8], Graziano Ceresini[9], Robin P. F. Dullaart[7], Luigi Ferrucci[10], Hans Grabe[11], J. Wouter Jukema[12], Matthias Nauck[13,14], Stella Trompet[15], Henry Völzke[16], Rudi Westendorp[17], Jacobijn Gussekloo[15,18], Stefan Klöppel[19], Drahomir Aujesky[2], Douglas Bauer[20], Robin Peeters[21], Martin Feller[1,2] & Nicolas Rodondi[1,2]

In subclinical hypothyroidism, the presence of depressive symptoms is often a reason for starting levothyroxine treatment. However, data are conflicting on the association between subclinical thyroid

[1]Institute of Primary Health Care (BIHAM), University of Bern, Mittelstrasse 43, 3012 Bern, Switzerland. [2]Department of General Internal Medicine, Inselspital, Bern University Hospital, University of Bern, Freiburgstrasse 15, 3010 Bern, Switzerland. [3]Service of Endocrinology, Diabetes and Metabolism, Department of Medicine, Lausanne University Hospital and University of Lausanne, Rue du Bugnon 46, 1011 Lausanne, Switzerland. [4]Division of Endocrinology, Diabetes, and Metabolism, Department of Medicine, University of Pennsylvania School of Medicine, 3400 Civic Center Boulevard, Philadelphia, PA 19104, USA. [5]K.G. Jebsen Center for Genetic Epidemiology, Department of Public Health and Nursing, NTNU, Norwegian University of Science and Technology, Postboks 8905 MTFS, 7491 Trondheim, Norway. [6]Department of Endocrinology, St. Olavs Hospital, Trondheim University Hospital, Postbox 3250 Torgarden, 7006 Trondheim, Norway. [7]Department of Internal Medicine, University Medical Center Groningen, University of Groningen, Antonius Deusinglaan 1, 9713 AV Groningen, The Netherlands. [8]Medical School, University of Western Australia Perth, The University of Western Australia (M582), 35 Stirling Highway, Crawley, WA 6009, Australia. [9]Department of Medicine and Surgery, Unit of Internal Medicine and Onco-Endocrinology, University Hospital of Parma, Via Gramsci, 14 - 43126 Parma, Italy. [10]Longitudinal Studies Section, Translational Gerontology Branch, National Institute on Aging, 251 Bayview Boulevard, Suite 100, Baltimore, MD 21224, USA. [11]Institute for Community Medicine, Clinical-Epidemiological Research, University Medicine Greifswald, Walter Rathenau Str. 48, 17475 Greifswald, Germany. [12]Department of Cardiology, Leiden University Medical Center, Postbus 9600, 2300 RC Leiden, The Netherlands. [13]Institute of Clinical Chemistry and Laboratory Medicine, University Medicine Greifswald, Ferdinand-Sauerbruch-Straße, 17475 Greifswald, Germany. [14]DZHK (German Centre for Cardiovascular Research), Partner Site Greifswald, University Medicine, Ferdinand-Sauerbruch-Straße, 17475 Greifswald, Germany. [15]Section Gerontology and Geriatrics, Department of Internal Medicine, Leiden University Medical Center, Albinusdreef 2, 2333 ZA Leiden, The Netherlands. [16]Department of Psychiatry and Psychotherapy, University Medicine Greifswald, Ellernholzstrasse 1-2, 17489 Greifswald, Germany. [17]Department of Public Health and Center for Healthy Aging, University of Copenhagen, Gothersgade 160, 1123 København K, Mærsk Tower, Copenhagen, Denmark. [18]Department of Public Health and Primary Care, Leiden University Medical Center, LUMC Education Building, Hippocratespad 21, 2333 ZD Leiden, the Netherlands. [19]University Hospital of Old Age Psychiatry, University of Bern, Murtenstrasse 21, 3008 Bern, Switzerland. [20]Departments of Medicine and Epidemiology and Biostatistics, University of California, San Francisco, 550 16th St., Box 0560, San Francisco, CA 94158, USA. [21]Department of Medicine, Erasmus Medical Center, Postbus 2040, 3000 CA Rotterdam, The Netherlands. [22]Graduate School for Health Sciences, University of Bern, Mittelstrasse 43, 3012 Bern, Switzerland. ✉email: lea.wildisen@biham.unibe.ch

dysfunction and depressive symptoms. We aimed to examine the association between subclinical thyroid dysfunction and depressive symptoms in all prospective cohorts with relevant data available. We performed a systematic review of the literature from Medline, Embase, Cumulative Index to Nursing and Allied Health Literature, and the Cochrane Library from inception to 10th May 2019. We included prospective cohorts with data on thyroid status at baseline and depressive symptoms during follow-up. The primary outcome was depressive symptoms measured at first available follow-up, expressed on the Beck's Depression Inventory (BDI) scale (range 0–63, higher values indicate more depressive symptoms, minimal clinically important difference: 5 points). We performed a two-stage individual participant data (IPD) analysis comparing participants with subclinical hypo- or hyperthyroidism versus euthyroidism, adjusting for depressive symptoms at baseline, age, sex, education, and income (PROSPERO CRD42018091627). Six cohorts met the inclusion criteria, with IPD on 23,038 participants. Their mean age was 60 years, 65% were female, 21,025 were euthyroid, 1342 had subclinical hypothyroidism and 671 subclinical hyperthyroidism. At first available follow-up [mean 8.2 (± 4.3) years], BDI scores did not differ between participants with subclinical hypothyroidism (mean difference = 0.29, 95% confidence interval = − 0.17 to 0.76, $I^2 = 15.6$) or subclinical hyperthyroidism (− 0.10, 95% confidence interval = − 0.67 to 0.48, $I^2 = 3.2$) compared to euthyroidism. This systematic review and IPD analysis of six prospective cohort studies found no clinically relevant association between subclinical thyroid dysfunction at baseline and depressive symptoms during follow-up. The results were robust in all sensitivity and subgroup analyses. Our results are in contrast with the traditional notion that subclinical thyroid dysfunction, and subclinical hypothyroidism in particular, is associated with depressive symptoms. Consequently, our results do not support the practice of prescribing levothyroxine in patients with subclinical hypothyroidism to reduce the risk of developing depressive symptoms.

Abbreviations

BDI	Beck depression inventory
CESD	Center for epidemiological studies depression
CI	Confidence interval
CINAHL	Cumulative index to nursing and allied health literature
DSM-5	Diagnostic and statistical manual of mental disorders, fifth edition
fT3	Free triiodothyronine
fT4	Free thyroxine
GDS	Geriatric depression scale
HADS	Hospital anxiety and depression scale
ICD	International statistical classification of diseases and related health problems
IPD	Individual participant data/individual patient data
MD	Mean difference
MMSE	Mini-mental state examination
N	Number
NOS	Newcastle–Ottawa scale
PRISMA	Preferred reporting items for systematic reviews and meta-analyses
PROSPERO	International prospective register of systematic reviews
SD	Standard deviation
SHyper	Subclinical hyperthyroidism
SHypo	Subclinical hypothyroidism
SMD	Standardized mean difference
TSH	Thyroid-stimulating hormone
y	Year

Subclinical thyroid dysfunction is common in the adult population and its prevalence increases with age, affecting up to 10–15% of older adults[1]. Patients are diagnosed with subclinical thyroid dysfunction when their serum thyroid-stimulating hormone (TSH) levels are below or above the reference range, but when their serum free thyroxine (fT4) levels are still within the reference range. Only a few symptoms are usually linked to subclinical thyroid dysfunction. Several guidelines discuss the association between depressive symptoms and thyroid dysfunction, and the potential benefit of levothyroxine for patients with the two diagnoses[2–5]. However, as the evidence is low, guidelines do not recommend to routinely treat patients with subclinical hypothyroidism and depressive symptoms with levothyroxine. Nevertheless, a study among GPs reported that the presence of depressive symptoms or low mood influence their decision whether or not to start treatment for subclinical hypothyroidism[6].

The association between subclinical thyroid dysfunction and depressive symptoms is unclear because studies to date have yielded conflicting results. Several studies showed that participants with subclinical hypothyroidism or subclinical hyperthyroidism had more severe depressive symptoms, but other studies reported no differences[7–15]. The largest prospective study published showed no association between subclinical hypothyroidism and incidence of depression after 2 years of follow-up[16], whereas depressive symptoms were associated with subclinical hyperthyroidism (but not subclinical hypothyroidism) in another prospective study[11].

These conflicting results may be explained by differences in outcome definition and in the statistical methods used to analyse data. Individual participant data (IPD) help researchers to standardise the analyses and definitions used across studies and make it possible to identify the effects for different subgroups in large study populations[17]. For example, IPD allow the use of uniform cut-off levels of TSH to define thyroid status for each study, the same model for the analysis in each study, and stratification by age and sex.

We thus aimed to assess the association between subclinical hypothyroidism or subclinical hyperthyroidism and future development of depressive symptoms by conducting an analysis of IPD from prospective cohort studies.

Methods

We registered this systematic review and IPD analysis in the international Prospective Register of Systematic Reviews PROSPERO (CRD42018091627) and published the study protocol[18]. This study adheres to the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) statement for IPD systematic reviews[19].

Search strategy and selection criteria.   We performed a systematic literature search in Ovid Medline, Ovid Embase, Cumulative Index to Nursing and Allied Health Literature (CINAHL), and in the Cochrane Library, from inception to 10th May 2019. We included publications from prospective studies that measured at least baseline TSH in adults and assessed depressive symptoms during follow-up on a validated continuous depression scale or diagnosis of depression (e.g. through ICD-10 or DSM-V codes). The following search items were used: thyroid diseases, hyperthyroidism, hypothyroidism, thyroid hormones, triiodothyronine, thyroxine, thyrotropin, subclinical, sub-clinical, mild, subnormal, pre-clinical, preclinical, depression. We did not include studies of only depressed patients, pregnant women, or women wanting to get pregnant. We included studies in any language and any publication year. We worked with two experienced librarians to develop the search strategy in Ovid Medline and then translated it to match subject headings and keywords for the other databases. See Appendix 1 for details of the Medline search strategy. Details of the systematic literature search, inclusion and exclusion criteria, and IPD analysis have been described in detail elsewhere[18]. We identified additional unpublished data by contacting the Thyroid Studies Collaboration, a consortium of cohort studies that investigate the association between subclinical thyroid dysfunction and clinical outcomes[20].

Data extraction and quality assessment.   We contacted investigators from all prospective cohort studies that met the inclusion criteria to collaborate in our IPD analysis by sharing their data. We requested IPD on thyroid status at baseline (TSH, fT4), and free triiodothyronine (fT3)), socioeconomic status (education, income), demographics (sex, age), medication use (levothyroxine, anti-thyroid, anti-depressant, thyroid-altering medication, including lithium, and amiodarone), and on depressive symptoms as measured on a validated continuous scale at baseline and at any follow-up. Each study was approved by its local ethics committee and all participants gave informed consent for the original studies. We used the Newcastle–Ottawa Scale (NOS) to assess the quality of the studies included[21]. The NOS contains eight items that focus on selection, comparability and outcome. The scale is scored from zero to nine stars; the highest score indicates the best methodological quality. We classed the studies as good, fair, and poor quality based on their star rating. We also assessed the certainty in the evidence with the GRADE tool (www.gradeworkinggroup.org). The certainty of evidence based on observational studies is «low», and may be decreased to «very low» for several reasons including study limitations (i.e. study quality), inconsistency, indirectness, imprecision, or increased for other considerations like particular design features of extremely rigorous well-conducted observational studies[22]. For assessing the study limitations, the final NOS score of the studies included in the main analysis was used. E.g. if all the included studies have a good NOS quality score the study limitations in the GRADE can be judged as "not serious" (Appendix 4).

Thyroid function testing.   Consistent with our previous IPD analyses, we used uniform TSH cut-off levels and study-specific fT4 cut-off values to define the thyroid status[20,23,24]. We defined euthyroidism as TSH levels between 0.45 and 4.50 mIU/L. Subclinical hyperthyroidism was defined as TSH levels < 0.45 mIU/L with normal fT4 values, and subclinical hypothyroidism as TSH levels > 4.50 and < 20 mIU/L and fT4 values within the reference range. For fT4, we used study specific cut-offs because these measurements show greater inter method variation than do sensitive 3rd generation TSH assays[20]. We excluded participants with fT4 values out of the reference range. When fT4 values were missing, we considered participants with TSH levels below 0.45 mIU/L to have subclinical rather than overt hyperthyroidism and participants with TSH levels above 0.45 mIU/L and below 20 mIU/L to have subclinical rather than overt hypothyroidism, because most adults with a TSH level in this range rather have subclinical than overt thyroid dysfunction[25,26]. We performed a sensitivity analysis excluding participants with missing fT4 to verify that the results remained robust, meaning that the effect size was not clinically different from the main result. We additionally performed a sensitivity analysis examining the difference in depressive symptoms between the subclinical hyperthyroidism and euthyroid participants, excluding participants with missing fT3 levels or values outside the reference range. We completed two sensitivity analyses excluding participants with thyroid medication (levothyroxine or anti-thyroid medication), and with thyroid altering medication (anti-thyroid drugs, levothyroxine, amiodarone, lithium) at baseline or follow-up.

Depressive symptoms.   Since studies used different continuous scales to measure depressive symptoms, we converted scores from different scales to the Beck Depression Inventory (BDI) scale, a commonly used depressive symptoms scale[27,28]. The BDI scale ranges from 0 to 63, with higher values indicating greater frequency and severity of depressive symptoms; the minimal clinically important difference is 5 points[28]. We calculated the conversion factor by dividing the range of the BDI scale by the range of the original scale. For example,

to convert the Center for Epidemiological Studies Depression (CESD) scale to the BDI, we used a conversion factor of 1.05 (63 (BDI range) ÷ 60 (CESD range))[28]. To transfer measurements from the CESD scale to the BDI scale, we then multiplied each individual's CESD value by the conversion factor. The primary outcome was depressive symptoms measured at first available follow-up, expressed on the BDI scale. As previously defined in the study protocol we converted measurements to the BDI scale instead of a standardised scale to facilitate the interpretation[18,28]. In a sensitivity analysis, we used the original scale to calculate the mean difference in each study and then we pooled the standardised mean differences (SMDs) across the studies. We coded SMDs so that positive values would indicate more severe depressive symptoms in participants with subclinical thyroid dysfunction than in euthyroid controls: < 0.40 was a small effect; 0.40–0.70 was a moderate effect, and > 0.70 was a large effect, respectively[29]. Since depressive symptoms could be influenced by medications, we conducted a sensitivity analysis excluding participants with antidepressant medication at baseline or follow-up. A secondary outcome was depressive symptoms at baseline, expressed on the BDI scale. An additional secondary outcome was depressive symptoms at the last available follow-up and at the third year of follow-up, expressed on the BDI scale. We chose a follow-up at year three because most of the cohorts had this follow-up time in common. Studies without a 3-year follow up were excluded from this analysis. Another secondary outcome was incidence of depressive symptoms at the first available follow-up. For the outcome of incidence of depression, we analysed data on diagnosis of depression or established cut-off points for presence of depression from the continuous depressive symptoms scales (cut-off points were defined as; (Geriatric Depression Scale 15-item (GDS-15): > 5 [30], CESD-20: > 21 [31], CESD-10: ≥ 8[32], Hospital Anxiety and Depression Scale (HADS): > 12[33]). In the analysis of incidence of depression, we excluded participants with diagnosed depression or with depressive symptoms score above the cut-off at baseline. In the primary outcome analyses, we excluded participants with missing data on depressive symptoms at baseline and follow-up. In a sensitivity analysis, we did not exclude those participants, but used multiple imputation for missing data on depressive symptoms. We additionally conducted a sensitivity analysis excluding participants with dementia (Mini-Mental State Examination (MMSE) score < 24, or diagnosis of dementia), since the relationship between dementia and depression is complex[34].

Analysis. We performed a two-stage IPD analysis. In the first stage, we estimated the effect size for each study separately. For the primary outcome, we calculated the mean difference in BDI score between participants with subclinical hypothyroidism or hyperthyroidism and euthyroid controls using a multivariable linear regression model adjusted for age, sex, depressive symptoms at baseline, education, and income. We only included studies with data on depressive symptoms at baseline because adjusting for depressive symptoms at baseline adjusts for imbalance and accounts for the correlation between baseline and follow-up, which makes the effect estimates more precise[35]. In a sensitivity analysis, we additionally included studies without baseline data on depressive symptoms. When comparing incidences of depression, we calculated odds ratio for incidence of depression at the first available follow-up between subclinical hypothyroidism or hyperthyroidism and euthyroid controls using a multivariable logistic regression model adjusted for age, sex, depressive symptoms at baseline, income, and education. For the cross-sectional analysis at baseline, we calculated mean difference in BDI score between participants with subclinical hypo- or hyperthyroidism and euthyroid control, adjusted for age, sex, education, and income. In the second stage of the IPD analysis, we pooled derived mean differences or odds ratios from all the different studies from the first stage using a random effects model.

To identify sub-populations at risk and possible sources of heterogeneity, we conducted predefined subgroup and sensitivity analyses on the primary outcome. We performed predefined subgroup analyses by age (younger and older than 75 years old), by sex, by TSH levels (4.51–6.99 mIU/L, 7.00–9.99 mIU/L, 10.0–20.0 mIU/L)[20], and by levothyroxine use at baseline.

We performed a sensitivity analysis excluding the HUNT study as this was the biggest study included in the analysis and therefore had the biggest weight for the overall result.

Heterogeneity was estimated with I^2 and the Q test. Statistical significance testing was 2-sided and $P < 0.05$ was considered statistically significant. All analyses were conducted with Stata, release 15 (StataCorp).

Ethics approval and consent to participate. *Statement on ethical approval from ethics committee* This is a manuscript that analysed existing cohort data. Each study in the individual participant data set received local ethical approval. Our analysis did not include identifiable data.

Statement on guidelines followed The manuscript adheres to the preferred reporting items for systematic reviews and meta-analyses (PRISMA) statement for IPD systematic reviews and to the PRISMA statement for systematic review protocols.

Statement on written informed consent from the participants This manuscript analysed existing cohort data, each study in the individual participant data set received informed consent from the participants. Our analysis did not include identifiable data.

Results

Of the 1047 studies we identified through the literature search, four studies met our inclusion criteria (Appendix 2)[11,13,16,36]. From within the Thyroid Studies Collaboration we identified seven additional studies where data on subclinical thyroid dysfunction and depressive symptoms had not been published. We invited the investigators of eligible studies to collaborate in this IPD analysis; only one investigator, from a study identified by the literature review, declined to participate[16]. This study only presented a dichotomised result in the publication, so we could not combine this aggregate data with our main IPD analysis. We received IPD from ten studies with 33,769 participants that met our inclusion criteria. For the primary outcome analysis, we included only studies with a continuous outcome, and in which depressive symptoms had been measured at baseline. Six studies met

Study, Place	No. of Participants	Age, mean (range), y	Women, No. (%)	Original depression scale, (range)	Dementia, No. (%)	Thyroid medication, No. (%)	Depression medication, no. (%)	Median TSH, mIU/l	Normal range FT4, pmol/l	First/last available follow-up (mean), y
Studies included in the main analyses										
Leiden 85 + Study, The Netherlands	281	85 (85–85)	185 (66)	GDS-15 (0–15)	37 (13)	11 (4)	8 (3)	1.57	13–23	1.0/3.7
PROSPER Study, The Netherlands	405	75 (70–83)	190 (47)	GDS-15 (0–15)	0 (0)	10 (2)	12 (3)	2.04	12–18	3.0/3.0
InChianti Study, Italy	952	66 (21–98)	515 (54)	CESD-20 (0–60)	N.A	24 (2)	29 (3)	1.34	10–27	3.3/3.6
Health ABC Study, United States	2250	75 (69–81)	1139 (51)	CESD-20 (0–60)	481 (21)	212 (9)	7 (< 1)	2.15	10–23	2.9/7.8
CHS, United States	3419	75 (64–98)	2013 (59)	CESD-10 (0–30)	451 (13)	280 (8)	57 (2)	2.14	9–22	1.1/8.2
HUNT, Norway	15,731	53 (19–86)	10,963 (70)	HADS (0–21)	N.A	684 (4)	N.A	1.50	8–20	11.2/11.2
Overall	23,038	60 (19–98)	15,005 (65)	N.A	969 (4)	1221 (5)	113 (< 1)	1.63	N.A	8.2/9.9
Studies only included in sensitivity analyses or secondary outcomes										
MrOs*, United States	1356	73 (65–92)	0 (0)	GDS-15 (0–15)	29 (2)	93 (7)	71 (5)	2.04	9–24	3.9/**
PREVEND*, The Netherlands	2124	48 (28–75)	1086 (51)	N.A.‡	N.A	36 (2)	N.A	1.37	12–22	4.2/**
SHIP*, Germany	2139	46 (20–80)	1133 (53)	BDI (0–63)	46 (12)	105 (5)	31 (1)	0.68§	8–19	9.5/**
HIMS†, Australia	4032	77 (71–89)	0 (0)	N.A	212 (5)	126 (3)	261 (7)	2.00	10–23	N.A¶

Table 1. Study characteristics and baseline data. *Leiden 85 + Study*[13] Leiden 85 plus Study; *PROSPER*[11] prospective study of Pravastatin in the elderly at risk; *Health ABC Study*[46] The health, aging and body composition study; *CHS*[32] cardiovascular health study; *InChianti Study*[47] Invecchiare in Chianti Study; *HUNT*[48] Nord-Trøndelag Health Study; *PREVEND*[49] prevention of renal and vascular end-stage disease; *MrOS*[50] osteoporotic fractures in men study; *SHIP*[51] study of health in pomerania, *HIMS*[36] health in men study; *y* year; *GDS-15* geriatric depression scale 15-item; *CESD-10/-20* center for epidemiologic studies depression 10/20-item scale; *HADS* hospital anxiety and depression scale; *BDI* beck depression inventory scale; *TSH* thyroid stimulating hormone; *FT4* free thyroxine; *N.A.,* not available; *No.* number. *Not included in the main analysis because no data on depressive symptoms at baseline available. †Not included in the main analysis because no continuous scale used to measure depressive symptoms during follow up. Incidence of depression available (via Data Linkage). ‡No validated depression scale used. §SHIP includes participants from Pomerania, where an iodine supplementation program began in the mid-1990s. This shifted the distribution of TSH values towards the left in its first years, which lowered TSH values in the population of the SHIP Study during baseline examinations in 1997–2001. ¶follow-up via data-linkage from baseline 2001–2004, censor date 2013. **In the sensitivity analyses and subgroup analyses we only assessed the effect estimate at the first available follow-up.
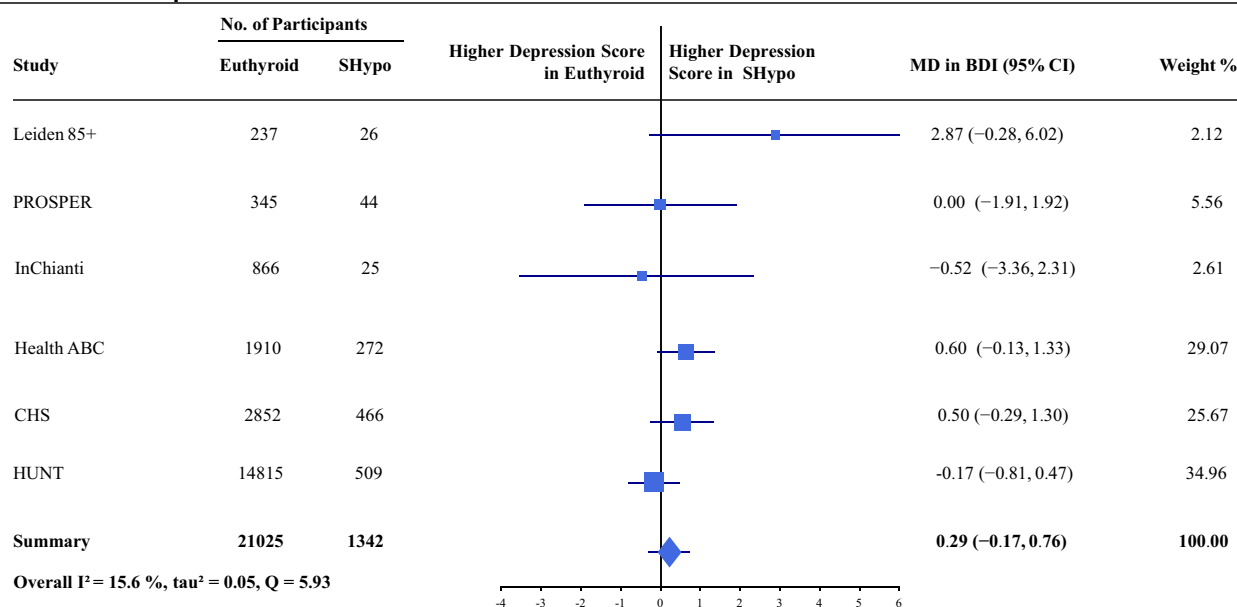
these criteria, including 23,038 participants. Mean age (± SD) was 60 years (± 13), 65% were female, and median TSH was 1.63 mIU/L (Table 1). 21,025 (91%) participants were euthyroid, 1342 (6%) participants had subclinical hypothyroidism, and, 671 (3%) had subclinical hyperthyroidism. In sensitivity analyses, we additionally included the three studies that did not measure depressive symptoms at baseline[37–39]. For the secondary outcome of incidence of depression, we additionally included the Health in Men Study, which did not use a continuous scale to measure depressive symptoms but assessed incidence of depression via data linkage[36]. Depressive symptoms scores at baseline were balanced between groups in each cohort and on average the correlation between depressive symptoms at baseline and follow-up was 0.6 across studies (Appendix 6).

Quality assessment. Based on the NOS, the quality of all studies that we included in the primary outcome analysis was good (Appendix 3). Certainty in the evidence assessed with the GRADE tool for the primary outcome was low (Appendix 4). Because of the low number of studies, we did not assess publication bias[40].

Subclinical hypothyroidism. At first available follow-up (mean 8.2 (± 4.3) years), there was no difference in the primary outcome of BDI score between subclinical hypothyroidism and euthyroid controls (pooled mean difference (MD) 0.29, 95% CI − 0.17 to 0.76) with a low heterogeneity ($I^2 = 15.6\%$) (Fig. 1).

Secondary outcomes are shown in Fig. 2. At baseline there was neither relevant difference in depressive symptoms between euthyroid participants (mean BDI = 10.28) and participants with subclinical hypothyroidism (mean BDI = 9.63), nor in multivariable analysis adjusted for age, sex, education, and income (MD in BDI

a) Difference in BDI score between participants with subclinical hypothyroidism and euthyroid participants after the first available follow up*
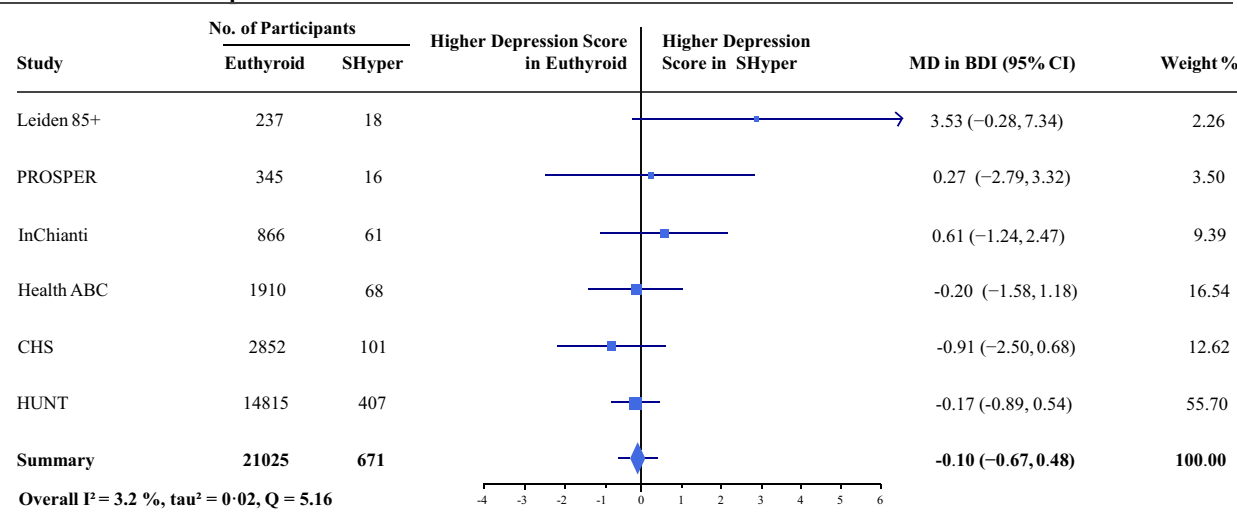
Study	No. of Participants		Higher Depression Score in Euthyroid	Higher Depression Score in SHypo	MD in BDI (95% CI)	Weight %
	Euthyroid	SHypo				
Leiden 85+	237	26			2.87 (−0.28, 6.02)	2.12
PROSPER	345	44			0.00 (−1.91, 1.92)	5.56
InChianti	866	25			−0.52 (−3.36, 2.31)	2.61
Health ABC	1910	272			0.60 (−0.13, 1.33)	29.07
CHS	2852	466			0.50 (−0.29, 1.30)	25.67
HUNT	14815	509			-0.17 (−0.81, 0.47)	34.96
Summary	**21025**	**1342**			**0.29 (−0.17, 0.76)**	**100.00**

Overall I² = 15.6 %, tau² = 0.05, Q = 5.93

b) Difference in BDI score between participants with subclinical hyperthyroidism and euthyroid participants after the first available follow up*

Study	No. of Participants		Higher Depression Score in Euthyroid	Higher Depression Score in SHyper	MD in BDI (95% CI)	Weight %
	Euthyroid	SHyper				
Leiden 85+	237	18			3.53 (−0.28, 7.34)	2.26
PROSPER	345	16			0.27 (−2.79, 3.32)	3.50
InChianti	866	61			0.61 (−1.24, 2.47)	9.39
Health ABC	1910	68			-0.20 (−1.58, 1.18)	16.54
CHS	2852	101			-0.91 (−2.50, 0.68)	12.62
HUNT	14815	407			-0.17 (−0.89, 0.54)	55.70
Summary	**21025**	**671**			**-0.10 (−0.67, 0.48)**	**100.00**

Overall I² = 3.2 %, tau² = 0·02, Q = 5.16

Figure 1. (**a**) Difference in BDI score between participants with subclinical hypothyroidism and euthyroid participants after the first available follow up*. (**b**) Difference in BDI score between participants with subclinical hyperthyroidism and euthyroid participants after the first available follow up*. * Analysis adjusted for depressive symptoms at baseline, sex, age, and education (The CHS, Health ABC Study, and the InChianti Study were additionally adjusted for income). Overall mean BDI score at first follow-up was 10.67 with a standard deviation of 8.99. Abbreviations: *SHypo* Subclinical Hypothyroidism; *SHyper* Subclinical Hyperthyroidism; *Leiden 85+* (9) Leiden 85 plus Study; *PROSPER (7)* Prospective Study of Pravastatin in the Elderly at risk; *Health ABC Study (40)* The Health, Ageing and Body Composition Study; *CHS (26)* Cardiovascular Health Study; *InChianti Study (41)* Invecchiare in Chianti Study; *HUNT (42)* Nord-Trøndelag Health Study; *BDI* Beck Depression Inventory Score (range 0–63, minimal clinically important difference 5 points); *MD* Mean Difference; *CI* Confidence Interval, *No.* Number.

scores − 0.05 95% CI = − 0.50–0.39 I² = 0.0%). In the five cohorts (N = 6393) with data on depressive symptoms at 3 years follow-up, BDI scores did not differ between participants with subclinical hypothyroidism and euthyroid controls (MD 0.36, 95% CI − 0.18–0.90, I² = 0.0%). At last available follow-up (9.9 ± 3.3 years) there was no association in BDI score between subclinical hypothyroidism and euthyroid control (MD 0.05, 95% CI − 0.38–0.48,

Secondary outcomes - Association between subclinical hypothyroidism and depressive symptoms*
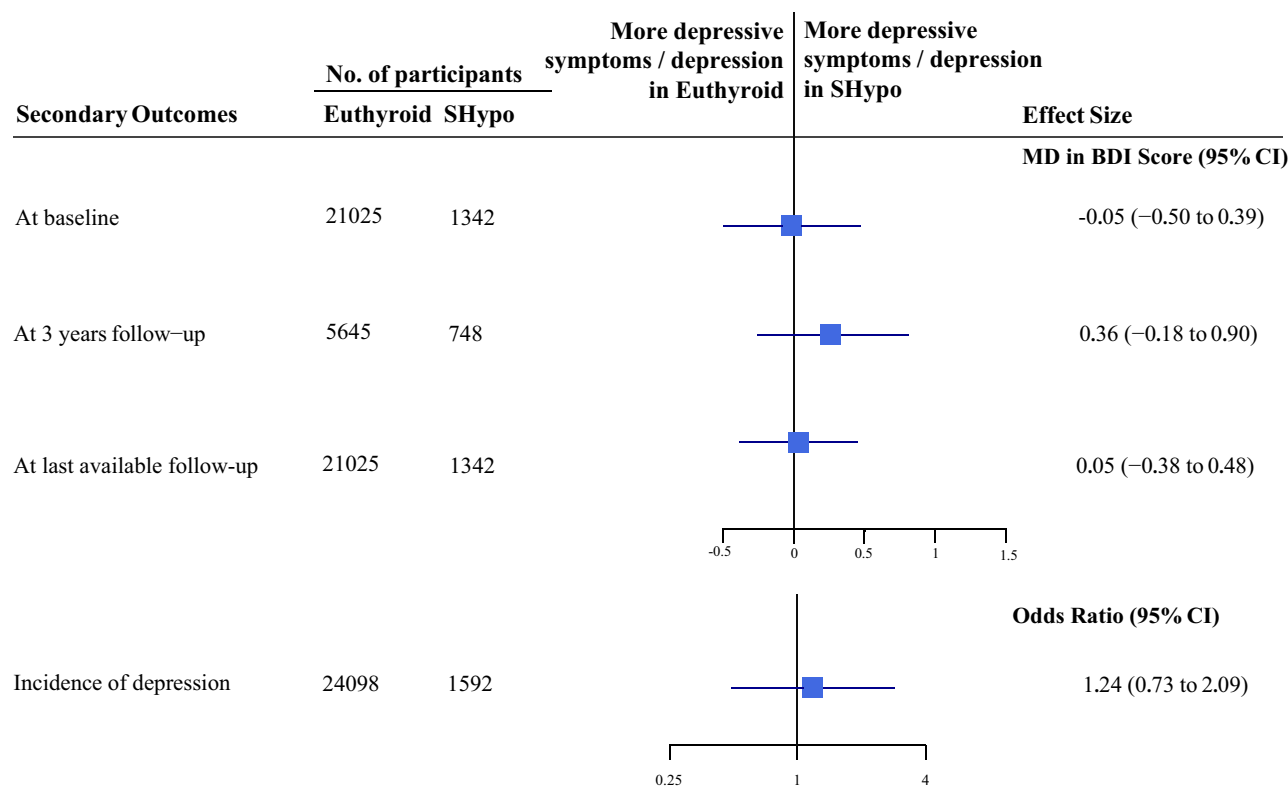

Secondary Outcomes	No. of participants		More depressive symptoms / depression in Euthyroid	More depressive symptoms / depression in SHypo	
	Euthyroid	SHypo			Effect Size
					MD in BDI Score (95% CI)
At baseline	21025	1342			-0.05 (−0.50 to 0.39)
At 3 years follow−up	5645	748			0.36 (−0.18 to 0.90)
At last available follow-up	21025	1342			0.05 (−0.38 to 0.48)
					Odds Ratio (95% CI)
Incidence of depression	24098	1592			1.24 (0.73 to 2.09)

Figure 2. Secondary outcomes—association between subclinical hypothyroidism and depressive symptoms*. * Analysis adjusted for depressive symptoms at baseline, sex, age, and education (The CHS, Health ABC Study, and the InChianti Study were additionally adjusted for income). † Analysis includes the same studies as for the primary outcome analysis: Leiden 85+ (9), PROSPER (7), Health ABC Study (40), CHS (26), InChianti Study (41), HUNT (42). ‡ Analysis includes the same studies as in the primary outcome analysis except of HUNT (42). § Analysis includes the same studies as in the primary outcome analysis plus the HIMS (30) which only had data on incidence of depression and no continuous measurement. Abbreviations: *SHypo* Subclinical Hypothyroidism; *MD* Mean Difference; *BDI Score* Beck Depression Inventory Score (range 0–63, minimal clinically important difference 5 points); *CI* Confidence Interval; *No.* Number.

$I^2 = 0.0\%$). The odds ratio for incident depression comparing subclinical hypothyroid and euthyroid participants was 1.24 (95% CI 0.73–2.09, $I^2 = 56.4\%$).

Results of sensitivity analyses in which we excluded participants taking thyroid medication (N = 21,391), thyroid altering medication (N = 21,384) or antidepressant medication (N = 23,175), participants with dementia (N = 22,224), or participants without fT4 measurements at baseline (N = 5618), or in which we excluded the study with the biggest weight (HUNT) (N = 7043), were similar comparable to those from our primary analysis (Table 2). When we used multiple imputation, depressive symptoms between subclinical hypothyroidism and euthyroidism did also not differ in a sensitivity analysis that included participants with t4 data on outcome and depressive symptoms at baseline (N = 45,398). Results of a sensitivity analysis in which we included studies without baseline information on depressive symptoms (N = 27,361) were similar to those of the main analysis. In a sensitivity analysis that used the original scale from each study, we found SMD between groups of 0.04 (95% CI = −0.02–0.09). Figure 3 shows the results of several subgroup analyses. After stratifying the population that was included in the primary outcome analysis by age (participants older and younger than 75), by sex, by levothyroxine treatment at baseline, and by different TSH levels, we found no significant differences in depressive symptoms scores.

Subclinical Hyperthyroidism. There was no difference in the primary outcome of depressive symptoms between subclinical hyperthyroidism and euthyroid controls (MD 0.10, 95% CI − 0.67–0.48, $I^2 = 3.2\%$) at first available follow-up (Fig. 1), as well as at year three follow-up, and at last available follow-up (Appendix 5a). At baseline, there was no difference in depressive symptoms expressed on the BDI scale between euthyroid participants (Mean BDI = 10.26) and participants with subclinical hyperthyroidism (Mean BDI = 10.28) (Appendix 5). Odds for incidence of depression were not higher for participants with subclinical hyperthyroidism than

Sensitivity analysis	No. of participants		No. of included studies†	MD in BDI score (95% CI), I²
	Euthyroid	SHypo		
Main analysis	21,025	1342	6	0.29 (−0.17–0.76) §, 15.6%
1) Excluding participants with thyroid medication	20,268	1123	6	0.30 (−0.15–0.75), 4.7%
2) Excluding participants with thyroid altering medication	20,261	1123	6	0.30 (−0.15–0.76), 4.8%
3) Excluding participants with antide-pressant medication	20,897	1327	6	0.32 (−0.19–0.83), 26.1%
4) Excluding participants with dementia	20,203	1222	6	0.24 (−0.17–0.65), 0.0%
5) Excluding participants without FT4 measurement	4550	1068	5	0.17 (−0.49–0.83), 11.8%
6) Using multiple imputation to impute missing outcome data	42,759	2639	6	0.35 (−0.10–0.81), 10.7%
7) Including studies without depressive symptoms at baseline	25,851	1510	9	0.25 (−0.13–0.63), 0.0%
8) Not adjusted for income	21,509	1413	6	0.22 (−0.22, 0.65), 11.1%
9) Excluding HUNT	6210	833	5	0.54 (0.04, 1.05), 0.0%
			SMD in depressive symptoms (95% CI), I²	
10) Using original scale for depressive symptoms‡	21,025	1342	6	0.04 (−0.02–0.09), 27.2%

Table 2. Sensitivity analyses on subclinical hypothyroidism and depressive symptoms at first available follow-up* *SHypo* subclinical hypothyroidism; *TSH* thyroid-stimulating hormone; *FT4* free thyroxine; *MD* mean difference; *SMD* standardised mean difference; *BDI Score* Beck DEPRESSION INVENTORY SCORE (range 0–63, minimal clinically important difference 5 points); *CI* confidence Interval; No. number. *Analyses adjusted for depressive symptoms at baseline (In sensitivity analysis 7: except studies without measurement), sex, age, income, and education (The CHS, Health ABC Study, and the InChianti Study were additionally adjusted for income). †The main analysis includes the same studies as for the primary outcome analysis: Leiden 85+[13], PROSPER[11], Health ABC Study[46], CHS[32], InChianti Study[47], HUNT[48]. **Sensitivity analyses 1–4, 6, 8, 10**: the same studies as in the main analysis were included, only participants with a certain measurement missing were excluded. **Sensitivity analysis 5:** the same studies as in the main analysis without the Health ABC Study [32] because in this study FT4 was not measured in the euthyroid group. **Sensitivity analysis 7:** the same studies as in the main analysis plus 3 studies that did not have data for depressive symptoms at baseline were included (PREVEND (Prevention of Renal and Vascular end-stage Disease)[49], MrOS (Osteoporotic Fractures in Men)[50], SHIP (Study of Health in Pomerania)[51]). **Sensitivity analysis 9:** the same studies as in the main analysis without the HUNT[48] study as the HUNT study has the biggest weight in the summarized result of the main outcome (34.96%). ‡Mean differences using the original scale for depressive symptoms within each study were pooled. §Overall mean BDI score at first follow-up of all 23,038 participants was 10.67 with a standard deviation of 8.97.

for euthyroid controls (Appendix 5a). Our results remained robust in several sensitivity and subgroup analyses (Appendix 1b,c).

Discussion

In this IPD analysis of 23,038 participants from six prospective cohort studies, we found no clinically relevant differences in depressive symptoms during follow-up between subclinical hypothyroidism or hyperthyroidism and euthyroid controls. Depressive symptoms of participants with subclinical hypothyroidism or hyperthyroidism were not different from symptoms of euthyroid control participants at baseline or at any follow-up. Participants with subclinical hypothyroidism were not at increased risk for incidence of depression. Our results were robust across all sensitivity analyses.

To our knowledge, no pooled IPD analysis has previously assessed the association of subclinical thyroid dysfunction and depressive symptoms. The results are in contrast with two previous meta-analyses of cross-sectional studies[7, 8] which found a positive association between subclinical hypothyroidism and depression. Reasons for difference in results could be that we included published and unpublished studies, that we did not include studies only on depressed patients, that we did not include case–control studies and cross-sectional studies (only cross-sectional analysis of prospective studies in our analysis, as they are considered of higher validity), and that we analyzed individual participant data, which leads to far more reliable results[17,35]. With individual participant data, we could use standardized definition of subclinical hypo- and hyperthyroidism for all studies, which was not possible in these study-level meta-analyses. Our results are in line with the results found by the Kangbuk Samsung Health Study showing that participants with subclinical hypothyroidism had no higher incidence of depression than euthyroid controls[16].

Our study has some limitations. First, younger people were underrepresented because three of six studies included participants only over 64. However, our sensitivity analysis that excluded participants over 75 also yielded no association, but we were able to include too few participants to assess risk among middle-aged adults. Second, we were limited by measurement of depressive symptoms using different scales across studies, so that we

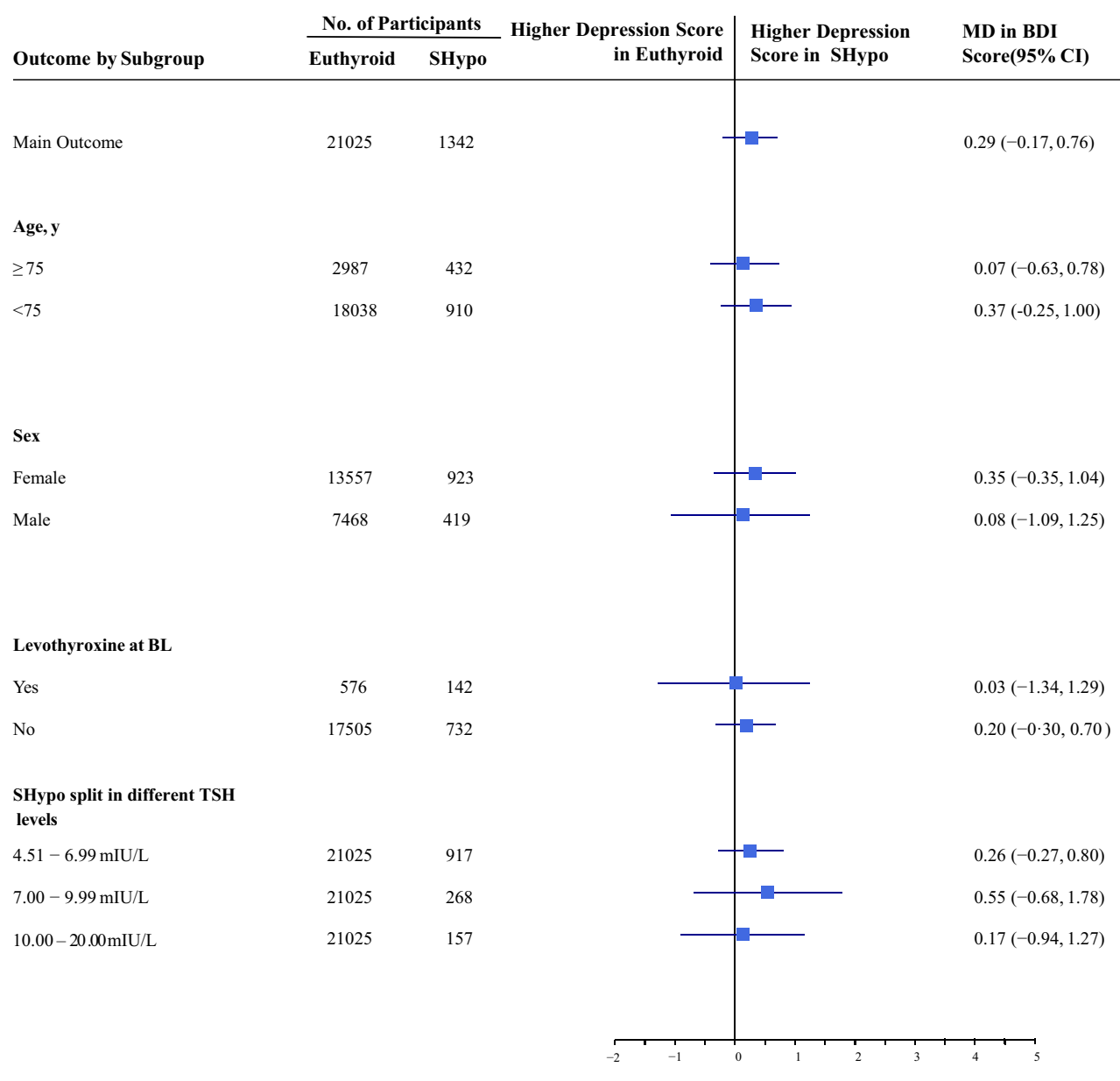**The association between subclinical hypothyroidism and depressive symptoms by subgroups***

Outcome by Subgroup	No. of Participants		Higher Depression Score in Euthyroid	Higher Depression Score in SHypo	MD in BDI Score(95% CI)
	Euthyroid	SHypo			
Main Outcome	21025	1342			0.29 (−0.17, 0.76)
Age, y					
≥75	2987	432			0.07 (−0.63, 0.78)
<75	18038	910			0.37 (-0.25, 1.00)
Sex					
Female	13557	923			0.35 (−0.35, 1.04)
Male	7468	419			0.08 (−1.09, 1.25)
Levothyroxine at BL					
Yes	576	142			0.03 (−1.34, 1.29)
No	17505	732			0.20 (−0·30, 0.70)
SHypo split in different TSH levels					
4.51 − 6.99 mIU/L	21025	917			0.26 (−0.27, 0.80)
7.00 − 9.99 mIU/L	21025	268			0.55 (−0.68, 1.78)
10.00 − 20.00 mIU/L	21025	157			0.17 (−0.94, 1.27)

Figure 3. The association between subclinical hypothyroidism and depressive symptoms by subgroups*. * Analysis adjusted for depressive symptoms at baseline, sex, age, and education (The CHS, Health ABC Study, and the InChianti Study were additionally adjusted for income). Abbreviations: *SHypo* Subclinical Hypothyroidism; *TSH* Thyroid-Stimulating Hormone; *MD* Mean Difference; *BDI Score* Beck Depression Inventory Score (range 0–63, minimal clinically important difference 5 points); *CI* Confidence Interval; *No.* Number.

were unable to combine the effect estimates from different studies using their original scales. To standardise the scale between studies, we converted scores from the various scales to the BDI scale, a common scale whose scores are easy to interpret[28]. As there was no validated conversion factor, we examined whether our results remained robust when we converted the original scores to a standardised scale, which yielded similar findings. Third, we did not have access to information about treatment prior to baseline; patients with subclinical thyroid dysfunction and depressive symptoms may have been more frequently diagnosed with subclinical thyroid dysfunction and been treated to restore euthyroidism prior to baseline, in which case the subclinical thyroid dysfunction group at baseline would overrepresent the number of people who did not develop depressive symptoms. Fourth, the diagnosis of subclinical hypothyroidism was based on one assessment of TSH, and did not depend on a second

verified measurement, which is a limitation of most published large cohorts that have examined the risk of sub-clinical thyroid dysfunction[20,41]. Based on a single elevated or decreased TSH level, participants might revert to normal thyroid function over follow-up, which could have biased the results towards the null. However, previous IPD analyses including cohorts with just one TSH assessment documented an association between subclinical thyroid dysfunction and both coronary heart disease or fractures[20,24]. Inferring a causal relationship from observational data is challenging and, for this reason, we completed a series of sensitivity analysis to minimise the potential effects of residual confounding and bias.

Our study was strengthened by its IPD analysis, considered the most appropriate method in evidence synthesis since it offers many advantages over aggregate data analysis[42]. For example, our results do not suffer from the ecological bias of study-level meta-analyses. We could also standardise definitions of predictors and outcomes, use uniform adjustments for potential confounders to reduce heterogeneity across studies, and include unpublished data to increase the robustness of our results and our power to detect associations. Because our IPD analysis was large, we could assess the effects of age, sex, thyroid medication, antidepressant medication, and TSH levels in subgroup analyses.

Current guidelines for the management of people with subclinical hypothyroidism tend to recommend thyroid hormone substitution for adults with TSH levels > 10 mIU/L and for people with lower TSH values who are young or symptomatic, although some recent guidelines have more narrow indications[43,44]. As we found no association between subclinical hypothyroidism and depressive symptoms, one might infer that thyroid hormones would be of limited benefit for the treatment of depressive symptoms in affected people with subclinical hypothyroidism. This is in line with a previous meta-analysis of four small randomised clinical trials (total N = 278) that found no positive association between thyroid hormone therapy and depressive symptoms[45]. Overall, our results do not support increased risk of depressive symptoms in adults with subclinical thyroid dysfunction.

Data availability
The datasets analysed during this study are available from the corresponding author on reasonable request.

References

1. Biondi, B. & Cooper, D. S. The clinical significance of subclinical thyroid dysfunction. *Endocr. Rev.* **29**, 76–131. https://doi.org/10.1210/er.2006-0043 (2008).
2. Baskin, H. J. *et al.* American Association of Clinical Endocrinologists medical guidelines for clinical practice for the evaluation and treatment of hyperthyroidism and hypothyroidism. *Endocr. Pract.* **8**, 457–469 (2002).
3. Garber, J. R. *et al.* Clinical practice guidelines for hypothyroidism in adults: cosponsored by the American Association of Clinical Endocrinologists and the American Thyroid Association. *Endocr. Pract.* **18**, 988–1028. https://doi.org/10.4158/EP12280.GL (2012).
4. Jonklaas, J. *et al.* Guidelines for the treatment of hypothyroidism: prepared by the american thyroid association task force on thyroid hormone replacement. *Thyroid* **24**, 1670–1751. https://doi.org/10.1089/thy.2014.0028 (2014).
5. Okosieme, O. *et al.* Management of primary hypothyroidism: statement by the British Thyroid Association Executive Committee. *Clin. Endocrinol. (Oxf.)* **84**, 799–808. https://doi.org/10.1111/cen.12824 (2016).
6. Allport, J., McCahon, D., Hobbs, F. D. & Roberts, L. M. Why are GPs treating subclinical hypothyroidism? Case note review and GP survey. *Prim. Health Care Res. Dev.* **14**, 175–184. https://doi.org/10.1017/S1463423612000230 (2013).
7. Tang, R. *et al.* Subclinical hypothyroidism and depression: a systematic review and meta-analysis. *Front. Endocrinol. (Lausanne)* **10**, 340. https://doi.org/10.3389/fendo.2019.00340 (2019).
8. Loh, H. H., Lim, L. L., Yee, A. & Loh, H. S. Association between subclinical hypothyroidism and depression: an updated systematic review and meta-analysis. *BMC Psychiatry* **19**, 12. https://doi.org/10.1186/s12888-018-2006-2 (2019).
9. Demartini, B. *et al.* Depressive symptoms and major depressive disorder in patients affected by subclinical hypothyroidism: a cross-sectional study. *J. Nerv. Ment. Dis.* **202**, 603–607. https://doi.org/10.1097/NMD.0000000000000168 (2014).
10. Krysiak, R., Szkrobka, W. & Okopien, B. The effect of l-thyroxine treatment on sexual function and depressive symptoms in men with autoimmune hypothyroidism. *Pharmacol. Rep.* **69**, 432–437. https://doi.org/10.1016/j.pharep.2017.01.005 (2017).
11. Blum, M. R. *et al.* Subclinical thyroid dysfunction and depressive symptoms among the elderly: a prospective cohort study. *Neuroendocrinology* **103**, 291–299. https://doi.org/10.1159/000437387 (2016).
12. Park, Y. J. *et al.* Subclinical hypothyroidism (SCH) is not associated with metabolic derangement, cognitive impairment, depression or poor quality of life (QoL) in elderly subjects. *Arch. Gerontol. Geriatr.* **50**, e68-73. https://doi.org/10.1016/j.archger.2009.05.015 (2010).
13. Gussekloo, J. *et al.* Thyroid status, disability and cognitive function, and survival in old age. *JAMA* **292**, 2591–2599. https://doi.org/10.1001/jama.292.21.2591 (2004).
14. Hong, J. W., Noh, J. H. & Kim, D. J. Association between subclinical thyroid dysfunction and depressive symptoms in the Korean adult population: The 2014 Korea National Health and Nutrition Examination Survey. *PLoS ONE* **13**, e0202258. https://doi.org/10.1371/journal.pone.0202258 (2018).
15. Grabe, H. J. *et al.* Mental and physical complaints in thyroid disorders in the general population. *Acta Psychiatr. Scand.* **112**, 286–293. https://doi.org/10.1111/j.1600-0447.2005.00586.x (2005).
16. Kim, J. S. *et al.* Subclinical hypothyroidism and incident depression in young and middle-age adults. *J. Clin. Endocrinol. Metab.* **103**, 1827–1833. https://doi.org/10.1210/jc.2017-01247 (2018).
17. Burke, D. L., Ensor, J. & Riley, R. D. Meta-analysis using individual participant data: one-stage and two-stage approaches, and why they may differ. *Stat. Med.* **36**, 855–875. https://doi.org/10.1002/sim.7141 (2017).
18. Wildisen, L. *et al.* Subclinical thyroid dysfunction and depressive symptoms: protocol for a systematic review and individual participant data meta-analysis of prospective cohort studies. *BMJ. Open* **9**, e029716. https://doi.org/10.1136/bmjopen-2019-029716 (2019).
19. Stewart, L. A. *et al.* Preferred reporting items for systematic review and meta-analyses of individual participant data: the PRISMA-IPD statement. *JAMA* **313**, 1657–1665. https://doi.org/10.1001/jama.2015.3656 (2015).
20. Rodondi, N. *et al.* Subclinical hypothyroidism and the risk of coronary heart disease and mortality. *JAMA* **304**, 1365–1374. https://doi.org/10.1001/jama.2010.1361 (2010).
21. 21Wells G, S. B., O'Connell D, Peterson J, Welch V, Losos M, Tugwell P. *The Newcastle-Ottawa Scale (NOS) for assessing the quality of nonrandomised studies in meta-analyses* https://www.ohri.ca/programs/clinical_epidemiology/oxford.asp (2016).

22. Guyatt, G. H. *et al.* GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ* **336**, 924–926. https://doi.org/10.1136/bmj.39489.470347.AD (2008).

23. Segna, D. *et al.* Association between subclinical thyroid dysfunction and change in bone mineral density in prospective cohorts. *J. Intern. Med.* **283**, 56–72. https://doi.org/10.1111/joim.12688 (2018).

24. Blum, M. R. *et al.* Subclinical thyroid dysfunction and fracture risk: a meta-analysis. *JAMA* **313**, 2055–2065. https://doi.org/10.1001/jama.2015.5161 (2015).

25. Schneider, C. *et al.* Initial evaluation of thyroid dysfunction—are simultaneous TSH and fT4 tests necessary?. *PLoS ONE* **13**, e0196631. https://doi.org/10.1371/journal.pone.0196631 (2018).

26. Hollowell, J. G. *et al.* Serum TSH, T(4), and thyroid antibodies in the United States population (1988 to 1994): National Health and Nutrition Examination Survey (NHANES III). *J. Clin. Endocrinol. Metab.* **87**, 489–499. https://doi.org/10.1210/jcem.87.2.8182 (2002).

27. Zhou, X. *et al.* Comparative efficacy and tolerability of new-generation antidepressants for major depressive disorder in children and adolescents: protocol of an individual patient data meta-analysis. *BMJ Open* **8**, e018357. https://doi.org/10.1136/bmjopen-2017-018357 (2018).

28. Smarr, K. L. & Keefer, A. L. Measures of depression and depressive symptoms: Beck Depression Inventory-II (BDI-II), Center for Epidemiologic Studies Depression Scale (CES-D), Geriatric Depression Scale (GDS), Hospital Anxiety and Depression Scale (HADS), and Patient Health Questionnaire-9 (PHQ-9). *Arthritis Care Res. (Hoboken)* **63**(Suppl 11), S454-466. https://doi.org/10.1002/acr.20556 (2011).

29. GRADE Working Group. *About Re-expressing SMD*, https://pregnancy.cochrane.org/sites/pregnancy.cochrane.org/files/public/uploads/re-expressing%20SMD%20new.pdf. Accessed 2 September 2019

30. *Geriatric Depression Scale (Short Form)*, https://integrationacademy.ahrq.gov/sites/default/files/Update_Geriatric_Depression_Scale-15_0.pdf. Accessed 31 October 2019

31. Henry, S. K., Grant, M. M. & Cropsey, K. L. Determining the optimal clinical cutoff on the CES-D for depression in a community corrections sample. *J. Affect. Disord.* **234**, 270–275. https://doi.org/10.1016/j.jad.2018.02.071 (2018).

32. Win, S. *et al.* Depressive symptoms, physical inactivity and risk of cardiovascular mortality in older adults: the Cardiovascular Health Study. *Heart* **97**, 500–505. https://doi.org/10.1136/hrt.2010.209767 (2011).

33. Chan, Y. F., Leung, D. Y., Fong, D. Y., Leung, C. M. & Lee, A. M. Psychometric evaluation of the hospital anxiety and depression scale in a large community sample of adolescents in Hong Kong. *Qual. Life Res.* **19**, 865–873. https://doi.org/10.1007/s11136-010-9645-1 (2010).

34. Kalmijn, S. *et al.* Subclinical hyperthyroidism and the risk of dementia The Rotterdam study. *Clin. Endocrinol. (Oxf.)* **53**, 733–737. https://doi.org/10.1046/j.1365-2265.2000.01146.x (2000).

35. Riley, R. D. *et al.* Meta-analysis of randomised trials with a continuous outcome according to baseline imbalance and availability of individual participant data. *Stat. Med.* **32**, 2747–2766. https://doi.org/10.1002/sim.5726 (2013).

36. Almeida, O. P. *et al.* Thyroid hormones and depression: the Health in Men study. *Am. J. Geriatr. Psychiatry.* **19**, 763–770. https://doi.org/10.1097/JGP.0b013e31820dcad5 (2011).

37. Orwoll, E. *et al.* Design and baseline characteristics of the osteoporotic fractures in men (MrOS) study–a large observational study of the determinants of fracture in older men. *Contemp. Clin. Trials* **26**, 569–585. https://doi.org/10.1016/j.cct.2005.05.006 (2005).

38. Volzke, H. Study of health in pomerania (SHIP). Concept, design and selected results. *Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz* **55**, 790–794. https://doi.org/10.1007/s00103-012-1483-6 (2012).

39. Diercks, G. F. *et al.* Rationale, design, and baseline characteristics of a trial of prevention of cardiovascular and renal disease with fosinopril and pravastatin in nonhypertensive, nonhypercholesterolemic subjects with microalbuminuria (the Prevention of REnal and Vascular ENdstage Disease Intervention Trial [PREVEND IT]). *Am. J. Cardiol.* **86**, 635–638. https://doi.org/10.1016/s0002-9149(00)01042-0 (2000).

40. *Cochrane Handbook*, https://handbook-5-1.cochrane.org/chapter_10/10_4_3_1_recommendations_on_testing_for_funnel_plot_asymmetry.htm. Acessed 01 October 2019

41. Boekholdt, S. M. *et al.* Initial thyroid status and cardiovascular risk factors: the EPIC-Norfolk prospective population study. *Clin. Endocrinol. (Oxf)* **72**, 404–410. https://doi.org/10.1111/j.1365-2265.2009.03640.x (2010).

42. Debray, T. P. *et al.* Get real in individual participant data (IPD) meta-analysis: a review of the methodology. *Res.. Synth. Methods* **6**, 293–309. https://doi.org/10.1002/jrsm.1160 (2015).

43. Bekkering, G. E. *et al.* Thyroid hormones treatment for subclinical hypothyroidism: a clinical practice guideline. *BMJ* **365**, 2006. https://doi.org/10.1136/bmj.l2006 (2019).

44. Garber, J. R. *et al.* Clinical practice guidelines for hypothyroidism in adults: cosponsored by the American Association of Clinical Endocrinologists and the American Thyroid Association. *Thyroid* **22**, 1200–1235. https://doi.org/10.1089/thy.2012.0205 (2012).

45. Feller, M. *et al.* Association of thyroid hormone therapy with quality of life and thyroid-related symptoms in patients with sub-clinical hypothyroidism: a systematic review and meta-analysis. *JAMA* **320**, 1349–1359. https://doi.org/10.1001/jama.2018.13770 (2018).

46. Morsink, L. F. *et al.* Associations between sex steroid hormone levels and depressive symptoms in elderly men and women: results from the Health ABC study. *Psychoneuroendocrinology* **32**, 874–883. https://doi.org/10.1016/j.psyneuen.2007.06.009 (2007).

47. Vogelzangs, N. *et al.* Metabolic depression: a chronic depressive subtype? Findings from the InCHIANTI study of older persons. *J. Clin. Psychiatry* **72**, 598–604. https://doi.org/10.4088/JCP.10m06559 (2011).

48. Panicker, V. *et al.* A paradoxical difference in relationship between anxiety, depression and thyroid function in subjects on and not on T4: findings from the HUNT study. *Clin. Endocrinol. (Oxf.)* **71**, 574–580. https://doi.org/10.1111/j.1365-2265.2008.03521.x (2009).

49. Smink, P. A. *et al.* Albuminuria, estimated GFR, traditional risk factors, and incident cardiovascular disease: the PREVEND (prevention of renal and vascular endstage disease) study. *Am. J. Kidney. Dis.* **60**, 804–811. https://doi.org/10.1053/j.ajkd.2012.06.017 (2012).

50. Cawthon, P. M., Shahnazari, M., Orwoll, E. S. & Lane, N. E. Osteoporosis in men: findings from the osteoporotic fractures in men study (MrOS). *Ther. Adv. Musculoskelet. Dis.* **8**, 15–27. https://doi.org/10.1177/1759720X15621227 (2016).

51. Ittermann, T., Volzke, H., Baumeister, S. E., Appel, K. & Grabe, H. J. Diagnosed thyroid disorders are associated with depression and anxiety. *Soc. Psychiatry Psychiatr. Epidemiol.* **50**, 1417–1425. https://doi.org/10.1007/s00127-015-1043-0 (2015).

Acknowledgements

Author contributions

L.W., M.F. and N.R. have full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. L.W., M.F. and N.R. had the final responsibility for the decision to submit for publication. Concept and design: L.W., S.K., M.F., C.D.G., N.R. Acquisition, analysis, or interpretation of data: L.W., N.R., C.D.G., M.F. Drafting of the manuscript: L.W., C.D.G., M.F. Critical revision of the manuscript for important intellectual content: E.M., S.B., L.S., T.C., A.R.C, B.O.Å, S.J.L.B., B.B.Y., O.P.A., G.C., R.P.F.D., L.F., H.G., J.W.J., M.N., S.T., H.V., R.W., J.G., R.P., S.K., D.A., D.B., N.R.. Statistical analysis: L.W., L.S., C.D.G. Obtained funding: N.R. Administrative, technical, or material support: L.W., E.M., S.B., A.R.C., B.O.Å., S.J.L.B., B.B.Y., O.P.A., G.C., R.P.F.D., L.F., H.G., J.W.J., M.N., S.T., H.V., R.W., J.G., R.P., D.B. Supervision: N.R., C.D.G., M.F. All authors read and approved the final manuscript.

Funding

Competing interests

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41598-020-75776-1.

Correspondence and requests for materials should be addressed to L.W.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.