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

Lea Wildisen1,2, Cinzia Del Giovane1, Elisavet Moutzouri1,2, Shanthi Beglinger1,2, Lamprimi Syrogiannouli1, Tinh-Hai Collet3, Anne R. Cappola4, Bjørn O. Åsvold5,6, Stephan J. L. Bakker7, Bu B. Yeap8, Osvaldo P. Almeida8, Graziano Ceresini9, Robin P. F. Dullaart7, Luigi Ferrucci10, Hans Grabe11, J. Wouter Jukema12, Matthias Nauck13,14, Stella Trompet15, Henry Völzke16, Rudi Westendorp17, Jacobijn Gussekloo15,18, Stefan Klöppel19, Drahomir Aujesky2, Douglas Bauer20, Robin Peeters21, Martin Feller1,2 & Nicolas Rodondi1,2

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...
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, I² = 15.6) or subclinical hyperthyroidism (−0.10, 95% confidence interval = −0.67 to 0.48, I² = 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  
SHyper 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. 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. 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.

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. The largest prospective study published showed no association between subclinical hypothyroidism and incidence of depression after 2 years of follow-up, whereas depressive symptoms were associated with subclinical hyperthyroidism (but not subclinical hypothyroidism) in another prospective study.
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\cite{21,22}. 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\cite{23}. This study adheres to the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) statement for IPD systematic reviews\cite{24}.

**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\cite{25}. 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\cite{26}.

**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\cite{27}. 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 classified 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\cite{28}. 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\cite{29,30,31}. 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\cite{32}. 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\cite{33,34}. 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\cite{35,36}. 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\cite{37}. 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\(^{14,28}\). 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 or 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, 10.00–20.00 mIU/L)\(^{30}\), 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\(^{36}\). 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
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***7 The health, aging and body composition study; *CHS***32 cardiovascular health study; **InChianti Study***7 Invecchiare in Chianti Study; **HUNT**96 Nord-Trøndelag Health Study; *PREVEND***6 prevention of renal and vascular end-stage disease; **MrOS**90 osteoporotic fractures in men study; **SHIP**97 study of health in pomerania, **HIMS**98 health in men study; *y* year; **GDS-15** geriatric depression scale 15-item; **CESD-10/-20** 10/20-item scale; **HADS** hospital anxiety and depression scale; **BDI** Beck depression inventory scale; **TSH** thyroid stimulating hormone; **FT4** free thyroxine; **GDS-15** geriatric depression scale 15-item; **CESD-10/-20** 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 normal.*

| 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* |
|-------------|---------------------|------------------------|----------------|-----------------------------------|------------------|-----------------------------|-------------------------------|----------------|-------------------------|---------------------------------------------|
| 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 (51) | 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

| 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* |
|-------------|---------------------|------------------------|----------------|-----------------------------------|------------------|-----------------------------|-------------------------------|----------------|-------------------------|---------------------------------------------|
| 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 + Study13 Leiden 85 plus Study; PROSPER11 prospective study of Pravastatin in the elderly at risk; Health ABC Study7 The health, aging and body composition study; CHS32 cardiovascular health study; InChianti Study7 Invecchiare in Chianti Study; HUNT96 Nord-Trøndelag Health Study; PREVEND6 prevention of renal and vascular end-stage disease; MrOS60 osteoporotic fractures in men study; SHIP97 study of health in pomerania, HIMS98 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 normal. The hypothesis was tested using a meta-analysis. The mean difference (MD) in BDI from the baseline was calculated. The mean 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 (I2 = 15.6%) (Fig. 1).

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 bias60.

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 (I2 = 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 0.68***).
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 % |
|--------------------|---------------------|-------------------------------------|----------------------------------|-------------------|----------|
| 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 %, τ² = 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 % |
|--------------------|---------------------|-------------------------------------|----------------------------------|-------------------|----------|
| 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 %, τ² = 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,
I² = 0%). The odds ratio for incident depression comparing subclinical hypothyroid and euthyroid participants was 1.24 (95% CI 0.73–2.09, I² = 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, I² = 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 from 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...
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, 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. With individual participant data, we analyzed individual participant data, which leads to far more reliable results. We 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 were limited by measurement of depressive symptoms using different scales across studies, so that we were limited by measurement of depressive symptoms using different scales across studies, so that we

Table 2. Sensitivity analyses on subclinical hypothyroidism and depressive symptoms at first available follow-up. 

| Sensitivity analysis | No. of participants | No. of included studies | MD in BDI score (95% CI), I² |
|---------------------|----------------------|-------------------------|-----------------------------|
| Main analysis       | Euthyroid | SHypo | 6 | 0.29 (−0.17–0.76) \(^1\), 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 antidepressant medication | 20,897 | 1327 | 6 | 0.32 (−0.19−0.83), 26.1% |
| 4) Excluding participants with FT4 measurement | 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% |
| 10) Using original scale for depressive symptoms | 21,025 | 1342 | 6 | 0.04 (−0.02−0.09), 27.2% |

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, 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. With individual participant data, we could use standardized definition of subclinical hypothyroidism 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.

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...
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. 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

| Outcome by Subgroup | No. of Participants | Higher Depression Score | Higher Depression Score in SHypo | MD in BDI Score(95% CI) |
|---------------------|---------------------|--------------------------|----------------------------------|-------------------------|
|                     | Euthyroid           | SHypo                    | 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)      |
| Levotyroxine 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.
verified measurement, which is a limitation of most published large cohorts that have examined the risk of subclinical thyroid dysfunction\textsuperscript{20,24}. 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\textsuperscript{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\textsuperscript{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\textsuperscript{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\textsuperscript{39}. 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.

Received: 3 March 2020; Accepted: 20 October 2020

Published online: 05 November 2020

**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.206-0043 (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. *Endocrin. Pract.* **8**, 457–469 (2002).

3. Garber, J. E. et al. Clinical practice guidelines for hypothyroidism in adults: cosponsored by the American Association of Clinical Endocrinologists and the American Thyroid Association. *Endocrin. Pract.* **18**, 988–1028. https://doi.org/10.4188/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. Okoszie, 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., McCaohon, D., Hobbs, F. D. & Roberts, L. M. Why are GP's treating subclinical hypothyroidism? Case note review and GP survey. *Prim. Health Care Res. Dev.* **14**, 175–184. https://doi.org/10.1017/S1463423216000230 (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/s12888-018-1006-2 (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., Szkroba, 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. Guuskeloo, 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. et al. 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. Wells 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).
Acknowledgements

The authors thank Beatrice Minder and Doris Kopp (Institute of Social and Preventive Medicine (ISPM), University of Bern, Switzerland) for helping us to develop the literature search strategy, Kali Tal, PhD (Institute of Primary Health Care (BIHAM), University of Bern, Switzerland) for editing the manuscript, and the Thyroid Studies Collaboration (www.thyroid-studies.org) for their contribution in this study.
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., S.J.L.B., B.B.Y., O.P.A., G.C., R.P.F.D., L.E., H.G., J.W.J., M.N., S.T., H.V., R.W., J.G., R.F., 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., 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.F., D.B. Supervision: N.R., C.D.G., M.F. All authors read and approved the final manuscript.

Funding
This systematic review and IPD analysis was funded by a grant from the Swiss National Science foundation (SNSF 320030-172676 to Nicolas Rodondi). The HABC study was also supported in part by National Institute on Aging (NIA) Contracts N01-AG-6-2101; N01-AG-6-2103; N01-AG-6-2106; NIA grant R01-AG028050, and NINR grant R01-NR012459, and by the Intramural Research Program of the NIH, National Institute on Aging. This research was also supported in part by SHIP. SHIP is part of the Community Medicine Research net of the University of Greifswald, Germany, which is funded by the Federal Ministry of Education and Research (grants no. 01ZZ9603, 01ZZ0103, and 01ZZ0403), the Ministry of Cultural Affairs and the Social Ministry of the Federal State of Mecklenburg-West Pomerania. The SHIP study was further support by the German Research Foundation (DFG GR 1912/5-1). The HUNT Study did also in part support this study. HUNT is a collaboration between HUNT Research Centre (Faculty of Medicine and Health Sciences, NTNU, Norwegian University of Science and Technology), Trøndelag County Council, Central Norway Regional Health Authority, and the Norwegian Institute of Public Health. The Cardiovascular Health Study (CHS) is supported by contracts HHSN268201200036C, HHSN268201200036C, HHSN268201200036C, NIH Roadmap for Medical Research under the following grant numbers: U01 AG027810, U01 AG042124, U01 AG042139, U01 AG042140, U01 AG042143, U01 AG042145, U01 AG042168, U01 AR066160, and UL1 TR000128. The Health In Men Study is supported by research grants from the National Health and Medical Research Council of Australia. All agencies had no role in the design and conduct of the study, collection, management, analysis, or interpretation of the data, or preparation, review, or approval of the manuscript.

Competing interests
HG has received travel grants and speakers honoraria from Fresenius Medical Care, Neuraxpharm and Janssen Cilag as well as research funding from Fresenius Medical Care. No other author has declared 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.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2020