The Lancet Primary Care

# Weight-based levothyroxine dosing and long-term cardiovascular, skeletal, and mortality outcomes in older adults: an emulated target trial using UK primary care data

--Manuscript Draft--

| | |
|---|---|
| **Manuscript Number:** | thelancetprimarycare-D-26-00085R2 |
| **Article Type:** | Article (Original Research) |
| **Keywords:** | Hypothyroidism; Levothyroxine; Older adults; Weight-based dosing; Target trial emulation; Cardiovascular disease; Bone health; Mortality; Primary care; Electronic health records |
| **Corresponding Author:** | Mia Holley<br>University of Sunderland<br>UNITED KINGDOM OF GREAT BRITAIN AND NORTHERN IRELAND |
| **First Author:** | Mia Holley |
| **Order of Authors:** | Mia Holley |
| | Scott Wilkes |
| | Ellen Dorothea Moss |
| | Ian Maxwell |
| | Rosie Dew |
| | Salman Razvi |
| **Manuscript Region of Origin:** | UNITED KINGDOM OF GREAT BRITAIN AND NORTHERN IRELAND |
| **Abstract:** | Background: Levothyroxine (LT4) is widely prescribed in older adults with hypothyroidism, yet evidence guiding safe initial weight-based dosing is limited. Both under- and over-replacement may adversely affect cardiovascular health, bone integrity, and survival. Given the impracticality of conducting a long-term randomised trial in this population, this study aimed to emulate a target trial to assess associations between initial LT4 dose per body weight and long-term clinical outcomes in older adults, using routinely collected primary care data.<br>Methods: An emulated target trial was conducted using The Health Improvement Network, a UK primary care database, including data from 2006 to 2021. Eligible participants were adults aged >50 years diagnosed with hypothyroidism initiating LT4 during this period. The target trial compared initiation of LT4 at >1·0 versus ≤1·0mcg/kg/day using total and lean body weight definitions. Inverse probability of treatment weighting adjusted for baseline confounders. Associations with cardiovascular events, bone outcomes, and all-cause mortality were estimated using weighted Cox proportional hazards models, and 10-year absolute risk differences were calculated. Participants were followed for up to 10 years from LT4 initiation until outcome occurrence, death, or study end.<br>Findings: Among 15,463 patients (mean age 67 years, 78% female), 11,762 (76%) initiated LT4 at ≤1·0mcg/kg/day and 3,701 (24%) at >1·0mcg/kg/day (total body weight). Unweighted event rates were higher in the higher-dose group for cardiovascular events (19·7% vs 17·7%), bone outcomes (13·1% vs 9·1%), and all-cause mortality (18·5% vs 17·5%). Higher LT4 dosing was associated with increased risk of cardiovascular events (hazard ratio (HR) 1·18, 95% CI 1·08–1·29), bone outcomes (1·29, 1·16–1·43), and all-cause mortality (1·17, 1·09–1·25). The weighted 10-year absolute risk difference between higher and lower dose groups was 1·3% (95% CI 0·3–3·0) for cardiovascular events, 3·0% (1·6–4·5) for bone outcomes, and 3·4% (1·7–5·1) for all-cause mortality. Similar associations were observed using lean body weight-based dosing.<br>Interpretation: These findings suggest that higher initial weight-based LT4 dosing may be associated with avoidable long-term harm, supporting cautious initiation and titration in older adults in primary care. These findings describe associations and should be interpreted cautiously. |

# Weight-based levothyroxine dosing and long-term cardiovascular, skeletal, and mortality outcomes in older adults: an emulated target trial using UK primary care data.

Mia Holley, PhD[1*], Prof Scott Wilkes, PhD[1], Ellen Dorothea Moss, PhD[2], Ian Maxwell, MSc[1], Rosie Dew, PhD[1], Salman Razvi, MD[3].

1. School of Medicine, University of Sunderland, Sunderland, UK
2. Population Health Sciences Institute, Newcastle University, Newcastle upon Tyne, UK
3. Translational and Clinical Research Institute, Newcastle University, Newcastle upon Tyne, UK

* Corresponding author: mia.holley@sunderland.ac.uk

## Research in context

### Evidence before this study

PubMed was searched from January 1, 2000, to October 1, 2025, using combinations of terms for levothyroxine ("levothyroxine", "thyroxine", "LT4"), weight-based dosing ("dose per body weight", "weight-based dose", "mcg/kg", "μg/kg"), hypothyroidism, ageing ("older adults", "elderly", "aged"), and outcomes ("cardiovascular", "fracture", "osteoporosis", "bone", "mortality"). No language restrictions were applied. No studies were identified that examined the associations between initial levothyroxine dose per body weight and cardiovascular, skeletal, or mortality outcomes in adults aged 50 years or older. Existing evidence has focused primarily on the risks of overtreatment, defined by suppressed thyroid-stimulating hormone levels, rather than the prescribed dose per body weight. Current guidance from the United Kingdom National Institute for Health and Care Excellence (NICE) recommends fixed low starting doses of levothyroxine (25-50mcg/day) for adults aged 65 years and older, irrespective of body weight. In contrast, younger adults are initiated on weight-based doses. The Baltimore Longitudinal Study of Ageing reported a mean requirement of approximately 1·09mcg/kg/day to maintain euthyroidism in adults aged 65 years and older, suggesting that lower weight-adjusted levothyroxine doses may be appropriate; however, no large-scale study has directly examined their association with long-term outcomes.

### Added value of this study

Using primary care data from more than 15,000 older adults with hypothyroidism in the United Kingdom, this study emulated a target trial to assess the association between initial levothyroxine dose per body weight and cardiovascular, bone, and mortality outcomes. After adjustment for baseline confounding, initiating levothyroxine at doses greater than 1·0mcg/kg/day, whether based on total or lean body weight, was associated with higher risks of adverse cardiovascular events, skeletal events and all-cause mortality.

### Implications of all the available evidence

This evidence suggests that standard dosing may overestimate levothyroxine requirements in adults over 50 years old. Weight-adjusted, age-sensitive dosing, targeting approximately 1·0mcg/kg/day, may be associated with fewer adverse outcomes. These findings support revising clinical guidance to incorporate weight-based levothyroxine prescribing for people aged 65 years and older.

**Abstract**

**Background:** Levothyroxine (LT4) is widely prescribed in older adults with hypothyroidism, yet evidence guiding safe initial weight-based dosing is limited. Both under- and over-replacement may adversely affect cardiovascular health, bone integrity, and survival. Given the impracticality of conducting a long-term randomised trial in this population, this study aimed to emulate a target trial to assess associations between initial LT4 dose per body weight and long-term clinical outcomes in older adults, using routinely collected primary care data.

**Methods:** An emulated target trial was conducted using The Health Improvement Network, a UK primary care database, including data from 2006 to 2021. Eligible participants were adults aged >50 years diagnosed with hypothyroidism initiating LT4 during this period. The target trial compared initiation of LT4 at >1·0 versus ≤1·0mcg/kg/day using total and lean body weight definitions. Inverse probability of treatment weighting adjusted for baseline confounders. Associations with cardiovascular events, bone outcomes, and all-cause mortality were estimated using weighted Cox proportional hazards models, and 10-year absolute risk differences were calculated. Participants were followed for up to 10 years from LT4 initiation until outcome occurrence, death, or study end.

**Findings:** Among 15,463 patients (mean age 67 years, 78% female), 11,762 (76%) initiated LT4 at ≤1·0mcg/kg/day and 3,701 (24%) at >1·0mcg/kg/day (total body weight). Unweighted event rates were higher in the higher-dose group for cardiovascular events (19·7% vs 17·7%), bone outcomes (13·1% vs 9·1%), and all-cause mortality (18·5% vs 17·5%). Higher LT4 dosing was associated with increased risk of cardiovascular events (hazard ratio (HR) 1·18, 95% CI 1·08–1·29), bone outcomes (1·29, 1·16–1·43), and all-cause mortality (1·17, 1·09–1·25). The weighted 10-year absolute risk difference between higher and lower dose groups was 1·3% (95% CI 0·3–3·0) for cardiovascular events, 3·0% (1·6–4·5) for bone outcomes, and 3·4% (1·7–5·1) for all-cause mortality. Similar associations were observed using lean body weight-based dosing.

**Interpretation:** These findings suggest that higher initial weight-based LT4 dosing may be associated with avoidable long-term harm, supporting cautious initiation and titration in older adults in primary care. These findings describe associations and should be interpreted cautiously.

**Funding**: This study was funded by the National Institute for Health and Care Research (NIHR) Applied Research Collaboration (ARC) North East and North Cumbria (NENC) (NIHR200173).

## Introduction

Hypothyroidism is a common chronic endocrine disorder characterised by insufficient thyroid hormone production, particularly thyroxine (T4).[1] Its prevalence increases with age and is notably higher in women, affecting up to 21% of those over 74 years old.[2] This higher prevalence in older adults likely reflects age-related rises in thyroid-stimulating hormone (TSH) levels,[3] a greater susceptibility to autoimmune thyroid disease, and prior treatments that impair thyroid function, such as thyroid surgery and radioactive iodine therapy.[4]

Levothyroxine (LT4), a synthetic form of T4, remains the standard treatment and is among the most widely prescribed medications globally.[5] In the United Kingdom (UK), prescribing rates are highest in people aged over 80 years, where more than 10% receive LT4.[6] The treatment goal is to restore and maintain euthyroidism, typically assessed by TSH normalisation.[7]

Older adults, however, are particularly vulnerable to the effects of suboptimal LT4 dosing. Both under- and over-replacement, indicated by persistently abnormal TSH levels outside the population reference range, have been associated with adverse cardiovascular and bone outcomes.[8,9] Standard weight-based dosing (1·6mcg/kg/day) is derived from studies in younger adults and may overestimate LT4 needs in older individuals, who have lower lean body mass and heightened peripheral sensitivity to thyroid hormones.[10] Reflecting this, the UK's National Institute for Health and Care Excellence (NICE) recommends weight-based dosing (1·6mcg/kg/day) in adults under 65 years without cardiovascular disease, but substantially lower fixed starting doses (20-50mcg/day) with titration in those aged 65 years and older or with cardiovascular disease.[11] The American Thyroid Association similarly advises individualised dosing based on both age and weight.[12] In addition, a cautious approach to treatment in older adults is often advocated, particularly where biochemical abnormalities are mild.[13,14] Despite these differences, there remains limited evidence directly examining the relationship between initial LT4 dose per body weight and long-term clinical outcomes in ageing populations.

Long-term randomised trials comparing initial LT4 dosing strategies in older adults are unlikely to be feasible. Observational data structured within a target trial emulation framework therefore provide an opportunity to address this evidence gap using real-world clinical practice data. By explicitly defining eligibility criteria, treatment strategies, follow-up, and outcomes, target trial emulation can strengthen causal inference from observational data and reduce biases commonly encountered in observational studies.

Accordingly, a target trial was emulated using UK primary care data to examine the association between initial LT4 dose per body weight and major clinical outcomes in adults with primary hypothyroidism. Individuals aged over 50 years were included, reflecting evidence from the TEARS study demonstrating age-related increases in TSH in older adults.[3] Participants were stratified into two age groups: 51–64 years, for whom standard weight-based dosing (1·6mcg/kg/day) is generally recommended, and ≥65 years, for whom NICE advocates lower fixed starting doses (25–50mcg/day).[11] The causal question of interest was whether initiating LT4 at higher versus lower weight-based doses is associated with differences in long-term cardiovascular and skeletal outcomes, and mortality in routine clinical practice, with the objective of estimating the comparative benefits and harms of these dosing strategies in older adults with hypothyroidism.

Initial LT4 dose was expressed per unit of total body weight (TBW) and estimated lean body weight (LBW), allowing comparison across two clinically relevant metrics. Outcomes included (1) cardiovascular events (composite of angina, myocardial infarction, heart failure, stroke, peripheral vascular disease, atrial fibrillation, and revascularisation), (2) bone outcomes (composite of fragility fractures and osteoporosis), and (3) all-cause mortality. Higher versus lower dosing strategies were compared using a threshold of 1·0mcg/kg/day, informed by data from the Baltimore Longitudinal Study of Aging, which reported a mean LT4 requirement of 1·09mcg/kg/day in adults aged ≥65 years old.[12]

## Methods

### Study design and participants

Data for this study were obtained from The Health Improvement Network (THIN), a UK primary care database containing anonymised electronic health records from over 850 general practices using the Vision software system, representing approximately 6% of the UK

population. THIN includes longitudinal information on demographics, diagnoses, prescriptions, and laboratory test results recorded in routine primary care and is broadly representative of the UK population in terms of age, sex distribution, major disease prevalence, and mortality rates.[15] Data from January 1, 2006, to December 31, 2021, were available for analysis. THIN captures data recorded by general practices and does not include linked secondary care or hospital admission data for this study. The database has been validated and widely used in epidemiological research examining LT4 therapy, thyroid function, cardiovascular disease, fractures, and mortality.[9,16]

This retrospective observational study emulated a target trial to estimate the association between initial LT4 dose and long-term cardiovascular, skeletal, and mortality outcomes in adults with hypothyroidism. The design followed the target trial framework proposed by Hernán and Robins,[17] which uses observational data to reproduce key features of a hypothetical randomised controlled trial, including explicit specification of eligibility criteria, treatment strategies, assignment procedures, and follow-up. This study follows both the Strengthening the Reporting of Observational Studies in Epidemiology guidelines for Observational Studies Reporting Propensity Score Analysis and the Transparent Reporting of Observational Studies Emulating a Target Trial checklists (Appendix pages 1-7).[18,19] The primary causal contrast of interest compared participants according to the initial prescribed LT4 dose at treatment initiation. Participants were classified according to their initial prescribed dose at treatment initiation, irrespective of subsequent dose modifications. An overview of the target trial specification and its emulation is presented in Table 1.

Patients aged over 50 years with at least one recorded TSH measurement exceeding 4·0 mU/L between January 1, 2006, and January 1, 2022, were extracted from THIN. Eligibility was assessed at the time of LT4 initiation (time zero). Participants were eligible if they met all of the following:

- Aged over 50 years at the time of LT4 initiation.
- Recorded diagnosis of primary hypothyroidism.
- Initiated LT4 therapy between January 1, 2006 and December 31, 2021.
- At least one serum TSH measurement recorded within three months before or after LT4 initiation.

- At least one body weight measurement within three months before or after LT4 initiation.
- At least one height measurement (*lean body mass analyses only*).
- At least six months of follow-up data available after LT4 initiation. Six months of follow-up data refers to the availability of routine primary care records for at least six months after LT4 initiation, ensuring sufficient observation time for outcome events.

Participants were excluded if they met any of the following:

- History of thyroid cancer, hyperthyroidism, or pituitary disease.
- Previous thyroid surgery or radioiodine therapy.
- Use of antithyroid drugs, liothyronine, or medications known to substantially affect thyroid function (including lithium and amiodarone) during the study period.
- Pre-existing cardiovascular or bone disease relevant to the outcome under assessment.

Each of the three outcomes (cardiovascular disease, bone health, and all-cause mortality) were analysed separately, using bespoke patient cohorts. Diagnostic, prescription, and procedure codes used to define inclusion and exclusion criteria were derived from the International Classification of Diseases, Tenth Revision (ICD-10), dictionary of medicines and devices (dm+d), and Office of Population Censuses and Surveys Classification of Interventions and Procedures, 4th revision (OPCS-4) (Appendix pages 8-18).

Data cleaning included removal of duplicate records, verification of measurement units, checks for chronological consistency, and assessment of plausibility for key variables. TSH <0·1 or >20mU/L and body weight <35 or >200kg were excluded as extreme values. Records with missing baseline information or with inconsistent follow-up data were excluded.

Formal sample size calculations were not performed, as the study aimed to include the full eligible population defined by the target trial protocol to provide precise observational estimates rather than to test a single prespecified hypothesis.[17]

This study received ethical approval from the University of Sunderland Research Ethics Group (Reference: 011081). Scientific approval for the use of data from THIN was granted by the THIN Scientific Review Committee (Protocol Number 22-003). Individual patient

consent was waived because anonymised routinely collected electronic health record data were used.

The study protocol was prospectively developed and is available at https://sure.sunderland.ac.uk/id/eprint/19274/. No protocol amendments affecting eligibility criteria, treatment assignment, follow-up procedures, or study conduct occurred during the study period.

**Procedures**

THIN data are collected prospectively during routine primary care consultations by participating UK general practitioners and affiliated healthcare professionals using the Vision electronic health record system. Demographic data, diagnoses, prescriptions, laboratory test results, and consultation records entered during routine clinical care were available for analysis. Data on race and ethnicity were not consistently available within THIN and therefore could not be analysed.

Baseline demographic variables, including age and sex assigned at birth, were obtained from routinely recorded primary care electronic health records within THIN. Clinical covariates, including hypertension, Charlson Comorbidity Index, smoking status, medication use, serum TSH measurements, and body weight, were identified using diagnostic records, prescribing data, and laboratory test entries recorded during routine clinical care. Prescription data were recorded by general practitioners at the point of prescribing, and laboratory results were electronically transferred into the medical record from affiliated laboratories (Appendix pages 19-165). Cardiovascular, bone, and mortality outcomes were identified using ICD-10 and OPCS-4 codes recorded in THIN (Appendix pages 166-179).

To evaluate the impact of initial LT4 dose according to body composition, exposure was defined using two weight-based approaches based on the initial prescribed dose at LT4 initiation:

- Total Body Weight (TBW) dose (mcg/day/kg): initial LT4 dose (mcg/day) divided by TBW (kg).
- Lean Body Weight (LBW) dose (mcg/day/kg): initial LT4 dose (mcg/day) divided by estimated LBW (kg).

LBW was calculated using the Boer formula,[20] which incorporates sex assigned at birth, height, and weight:

- Males: LBW (kg) = (0·407 × weight (kg)) + (0·267 × height (cm)) − 19·2
- Females: LBW (kg) = (0·252 × weight (kg)) + (0·473 × height (cm)) − 48·3

For each dosing metric, participants were categorised into two groups:

1. **Higher Dose Group**: >1·0mcg/kg/day
2. **Lower Dose Group**: ≤1·0mcg/kg/day

These classifications were applied separately for TBW and LBW, with analyses conducted independently for each approach.

TSH and body weight were defined as the closest measurements recorded within three months before or after LT4 initiation. This window was selected to maximise data completeness and reflects routine UK practice.[11] When multiple measurements were available, the value closest to LT4 initiation was used. TSH was used as the primary biochemical marker as it is the most consistently recorded thyroid function test in routine UK clinical practice and is the primary parameter used to guide LT4 dose titration. Free T4 measurements were inconsistently available and therefore not suitable for inclusion in all analyses. Reported free T4 values are based on complete-case data only.

Potential confounders, defined as variables plausibly associated with both LT4 dose and study outcomes, were identified *a priori* using a directed acyclic graph specified in the study protocol and informed by clinical reasoning and prior evidence. The final adjustment set comprised baseline age, sex assigned at birth, Charlson Comorbidity Index, hypertension, serum TSH, smoking status, and use of proton-pump inhibitors, oestrogen, and testosterone.[21,22]

Participants were followed from the date of LT4 initiation until the earliest of:

1. Occurrence of the relevant outcome event (cardiovascular, bone, or mortality),
2. Death (for non-mortality outcomes), or
3. End of the study period.

Follow-up continued for up to ten years after LT4 initiation, with eligible participants contributing follow-up time between January 1, 2006, and December 31, 2021.

Several measures were undertaken to reduce potential sources of bias. Eligibility criteria, treatment strategies, and follow-up definitions were specified in advance according to a target trial framework to minimise selection and immortal time biases. Time zero was consistently defined as the date of LT4 initiation. Extreme biologically implausible values for TSH and body weight were excluded.

## Outcomes

*Primary Outcomes**:***

1. **Cardiovascular**: First occurrence of a composite outcome including angina, myocardial infarction, stroke, peripheral vascular disease, heart failure, atrial fibrillation, or any revascularisation procedures recorded during follow-up.
2. **Bone**: First recorded diagnosis of osteoporosis or a fragility fracture during follow-up.
3. **All-cause mortality**: Death recorded during follow-up. Data on cause-specific mortality were unavailable.

*Secondary Outcomes:*

1. **Healthcare utilisation:** Total number of primary care encounters (including face-to-face, telephone, or online consultations) per patient during follow-up.
2. **Thyroid function control:** Proportion of participants whose latest thyroid function test during follow-up showed an abnormal TSH level (<0·4 or >4·0mU/L).

## Statistical analysis

The primary causal contrast of interest was comparing higher versus lower LT4 dose groups based on dose at treatment initiation. Because treatment allocation was not randomised and longitudinal adherence data were incomplete, analyses estimated the observational analogue of the effect of treatment initiation according to baseline prescribed dose group.

Baseline characteristics of participants in the higher- and lower-dose LT4 groups were summarised using descriptive statistics. Continuous variables are reported as mean (standard

deviation) or median (interquartile range (IQR)), and categorical variables as counts and percentages. Covariate balance before and after weighting was assessed using standardised mean differences (SMDs). Baseline data were complete for included participants; therefore, no multiple imputation procedures were required.

To emulate randomisation and adjust for measured confounding, inverse probability of treatment weighting (IPTW) was applied to reduce confounding by indication and other baseline imbalances based on measured covariates. Propensity scores estimating the probability of receiving a higher LT4 dose were derived from logistic regression models.

Unstabilised weights were calculated as the inverse of the propensity score for participants in the higher-dose group and the inverse of one minus the propensity score for those in the lower-dose group, with extreme weights truncated at the 1st and 99th percentiles. Propensity score distributions were examined to assess overlap between treatment groups, ensuring adequate positivity (Appendix page 180). This created a weighted pseudo-population with balanced baseline covariates between treatment groups. E-values were calculated to assess the robustness of observed associations to potential unmeasured confounding.

Weighted Cox proportional hazards models estimated hazard ratios (HRs) and 95% confidence intervals (CIs) for each outcome. Time zero was defined as the date of LT4 initiation, and participants were followed until outcome occurrence, death, censoring, or study end. Proportional hazards assumptions were assessed using Schoenfeld residuals. Weighted Kaplan–Meier curves were used to illustrate cumulative incidence by LT4 dose group.[23] Absolute risk differences were estimated at 10 years.

To assess potential non-linear dose–response relationships, IPTW-adjusted Cox models incorporating natural cubic splines with three degrees of freedom were fitted. Healthcare utilisation was analysed using IPTW-weighted negative binomial regression, reporting incidence rate ratios (IRRs) with 95% CIs.

Four sensitivity analyses were conducted. First, statin use was additionally included in the inverse probability of treatment weighting model. Second, the calendar year was incorporated into the outcome model. Third, analyses were restricted to participants with TSH measurements recorded prior to LT4 initiation. Fourth, time zero was redefined as the

initiation of the second LT4 prescription among participants receiving a repeat prescription. Pre-specified subgroup analyses were conducted by sex assigned at birth, age group (51–64 years and ≥65 years), and TSH levels (≤10·0 and >10·0 mU/L). All analyses were conducted in R version 4.5.1, using the MatchIt package.[24]

## Role of the funding source

The funder of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report.

## Results

*Cardiovascular outcomes*

A total of 19,027 adults were eligible for analysis (Figure 1). Of those 19,027 patients, 15,463 patients were included in the primary cardiovascular analysis. Based on TBW, 11,762 participants (76%) initiated LT4 at a dose of ≤1·0mcg/kg/day and 3,701 (24%) at a dose of >1·0mcg/kg/day. After applying IPTW, the pseudo-population represented 15,430 (51%) and 14,899 (49%) participants in the low and high dose groups, respectively. Participants had a median follow-up of 10·0 years (IQR 8·6–10·0).

**Figure 1 Flow diagram illustrating cohort selection, inclusion and exclusion criteria, and allocation to outcome groups.**

After weighting, all covariates were well balanced between dose groups, with standardised mean differences <0·1 (Appendix page 181). The weighted mean age was 67 years in both treatment groups, with similar distributions of sex assigned at birth (22% versus 21% male), smoking status (43% in both), comorbidity burden, and concomitant medication use (Table 2, Appendix pages 182-184).

Median TSH was lower in the high dose group (5·7mU/L [2·6–10·2]) compared with the low-dose group (6·8mU/L [5·2–9·8]), and this difference persisted at the end of follow-up (2·3mU/L [1·2–4·6] versus 2·8mU/L [1·6–4·2]). Weighted mean free T4 concentrations were higher in the high-dose group both at baseline (15·2 versus 13·1pmol/L) and at the end of follow-up (17·3 versus 15·7pmol/L).

In analyses based on TBW, initiating LT4 at a higher dose was associated with a statistically significant increased risk of cardiovascular events compared with lower dose therapy (aHR 1·18, 95% CI 1·08–1·29) (Figure 2A). Consistent with these adjusted estimates, unweighted event rates were higher in the high-dose group (19·7% vs 17·7%). Similar findings were observed when dosing was defined by LBW (aHR 1·15, 95% CI 1·06–1·24) (Figure 2A). Subgroup analyses generally demonstrated consistent trends, although not all estimates reached statistical significance. The association was particularly evident among adults aged ≥65 years for both TBW (aHR 1·18, 95% CI 1·08–1·29) and LBW (aHR 1·15, 95% CI 1·06–1·24) based dosing. E-values for the cardiovascular outcomes ranged from 1·34 to 2·13, indicating that a moderately strong unmeasured confounder would be required to fully explain the observed associations (Appendix page 185). In sensitivity analyses, the association with cardiovascular events was attenuated and no longer statistically significant in the lag-adjusted analysis (Appendix pages 186-199).

**Figure 2 Adjusted hazard ratios with 95% confidence intervals (CI) and p-values for Cardiovascular, Bone, and All-Cause Mortality Events by Levothyroxine Dose per Total Body Weight and Lean Body Weight. (A) Cardiovascular events, (B) Bone outcomes, (C) All-cause mortality.**

Kaplan–Meier curves demonstrated a higher cumulative incidence of cardiovascular events among participants initiated on higher LT4 doses (>1·0mcg/kg/day, TBW) compared with lower doses (≤1·0mcg/kg/day, TBW) (Figure 3A). Similar patterns were observed when dosing was defined using LBW and split by sex assigned at birth (Appendix pages 200-201). The weighted 10-year absolute risk difference between higher and lower LT4 dose groups was 1·3% (95% CI -0·3-3·0) for the total body weight–based exposure definition and 0·7% (95% CI -0·7-2·0) for the lean body weight–based definition. Spline analyses indicate increasing cardiovascular risk with TBW-based LT4 dosing, whereas for LBW-based dosing, risk appeared to increase modestly at doses between approximately 1·3–3·0mcg/kg/day (Appendix pages 202-203).

**Figure 3 Kaplan-Meier Survival Curve for Cardiovascular Events, Bone Outcomes and All-Cause Mortality According to Levothyroxine Dose Based on Total Body Weight. (A) Cardiovascular events, (B) Bone outcomes, (C) All-cause mortality.**

The red line represents the higher-dose group (>1·0mcg/kg/day), and the black line represents the lower-dose group (≤1·0mcg/kg/day). Shaded areas represent 95% confidence intervals.

*Bone outcomes*

Higher initial doses of LT4 (>1·0mcg/kg/day) were associated with a statistically significant increased risk of bone outcomes compared with lower doses (≤1·0mcg/kg/day). For TBW-based dosing, the aHR was 1·29 (95% CI 1·16–1·43), and for LBW-based dosing, the aHR was 1·16 (95% CI 1·06–1·28) (Figure 2B). Similarly, unweighted event rates were higher in the higher-dose group (13·1% vs 9·1%). Subgroup analyses showed broadly consistent trends across age, sex assigned at birth, and TSH categories, although confidence intervals overlapped and not all estimates reached statistical significance. E-values ranged from 1·49 to 2·79, indicating that a moderately strong unmeasured confounder would be required to fully explain the observed associations (Appendix page 185). Sensitivity analyses produced comparable results (Appendix pages 186-199).

Kaplan–Meier curves demonstrated a higher cumulative incidence of bone events among participants initiated on higher LT4 doses (>1·0mcg/kg/day, TBW) compared with lower doses (≤1·0mcg/kg/day, TBW), although the confidence intervals partly overlapped (Figure 3B). Comparable trends were observed for LBW-based dosing and when split by sex assigned at birth (Appendix pages 204-205). The weighted 10-year absolute risk difference was 3·04% (95% CI 1·58-4·49) for total body weight–based dosing and 1·74% (95% CI 0·51-2·78) for lean body weight–based dosing. Spline analyses suggested increasing bone risk above approximately 1·6mcg/kg/day for TBW-based dosing and above 2·0mcg/kg/day for LBW-based dosing (Appendix pages 206-207).

*Mortality outcomes*

Higher initial LT4 doses were associated with a statistically significant increased risk of all-cause mortality compared with lower doses. Using TBW-based dosing, the aHR was 1·17 (95% CI 1·09–1·25), with a similar association observed using LBW-based dosing (aHR 1·12, 95% CI 1·05–1·19). Unadjusted event rates were consistent (Figure 2C). Subgroup analyses demonstrated broadly consistent patterns across age, sex assigned at birth, and TSH categories, although statistical significance varied between strata. E-values for mortality outcomes ranged from 1·28 to 2·21(Appendix pages 185). Sensitivity analyses produced comparable results (Appendix pages 186-199).

Kaplan–Meier curves indicated a modestly higher cumulative incidence of all-cause mortality among participants initiated on higher LT4 doses (>1·0mcg/kg/day, TBW) compared with lower doses (≤1·0mcg/kg/day, TBW), with overlapping confidence intervals (Figure 3C). Similar trends were observed for LBW-based dosing and when split by sex assigned at birth (Appendix pages 208-209). The weighted 10-year absolute risk difference was 3·43% (95% CI 1·74-5·13) for total body weight–based dosing and 2·17% (95% CI 0·76-3·47) for lean body weight–based dosing. Spline analysis demonstrated a gradual increase in mortality risk with TBW-based LT4 dosing above approximately 1·0 mcg/kg/day (Appendix page 210). For LBW-based dosing, the relationship was less clear, with risk appearing to increase modestly at doses between approximately 1·3–3·0 mcg/kg/day (Appendix page 211).

*Secondary outcomes*

Older adults with hypothyroidism who were initiated on higher LT4 doses had greater healthcare utilisation compared with those started on lower doses. The higher dose group had a statistically significantly higher number of visits compared with the lower dose group (incidence rate ratio 1·09, 95% CI 1·06–1·13). Median visits were 246 (IQR 152–378) in the lower dose group and 272 (IQR 164–422) in the higher dose group.

At study end, abnormal TSH levels (<0·4 or >4·0 mU/L) were recorded in 27% (4,166/15,430) of participants in the low-dose group and 30% (4,470/14,899) of the high-dose group. Within the low-dose group, 22% (917/4,166) had low, and 78% (3,249/4,166) had elevated TSH values, compared with 43% (1,935/4,470) and 57% (2,535/4,470), respectively, in the high-dose group. While the overall prevalence of abnormal TSH did not differ between groups, the distribution of low versus high TSH values differed between dosing groups, with low TSH observed more frequently among participants initiated on higher LT4 doses.

## Discussion

In this emulated target trial of adults aged over 50 years with primary hypothyroidism, initiation of LT4 at doses greater than 1·0mcg/kg/day, whether based on TBW or LBW, was associated with higher risks of cardiovascular events, bone outcomes, and all-cause mortality. Notably, stratified analyses suggested that the potential harm of higher initial LT4 doses was more pronounced in adults 65 years and over, whereas outcomes in the 50-64 years age group

were largely non-significant, highlighting age-related differences in sensitivity to LT4 dosing. Participants initiated on higher LT4 doses also had a greater number of primary care consultations, suggesting an association between higher initial dosing and increased healthcare utilisation. Together, these findings indicate that higher initial weight-based LT4 dosing in older adults is associated with less favourable long-term outcomes.

For LBW-based dosing, binary analyses comparing doses above versus below 1·0 mcg/kg/day suggested higher risks of cardiovascular events and mortality. In contrast, spline analyses treating LBW as a continuous variable showed hazard ratios mostly near 1, with only a modest increase in risk at doses between approximately 1·3–3·0mcg/kg/day. This pattern may reflect a threshold effect around 1·0mcg/kg/day, which the binary analysis captures, or the limited number of participants at very low or very high LBW doses, which can flatten the spline curve. TBW-based dosing consistently showed increasing risk across the relevant dose range in both binary and spline analyses.

These findings are consistent with previous observational studies linking LT4 overtreatment to increased cardiovascular morbidity and mortality in older adults.[8] For example, prior research has reported higher rates of atrial fibrillation among individuals receiving LT4 doses exceeding 75mcg/day.[25] Similarly, the increased risk of osteoporosis and fractures observed in this study aligns with earlier findings of adverse skeletal outcomes among older adults treated with higher LT4 doses.[26–28] Previous literature also suggests that multiple LT4 dose adjustments may be associated with increased healthcare utilisation, including more frequent monitoring, medication adjustments, and management of complications.[29]

Together, this literature supports that LT4 requirements derived from studies in younger adults, typically approximately at 1·6mcg/kg/day, may not be appropriate for older populations, who have lower lean body mass and increased hormone sensitivity. However, the present study examines the initial prescribed dose rather than achieved biochemical control and therefore complements rather than replaces evidence based on TSH-defined overtreatment.

Target trial emulation seeks to approximate the causal question that would be addressed in a hypothetical RCT. This is achieved by strictly pre-defining eligibility, treatment, and follow-up so that it matches as closely as possible to a target RCT, which minimises the biases usually found in observational studies. To emulate random assignment to treatment arms, we used IPTW.

This study's strengths include its use of a large population-based UK primary care cohort with long-term follow-up and the application of a target trial emulation framework to reduce bias through clearly defined eligibility, treatment strategies, and follow-up.[17] To emulate a randomised trial and improve comparability between treatment groups, IPTW was applied. This method adjusts the contribution of each participant so that baseline characteristics are balanced across groups, effectively creating a 'pseudo-population' whose size reflects the re-weighted contributions of participants. In addition, the evaluation of dosing based on both TBW and LBW provides a novel insight into practical alternatives for safer LT4 prescribing in older adults.

Several limitations should be considered. As with all observational analyses, residual confounding cannot be excluded, particularly confounding by indication, whereby patients prescribed higher initial doses may differ systematically in ways not fully captured by measured covariates. Although IPTW was used to balance observed baseline characteristics, unmeasured factors such as disease severity, frailty, malabsorption, adherence, or clinician prescribing preferences may have influenced both dose selection and outcomes. Race and ethnicity data were not available within THIN, and, therefore, could not be accounted for in the analyses, despite their potential association with healthcare access, treatment patterns, and health outcomes. In addition, the study population was highly selected, comprising approximately 5% of patients identified in the database who met eligibility criteria; this reflects the application of strict inclusion and exclusion criteria required for the emulated target trial design, but may limit generalisability to all LT4 users in routine clinical practice.

Longitudinal data on dose adjustments, adherence, and serial TSH measurements were incomplete, precluding per-protocol analyses. Analysis therefore focused on initial LT4 dose as a proxy for intended treatment strategy, based on exposure at treatment initiation, recognising that subsequent titration commonly occurs in clinical practice. For cardiovascular outcomes, the association observed in the primary analysis was attenuated and no longer

statistically significant when time zero was redefined at the second LT4 prescription. This finding suggests that part of the observed cardiovascular association could reflect early treatment changes, treatment persistence, or residual confounding. In contrast, associations with bone outcomes and all-cause mortality remained broadly consistent across sensitivity analyses.

Estimated lean body weight may be inaccurate in frail older adults, potentially causing misclassification.[30] Excluding patients without recorded weights could introduce selection bias, since weights are more often measured in those with multiple comorbidities or in individuals who are underweight or overweight. Categorising doses using a 1·0 mcg/kg/day threshold may group widely different doses together while separating similar doses across the cut-off, potentially obscuring dose–response relationships. Sensitivity analyses using natural cubic splines to model dose continuously produced similar associations, supporting the robustness of the findings.

Some baseline measurements, including TSH, may have been recorded shortly after LT4 initiation, which could influence the observed associations. TSH and body weight were defined as the closest measurements within three months before or after treatment, reflecting real-world UK primary care practice where some values may be entered electronically post-initiation. Consequently, the lower TSH and higher free T4 observed in the high-dose group may partly reflect early treatment effects rather than true pre-treatment thyroid function. Free T4 and free triiodothyronine were measured inconsistently and therefore not included in weighting. However, in sensitivity analyses restricted to participants with TSH measurements recorded prior to LT4 initiation, effect estimates were materially unchanged, suggesting that inclusion of post-initiation measurements did not substantially bias the findings.

Diagnosis was based on a coded record of primary hypothyroidism, and thyroid-stimulating hormone (TSH) was included as a baseline variable without systematic free T4 or free triiodothyronine measurements, as these were not routinely recorded. This approach captures both overt and subclinical disease. Nevertheless, initial LT4 dosing is generally recommended uniformly for elevated TSH, which aligns with the study's focus on initial dosing strategies. Patients with post-surgical or post-ablative hypothyroidism were excluded to maintain a clinically homogeneous cohort, as these individuals typically require immediate full replacement dosing and follow different treatment pathways.

THIN captures only UK primary care data and does not reliably include information on ethnicity, family history, consultation type, or secondary care. The cause of death was not available, precluding assessment of cause-specific mortality and limiting interpretation of mortality findings. Death represents a competing risk for cardiovascular and bone outcomes, and this limitation should be considered when interpreting the results. Healthcare utilisation was included as a secondary outcome and could also act as a mediator, potentially influencing dose titration and earlier detection of complications. The analyses conduced involved multiple outcomes, exposure definitions, and subgroup comparisons, and certain estimates, particularly in subgroups, may reflect chance rather than true associations. Residual confounding from unmeasured factors, as well as from variation between practices and care pathways, cannot be excluded. Prescribing practices for LT4 may have changed over the study period (2006–2021), and evolving clinical guidance and clinician behaviour could have introduced residual confounding. Diagnostic coding inaccuracies and the restriction to primary care may limit generalisability to other settings or healthcare systems.

These findings highlight potential risks associated with higher initial LT4 dosing in adults over 50 years, particularly in those aged 65 years and over, and underscore the need for careful dose selection and monitoring in this population. The results describe associations observed in routine clinical practice and should not be interpreted as prescriptive dosing recommendations. Clinical decisions should continue to balance symptom control, biochemical targets, comorbidities, and patient preferences.

While weight-based LT4 dosing may provide a more tailored approach, its implementation in routine primary care requires consideration of feasibility. Accurate weight measurement is essential, and calculation tools could be integrated into electronic prescribing systems or clinical decision support prompts to help facilitate safe and practical dosing in everyday practice.

Future research should aim to evaluate alternative LT4 initiation strategies in older adults, ideally through pragmatic randomised trials or prospective cohort studies incorporating adherence monitoring, serial thyroid function testing, cardiovascular and skeletal endpoints, and patient-reported outcomes. Further work is also needed to clarify optimal TSH targets

and the influence of frailty, multimorbidity, and polypharmacy on LT4 requirements in ageing populations.

## Data Sharing

The data that supports the findings of this study are available from THIN, a wholly owned subsidiary of Cegedim SA, which owns the proprietary rights to THIN data. Restrictions apply to the availability of these data, which were used under license for the current study and are not publicly available.

## Declaration of interest

SR has received speaker fees from Meck KGaA and Berlin Chemie AG, makers of Levothyroxine. All other authors declare that there were no conflicts of interest.

## Author contributions

SR developed the study concept. SW, SR, MH, RD, IM, and EDM designed the study. MH conducted all analyses, with guidance from EDM, SR, RD, IM, and SW. MH drafted the manuscript with feedback from SW, SR, RD, IM, and EDM. All authors read and approved the final manuscript and agree to be accountable for all aspects of the work.

## Acknowledgements

This study is funded by the National Institute for Health and Care Research (NIHR) Applied Research Collaboration (ARC) North East and North Cumbria (NENC) (NIHR200173). The views expressed are those of the authors and not necessarily those of the NIHR or the Department of Health and Social Care.

## Abbreviations

| | |
|---|---|
| aHR | Adjusted hazard ratio |
| CI | Confidence interval |
| dm+d | Dictionary of medicines and devices |
| ICD-10 | International Classification of Diseases, Tenth Revision |
| IPTW | Inverse probability of treatment weighting |
| IQR | Interquartile range |

| | | |
|---|---|---|
| LBW | Lean body weight |
| LT4 | Levothyroxine |
| OPCS-4 | Office of Population Censuses and Surveys Classification of Interventions and Procedures, 4th revision |
| RCT | Randomised controlled trial |
| SMD | Standardised mean differences |
| TBW | Total body weight |
| THIN | The Health Improvement Network |
| TSH | Thyroid-stimulating hormone |
| T4 | Thyroxine |
| UK | United Kingdom |

**References:**


1. Cooper DS, Biondi B. Subclinical thyroid disease. Lancet 2012;379(9821):1142–1154; doi: 10.1016/S0140-6736(11)60276-6.

2. Canaris GJ, Manowitz NR, Mayor G, et al. The Colorado thyroid disease prevalence study. Arch Intern Med 2000;160(4):526–534; doi: 10.1001/archinte.160.4.526.

3. Vadiveloo T, Donnan PT, Murphy MJ, et al. Age- and gender-specific TSH reference intervals in people with no obvious thyroid disease in Tayside, Scotland: the Thyroid Epidemiology, Audit, and Research Study (TEARS). J Clin Endocrinol Metab 2013;98(3):1147–1153; doi: 10.1210/jc.2012-3191.

4. Surks MI, Hollowell JG. Age-specific distribution of serum thyrotropin and antithyroid antibodies in the US population: implications for the prevalence of subclinical hypothyroidism. J Clin Endocrinol Metab 2007;92(12):4575–4582; doi: 10.1210/jc.2007-1499.

5. Fournier J, Barret L, Khouri C, et al. The evidence base of the 10 most prescribed drugs in England, France, and the United States: a scoping review. Journal of Clinical Epidemiology 2024;174:111478; doi: 10.1016/j.jclinepi.2024.111478.

6. Leng O, Razvi S. Hypothyroidism in the older population. Thyroid Res 2019;12:2; doi: 10.1186/s13044-019-0063-3.

7. Centanni M, Duntas L, Feldt-Rasmussen U, et al. ETA guidelines for the use of levothyroxine sodium preparations in monotherapy to optimize the treatment of hypothyroidism. European Thyroid Journal 2025;14(4); doi: 10.1530/ETJ-25-0123.

8. Feldt-Rasmussen U, Effraimidis G, Bliddal S, et al. Risks of suboptimal and excessive thyroid hormone replacement across ages. J Endocrinol Invest 2024;47(5):1083–1090; doi: 10.1007/s40618-023-02229-7.

9. Thayakaran R, Adderley NJ, Sainsbury C, et al. Thyroid replacement therapy, thyroid stimulating hormone concentrations, and long term health outcomes in patients with hypothyroidism: longitudinal study. BMJ 2019;366:l4892; doi: 10.1136/bmj.l4892.

10. Santini F, Pinchera A, Marsili A, et al. Lean body mass is a major determinant of levothyroxine dosage in the treatment of thyroid diseases. J Clin Endocrinol Metab 2005;90(1):124–127; doi: 10.1210/jc.2004-1306.

11. National Institute for Health and Care Excellence (NICE). Thyroid disease: assessment and management (NG145). 2019. Available from: https://www.nice.org.uk/guidance/ng145 [Last accessed: 25/06/2026].

12. Jonklaas J, Bianco AC, Bauer AJ, et al. Guidelines for the treatment of hypothyroidism: prepared by the american thyroid association task force on thyroid hormone replacement. Thyroid 2014;24(12):1670–1751; doi: 10.1089/thy.2014.0028.

13. Holley M, Razvi S, Farooq MS, et al. Cardiovascular and bone health outcomes in older people with subclinical hypothyroidism treated with levothyroxine: a systematic review and meta-analysis. Syst Rev 2024;13(1):123; doi: 10.1186/s13643-024-02548-7.

14. Holley M, Razvi S, Maxwell I, et al. Assessing the Cardiovascular Effects of Levothyroxine Use in an Ageing UK Population with Subclinical Hypothyroidism: Emulated Target Trial (ACEL-UK-ETT). Thyroid; 2026.; doi: 10.1177/10507256261426576.

15. Blak B, Thompson M, Dattani H, et al. Generalisability of The Health Improvement Network (THIN) database: demographics, chronic disease prevalence and mortality rates. Informatics in primary care 2011;19:251–5; doi: 10.14236/jhi.v19i4.820.

16. Holley M, Razvi S, Maxwell I, et al. Assessing the Cardiovascular Effects of Levothyroxine Use in an Ageing United Kingdom Population (ACEL-UK): Cohort Study. The Journal of Clinical Endocrinology & Metabolism 2025;dgaf208; doi: 10.1210/clinem/dgaf208.

17. Hernán MA, Robins JM. Using Big Data to Emulate a Target Trial When a Randomized Trial Is Not Available. American Journal of Epidemiology 2016;183(8):758–764; doi: 10.1093/aje/kwv254.

18. Hernán MA, Wang W, Leaf DE. Target Trial Emulation: A Framework for Causal Inference From Observational Data. JAMA 2022;328(24):2446–2447; doi: 10.1001/jama.2022.21383.

19. Yao XI, Wang X, Speicher PJ, et al. Reporting and Guidelines in Propensity Score Analysis: A Systematic Review of Cancer and Cancer Surgical Studies. J Natl Cancer Inst 2017;109(8):djw323; doi: 10.1093/jnci/djw323.

20. Boer P. Estimated lean body mass as an index for normalization of body fluid volumes in humans. American Journal of Physiology-Renal Physiology 1984; doi: 10.1152/ajprenal.1984.247.4.F632.

21. Flynn RW, Bonellie SR, Jung RT, et al. Serum thyroid-stimulating hormone concentration and morbidity from cardiovascular disease and fractures in patients on long-term thyroxine therapy. J Clin Endocrinol Metab 2010;95(1):186–193; doi: 10.1210/jc.2009-1625.

22. Vita R, Saraceno G, Trimarchi F, et al. Switching levothyroxine from the tablet to the oral solution formulation corrects the impaired absorption of levothyroxine induced by proton-pump inhibitors. J Clin Endocrinol Metab 2014;99(12):4481–4486; doi: 10.1210/jc.2014-2684.

23. Denz R, Klaaßen-Mielke R, Timmesfeld N. A comparison of different methods to adjust survival curves for confounders. Statistics in Medicine 2023;42(10):1461–1479; doi: 10.1002/sim.9681.

24. R Core Team. R: A Language and Environment for Statistical Computing. 2024.

25. Gong IY, Atzema CL, Lega IC, et al. Levothyroxine dose and risk of atrial fibrillation: A nested case-control study. American Heart Journal 2021; doi: 10.1016/j.ahj.2020.09.016.

26. Ko Y-J, Kim JY, Lee J, et al. Levothyroxine dose and fracture risk according to the osteoporosis status in elderly women. J Prev Med Public Health 2014;47(1):36–46; doi: 10.3961/jpmph.2014.47.1.36.

27. Leese GP, Flynn RV. Levothyroxine dose and fractures in older adults. BMJ 2011;342:d2250; doi: 10.1136/bmj.d2250.

28. Turner MR, Camacho X, Fischer HD, et al. Levothyroxine dose and risk of fractures in older adults: nested case-control study. BMJ 2011;342:d2238; doi: 10.1136/bmj.d2238.

29. Ernst FR, Barr P, Elmor R, et al. The Economic Impact of Levothyroxine Dose Adjustments: the CONTROL HE Study. Clin Drug Investig 2017;37(1):71–83; doi: 10.1007/s40261-016-0462-3.

30. Mitchell SJ, Kirkpatrick CMJ, Le Couteur DG, et al. Estimation of lean body weight in older community-dwelling men. Br J Clin Pharmacol 2010;69(2):118–127; doi: 10.1111/j.1365-2125.2009.03586.x.

table 1

**Table 1 Overview of the target trial and its emulation to estimate the effect of levothyroxine (LT4) dose per body weight on vascular, bone, and mortality outcomes.[13,14]**

| Protocol Component | Target Trial (ideal RCT) | Trial Emulation using observational data |
|---|---|---|
| Eligibility criteria | • Adults aged >50 years on January 1, 2006.<br>• Diagnosis of primary hypothyroidism.<br>• Prescribed LT4. | Similar as target trial, plus:<br>• ≥6 months of follow-up data available.<br>• LT4 therapy initiated during the observation period.<br>• Baseline TSH and body weight recorded within three months before or after LT4 initiation.<br>• No history of secondary hypothyroidism (pituitary disease), other thyroid disorders, or prior thyroid treatment.<br>Each outcome measure will be analysed separately using bespoke patient cohorts. Additional criteria for these are:<br>• Cardiovascular cohort: no history of pre-existing cardiovascular disease relevant to the outcome<br>• Bone health cohort: no history of bone disease relevant to the outcome |
| Treatment strategies | 1. LT4 initial dose >1mcg/kg/day.<br>2. LT4 initial dose ≤1mcg/kg/day. | Same as the target trial. |
| Assignment procedures | Random assignment to treatment strategy at baseline. | Randomisation emulated using inverse probability of treatment weighting (IPTW) based on baseline covariates. |
| Follow-up period | Begins at randomisation; ends at outcome event, death, or ten years. | Begins at LT4 initiation; ends at outcome event, death, administrative censoring, or ten years. |
| Outcome | Vascular, bone, or mortality event within ten years. | Each outcome will be analysed separately:<br>1. Cardiovascular disease<br>2. Bone health<br>3. All-cause mortality |
| Causal contrasts of interest | Intention-to-treat principle<br>estimating the effect of treatment assignment. | Effect of initial treatment dose assignment. |
| Identifying assumptions | Randomisation ensures exchangeability, positivity, and consistency. | Exchangeability assumed conditional on measured covariates; positivity checked via propensity score overlap and extreme weight truncation; consistency assumed with treatment defined by initial LT4 dose category. |
| Analysis plan | Cox proportional hazards models (for vascular, bone, and mortality outcomes). | Same as target trial, weighted by IPTW. |

Table 2

**Table 2 Baseline characteristics of the study population by levothyroxine dose group based on total body weight, after weighting.**

| Characteristic | Cardiovascular cohort | | Bone cohort | | Mortality cohort | |
|---|---|---|---|---|---|---|
| | Low Dose (Weighted) | High Dose (Weighted) | Low Dose (Weighted) | High Dose (Weighted) | Low Dose (Weighted) | High Dose (Weighted) |
| **N** | **15,430** | **14,899** | **14,939** | **14,609** | **18,990** | **18,329** |
| **Sex, n (%)** | | | | | | |
| Male | 3,420 (22%) | 3,177 (21%) | 3,955 (26%) | 3,733 (26%) | 4,731 (25%) | 4,390 (24%) |
| Female | 12,010 (78%) | 11,722 (79%) | 10,984 (74%) | 10,876 (74%) | 14,259 (75%) | 13,939 (76%) |
| **Smoking status, n (%)** | | | | | | |
| Smoker | 6,687 (43%) | 6,483 (44%) | 6,756 (45%) | 6,613 (45%) | 8,575 (45%) | 8,307 (45%) |
| Non-smoker | 8,743 (57%) | 8,416 (56%) | 8,183 (55%) | 7,996 (55%) | 10,415 (55%) | 10,022 (55%) |
| **Proton pump inhibitor use, n (%)** | | | | | | |
| Yes | 5,262 (34%) | 5,053 (34%) | 5,420 (36%) | 5,228 (36%) | 7,173 (38%) | 6,876 (38%) |
| No | 10,168 (66%) | 9,846 (66%) | 9,519 (64%) | 9,381 (64%) | 11,817 (62%) | 11,453 (62%) |
| **Hormone therapy, n (%)** | | | | | | |
| Yes | 213 (1%) | 205 (1%) | 198 (1%) | 191 (1%) | 259 (1%) | 247 (1%) |
| No | 15,217 (99%) | 14,694 (99%) | 14,741 (99%) | 14,418 (99%) | 18,731 (99%) | 18,082 (99%) |
| **Statin use, n (%)** | | | | | | |
| Yes | 6,244 (40%) | 6,190 (42%) | 7,099 (48%) | 7,061 (48%) | 9,188 (48%) | 9,035 (49%) |
| No | 9,186 (60%) | 8,709 (58%) | 7,840 (52%) | 7,548 (52%) | 9,802 (52%) | 9,294 (51%) |
| **Chronic pulmonary disease, n (%)** | | | | | | |
| Yes | 2,266 (15%) | 2,153 (14%) | 2,333 (16%) | 2,270 (16%) | 3,154 (17%) | 2,981 (16%) |
| No | 13,164 (85%) | 12,746 (86%) | 12,606 (84%) | 12,339 (84%) | 15,836 (83%) | 15,348 (84%) |
| **Hypertension, n (%)** | | | | | | |
| Yes | 8,431 (55%) | 8,155 (55%) | 8,908 (60%) | 8,695 (60%) | 11,619 (61%) | 11,225 (61%) |
| No | 6,999 (45%) | 6,744 (45%) | 6,031 (40%) | 5,914 (40%) | 7,371 (39%) | 7,104 (39%) |
| **Asthma, n (%)** | | | | | | |
| Yes | 1,548 (10%) | 1,463 (10%) | 1,522 (10%) | 1,489 (10%) | 2,051 (11%) | 1,964 (11%) |
| No | 13,882 (90%) | 13,436 (90%) | 13,417 (90%) | 13,120 (90%) | 16,939 (89%) | 16,365 (89%) |
| **Dyslipidaemia, n (%)** | | | | | | |
| Yes | 1,864 (12%) | 1,801 (12%) | 2,240 (15%) | 2,211 (15%) | 2,942 (15%) | 2,791 (15%) |
| No | 13,566 (88%) | 13,098 (88%) | 12,699 (85%) | 12,398 (85%) | 16,048 (85%) | 15,538 (85%) |
| **Diabetes, n (%)** | | | | | | |
| Yes | 342 (2%) | 417 (3%) | 383 (3%) | 501 (3%) | 521 (3%) | 606 (3%) |
| No | 15,088 (98%) | 14,482 (97%) | 14,556 (97%) | 14,108 (97%) | 18,469 (97%) | 17,723 (97%) |
| **Cancer, n (%)** | | | | | | |

| Characteristic | Cardiovascular cohort | | Bone cohort | | Mortality cohort | |
|---|---|---|---|---|---|---|
| | Low Dose (Weighted) | High Dose (Weighted) | Low Dose (Weighted) | High Dose (Weighted) | Low Dose (Weighted) | High Dose (Weighted) |
| Yes | 776 (5%) | 667 (4%) | 827 (6%) | 683 (5%) | 1,044 (5%) | 913 (5%) |
| No | 14,654 (95%) | 14,232 (96%) | 14,112 (94%) | 13,926 (95%) | 17,946 (95%) | 17,416 (95%) |
| **Continuous Variables** | | | | | | |
| Age (years), mean (SD) | 67 (10) | 67 (10) | 68 (10) | 68 (10) | 68 (10) | 68 (10) |
| Weight (kg), mean (SD) | 79 (19) | 75 (16) | 80 (19) | 75 (16) | 79 (19) | 75 (16) |
| BMI (kg/m²), mean (SD) | 30 (6) | 28 (6) | 29.6 (6.3) | 28.2 (5.5) | 29.6 (6.4) | 28.1 (5.6) |
| TSH (μIU/mL), median [IQR] | 6.9 [5.2, 9.8] | 5.7 [2.6, 10.2] | 6.9 [5.2, 10.4] | 5.1 [5.4, 10.4] | 6.9 [5.2, 10.5] | 4.5 [2.0,8.0] |
| fT4 (pmol/L), mean (SD) | 13.1 (2.8) | 15.2 (3.6) | 13.1 (2.7) | 15.1 (3.6) | 13.1 (2.8) | 15.3 (3.6) |
| Levothyroxine dose per weight (mcg/day/kg), median [IQR] | 0.5 [0.3, 0.7] | 1.4 [1.2, 1.6] | 0.5 [0.3, 0.7] | 1.4 [1.2, 1.6] | 0.5 [0.3, 0.7] | 1.4 [1.2, 1.6] |

LT4, levothyroxine; FT4, free thyroxine; TSH, thyroid-stimulating hormone; SD, standard deviation; IQR, interquartile range (25th–75th percentile). Low dose defined as LT4 dose <1·0 mcg/kg/day; high dose defined as LT4 dose ≥1·0 mcg/kg/day. FT4 and TSH were defined as the closest available measurements within 3 months before or after treatment initiation, selecting the value nearest to time zero.

Figure 1

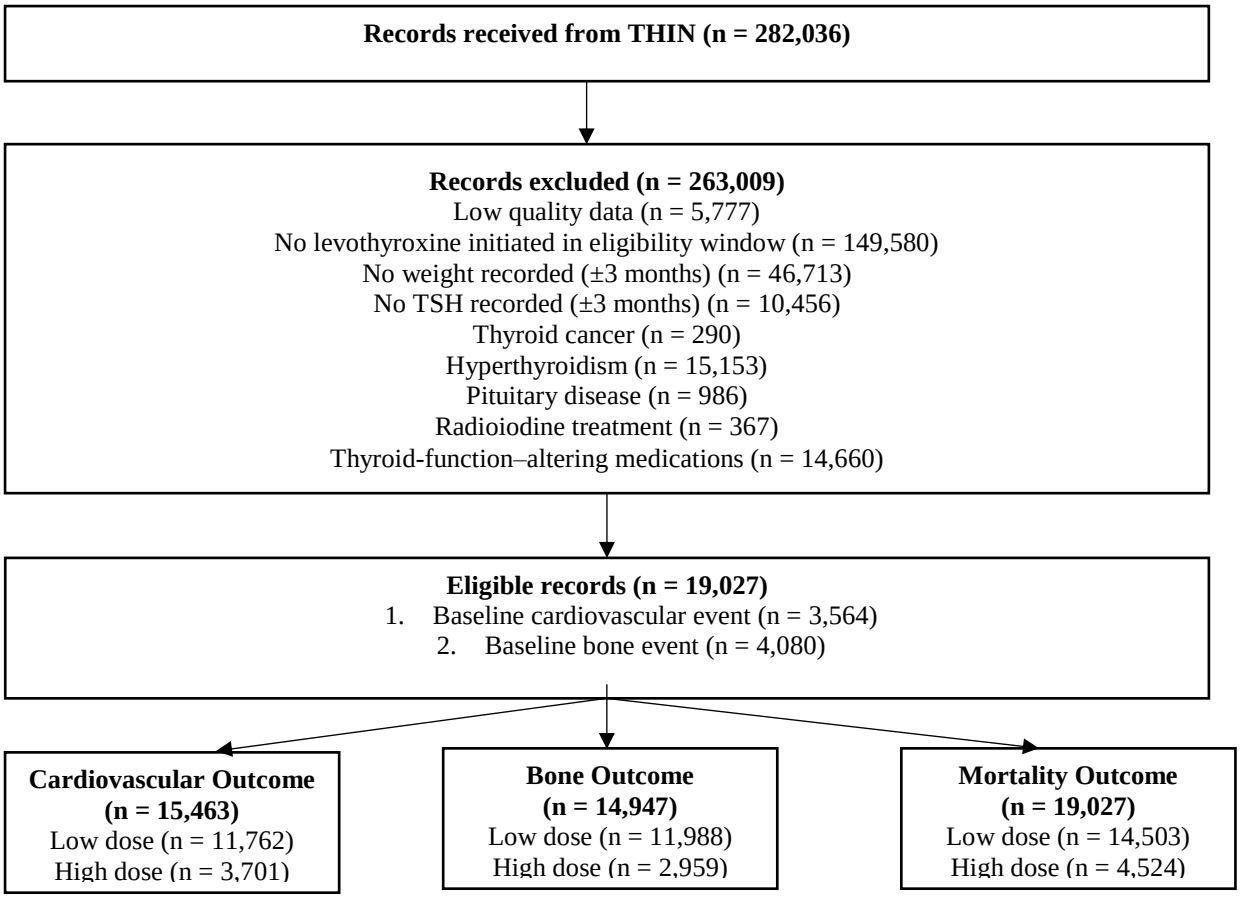
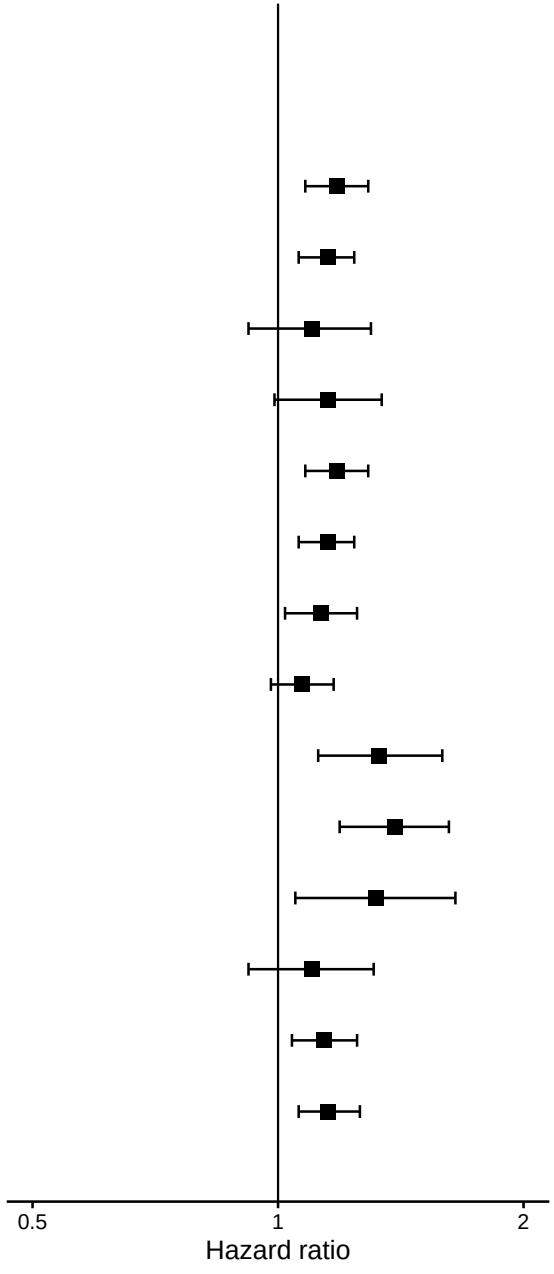

Records received from THIN (n = 282,036)
Records excluded (n = 263,009)
Low quality data (n = 5,777)
No levothyroxine initiated in eligibility window (n = 149,580)
No weight recorded (±3 months) (n = 46,713)
No TSH recorded (±3 months) (n = 10,456)
Thyroid cancer (n = 290)
Hyperthyroidism (n = 15,153)
Pituitary disease (n = 986)
Radioiodine treatment (n = 367)
Thyroid-function–altering medications (n = 14,660)
Eligible records (n = 19,027)
1. Baseline cardiovascular event (n = 3,564)
2. Baseline bone event (n = 4,080)
Cardiovascular Outcome (n = 15,463)
Low dose (n = 11,762)
High dose (n = 3,701)
Bone Outcome (n = 14,947)
Low dose (n = 11,988)
High dose (n = 2,959)
Mortality Outcome (n = 19,027)
Low dose (n = 14,503)
High dose (n = 4,524)

Figure 2

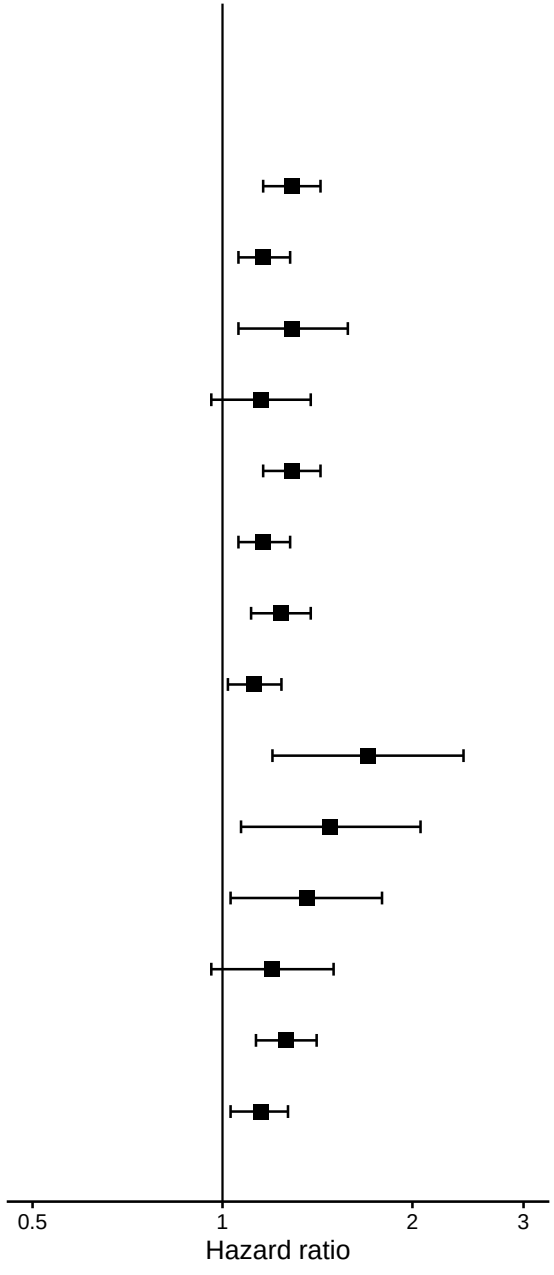
**A. Cardiovascular events**

| Subgroup | | Low dose Events/patients | High dose Events/patients | Hazard ratio (95% CI) | p value |
|---|---|---|---|---|---|
| **Overall** | Total body weight | 2079/11762 | 728/3701 | 1.18 (1.08−1.29) | <0.001 |
| | Lean body weight | 1458/8025 | 1291/7129 | 1.15 (1.06−1.24) | <0.001 |
| **Age 51−65 years** | Total body weight | 562/5421 | 180/1674 | 1.10 (0.92−1.30) | 0.295 |
| | Lean body weight | 395/3669 | 336/3299 | 1.15 (0.99−1.34) | 0.073 |
| **Age >=65 years** | Total body weight | 1517/6341 | 548/2027 | 1.18 (1.08−1.29) | <0.001 |
| | Lean body weight | 1063/4356 | 955/3830 | 1.15 (1.06−1.24) | <0.001 |
| **Female** | Total body weight | 1417/8889 | 564/3154 | 1.13 (1.02−1.25) | 0.015 |
| | Lean body weight | 873/5456 | 1064/6357 | 1.07 (0.98−1.17) | 0.141 |
| **Male** | Total body weight | 662/2873 | 164/547 | 1.33 (1.12−1.59) | 0.001 |
| | Lean body weight | 585/2569 | 227/772 | 1.39 (1.19−1.62) | <0.001 |
| **High TSH** | Total body weight | 450/3071 | 100/558 | 1.32 (1.05−1.65) | 0.016 |
| | Lean body weight | 317/2077 | 216/1456 | 1.10 (0.92−1.31) | 0.319 |
| **Low TSH** | Total body weight | 1629/8691 | 628/3143 | 1.14 (1.04−1.25) | 0.007 |
| | Lean body weight | 1141/5948 | 1075/5673 | 1.15 (1.06−1.26) | 0.001 |



## B. Bone health events

| Subgroup | | Low dose Events/patients | High dose Events/patients | Hazard ratio (95% CI) | p value |
|---|---|---|---|---|---|
| **Overall** | Total body weight | 1089/11988 | 387/2959 | 1.29 (1.16−1.43) | <0.001 |
| | Lean body weight | 730/8606 | 724/6047 | 1.16 (1.06−1.28) | 0.002 |
| **Age 51−65 years** | Total body weight | 320/5153 | 112/1257 | 1.29 (1.06−1.58) | 0.011 |
| | Lean body weight | 207/3574 | 222/2719 | 1.15 (0.96−1.38) | 0.131 |
| **Age >=65 years** | Total body weight | 769/6835 | 275/1702 | 1.29 (1.16−1.43) | <0.001 |
| | Lean body weight | 523/5032 | 502/3328 | 1.16 (1.06−1.28) | 0.002 |
| **Female** | Total body weight | 972/8547 | 361/2448 | 1.24 (1.11−1.38) | <0.001 |
| | Lean body weight | 622/5476 | 692/5314 | 1.12 (1.02−1.24) | 0.021 |
| **Male** | Total body weight | 117/3441 | 26/511 | 1.70 (1.20−2.41) | 0.003 |
| | Lean body weight | 108/3130 | 32/733 | 1.48 (1.07−2.06) | 0.018 |
| **High TSH** | Total body weight | 258/3356 | 58/552 | 1.36 (1.03−1.79) | 0.032 |
| | Lean body weight | 163/2355 | 146/1457 | 1.20 (0.96−1.50) | 0.104 |
| **Low TSH** | Total body weight | 831/8632 | 329/2407 | 1.26 (1.13−1.41) | <0.001 |
| | Lean body weight | 567/6251 | 578/4590 | 1.15 (1.03−1.27) | 0.012 |



## C. All−cause mortality

| Subgroup | | Low dose Events/patients | High dose Events/patients | Hazard ratio (95% CI) | p value |
|---|---|---|---|---|---|
| **Overall** | Total body weight | 2532/14503 | 836/4524 | 1.17 (1.09−1.25) | <0.001 |
| | Lean body weight | 1879/10151 | 1374/8511 | 1.12 (1.05−1.19) | <0.001 |
| **Age 51−65 years** | Total body weight | 321/5991 | 96/1822 | 1.15 (0.96−1.39) | 0.135 |
| | Lean body weight | 233/4070 | 171/3609 | 1.05 (0.90−1.24) | 0.524 |
| **Age >=65 years** | Total body weight | 2211/8512 | 740/2702 | 1.17 (1.09−1.25) | <0.001 |
| | Lean body weight | 1646/6081 | 1203/4902 | 1.12 (1.05−1.19) | <0.001 |
| **Female** | Total body weight | 1605/10526 | 648/3770 | 1.19 (1.10−1.29) | <0.001 |
| | Lean body weight | 1053/6566 | 1124/7467 | 1.10 (1.02−1.18) | 0.010 |
| **Male** | Total body weight | 927/3977 | 188/754 | 1.14 (0.99−1.31) | 0.072 |
| | Lean body weight | 826/3585 | 250/1044 | 1.17 (1.04−1.33) | 0.011 |
| **High TSH** | Total body weight | 616/3704 | 152/670 | 1.43 (1.22−1.69) | <0.001 |
| | Lean body weight | 467/2594 | 267/1669 | 1.09 (0.95−1.26) | 0.211 |
| **Low TSH** | Total body weight | 1916/10799 | 684/3854 | 1.10 (1.02−1.19) | 0.013 |
| | Lean body weight | 1412/7557 | 1107/6842 | 1.12 (1.05−1.20) | 0.001 |

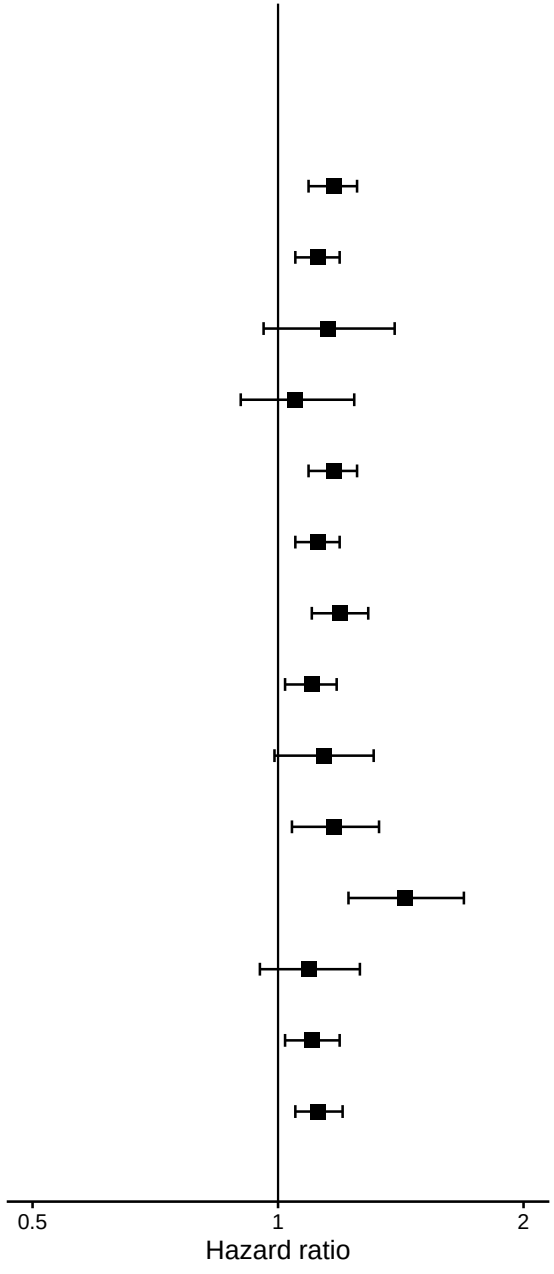
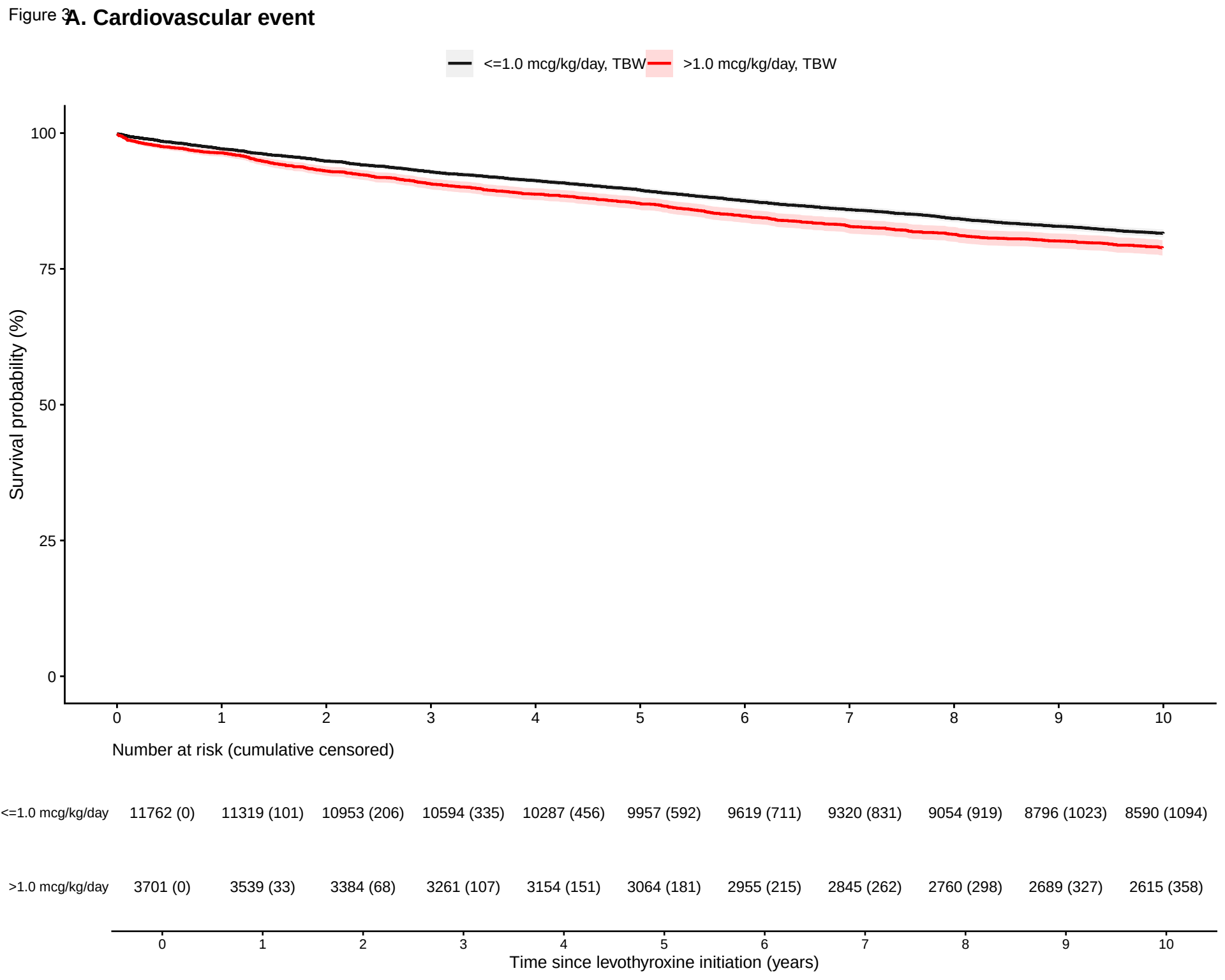
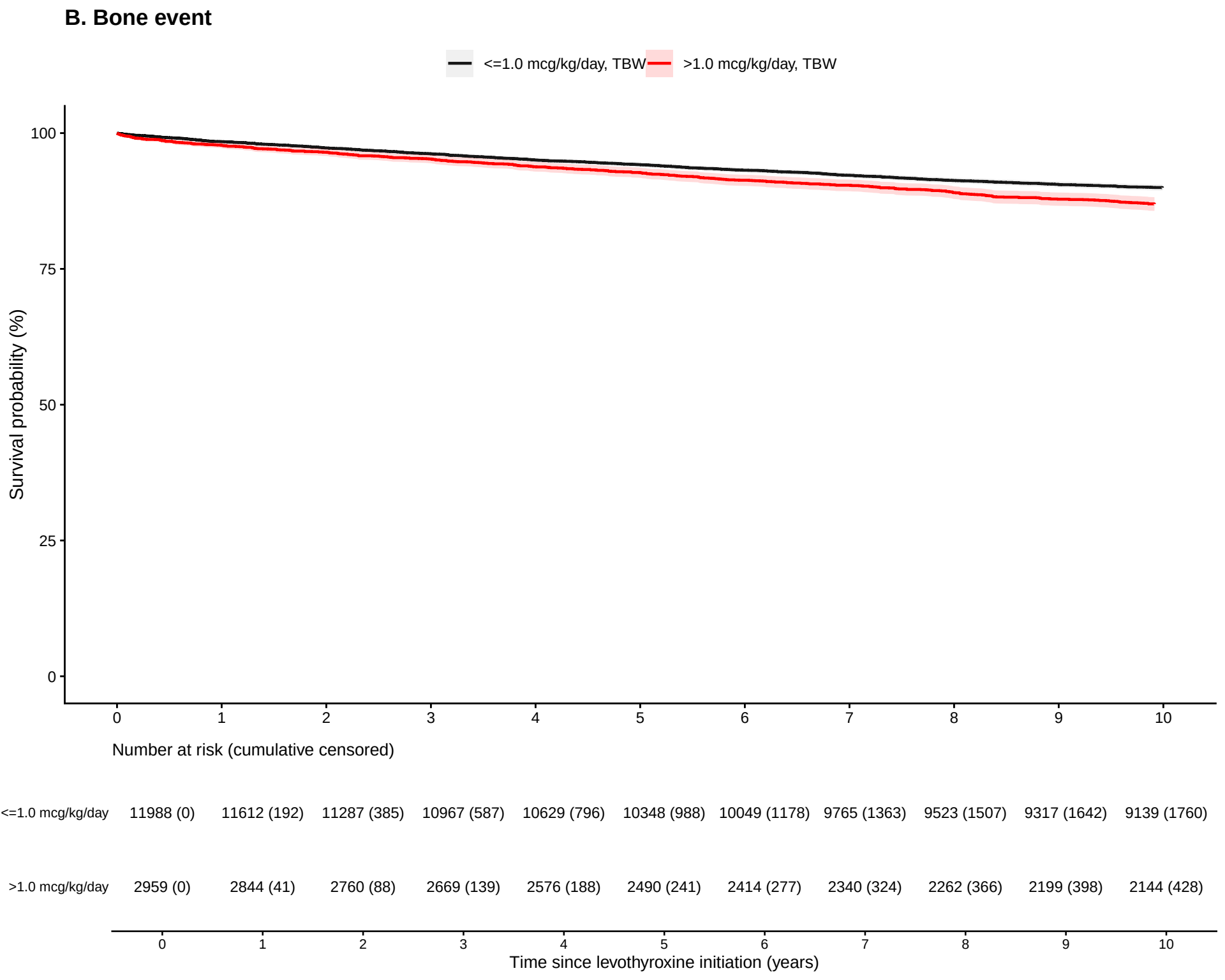


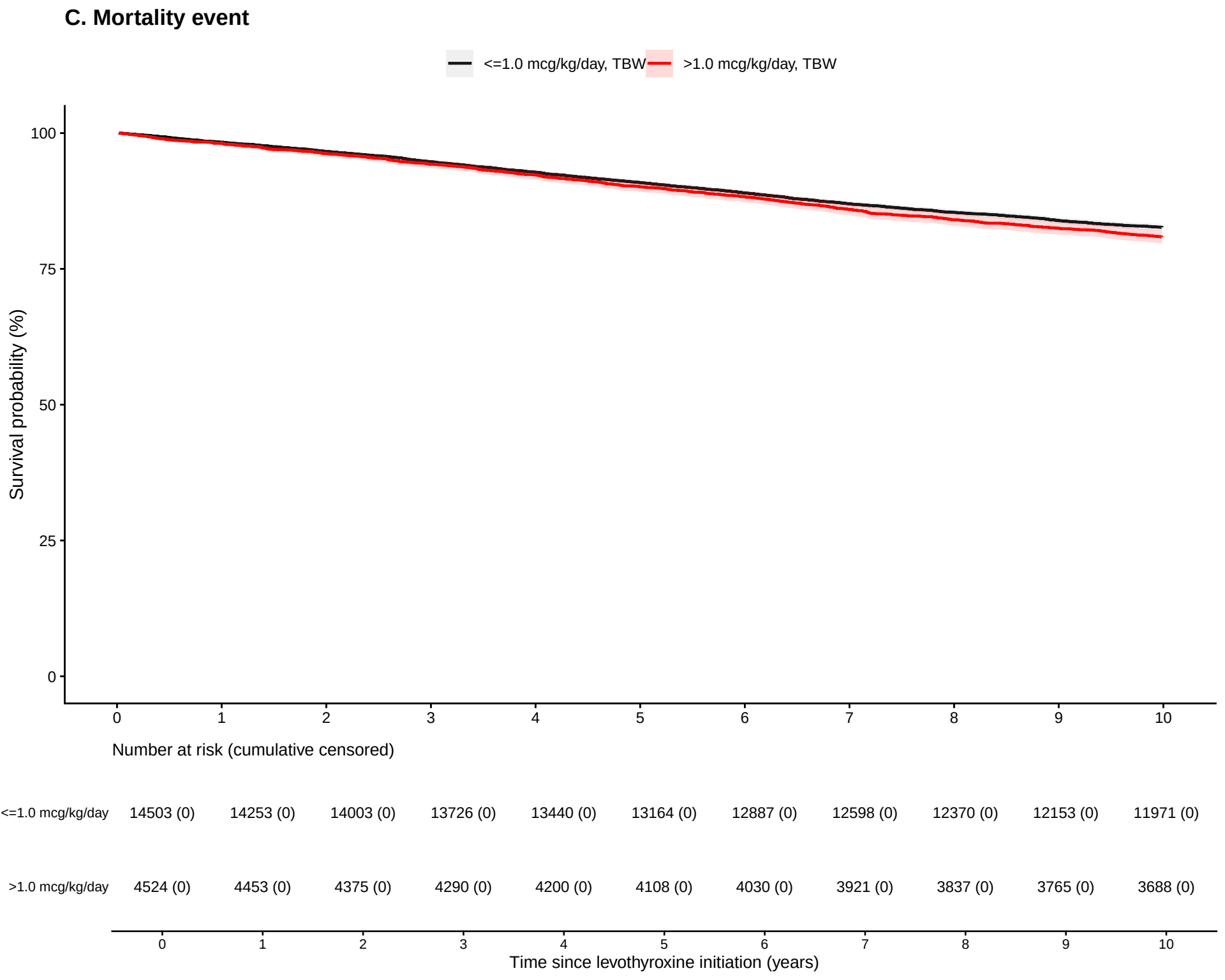
Figure 3


A. Cardiovascular event
<=1.0 mcg/kg/day, TBW
>1.0 mcg/kg/day, TBW
Survival probability (%)
100
75
50
25
0
0
1
2
3
4
5
6
7
8
9
10
Number at risk (cumulative censored)
<=1.0 mcg/kg/day
11762 (0)
11319 (101)
10953 (206)
10594 (335)
10287 (456)
9957 (592)
9619 (711)
9320 (831)
9054 (919)
8796 (1023)
8590 (1094)
>1.0 mcg/kg/day
3701 (0)
3539 (33)
3384 (68)
3261 (107)
3154 (151)
3064 (181)
2955 (215)
2845 (262)
2760 (298)
2689 (327)
2615 (358)
0
1
2
3
4
5
6
7
8
9
10
Time since levothyroxine initiation (years)

## B. Bone event


<=1.0 mcg/kg/day, TBW
>1.0 mcg/kg/day, TBW
Survival probability (%)
100
75
50
25
0
0
1
2
3
4
5
6
7
8
9
10
Number at risk (cumulative censored)
<=1.0 mcg/kg/day 11988 (0) 11612 (192) 11287 (385) 10967 (587) 10629 (796) 10348 (988) 10049 (1178) 9765 (1363) 9523 (1507) 9317 (1642) 9139 (1760)
>1.0 mcg/kg/day 2959 (0) 2844 (41) 2760 (88) 2669 (139) 2576 (188) 2490 (241) 2414 (277) 2340 (324) 2262 (366) 2199 (398) 2144 (428)
0
1
2
3
4
5
6
7
8
9
10
Time since levothyroxine initiation (years)

## C. Mortality event



appendix final

# ICMJE DISCLOSURE FORM

**Date:_4th June 2026_____**
**Your Name: Salman Razvi_____**

**Manuscript Title: Weight-based levothyroxine dosing and long-term cardiovascular, skeletal, and mortality outcomes in older adults: an emulated target trial using UK primary care data**
**Manuscript number (if known): thelancetprimarycare-D-26-00085R1**

**In the interest of transparency, we ask you to disclose all relationships/activities/interests listed below that are related to the content of your manuscript. "Related" means any relation with for-profit or not-for-profit third parties whose interests may be affected by the content of the manuscript. Disclosure represents a commitment to transparency and does not necessarily indicate a bias. If you are in doubt about whether to list a relationship/activity/interest, it is preferable that you do so.**

**The following questions apply to the author's relationships/activities/interests as they relate to the current manuscript only.**

**The author's relationships/activities/interests should be defined broadly. For example, if your manuscript pertains to the epidemiology of hypertension, you should declare all relationships with manufacturers of antihypertensive medication, even if that medication is not mentioned in the manuscript.**

**In item #1 below, report all support for the work reported in this manuscript without time limit. For all other items, the time frame for disclosure is the past 36 months.**

| | | Name all entities with whom you have this relationship or indicate none (add rows as needed) | Specifications/Comments (e.g., if payments were made to you or to your institution) |
|---|---|---|---|
| **Time frame: Since the initial planning of the work** | | | |
| 1 | All support for the present manuscript (e.g., funding, provision of study materials, medical writing, article processing charges, etc.) **No time limit for this item.** | __X_None | |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |
| **Time frame: past 36 months** | | | |
| 2 | Grants or contracts from any entity (if not indicated in item #1 above). | Merck KGaA | Investigator initiated trial assessing the feasibility of using T3 in patients with heart failure. Paid to Newcastle University. |
| | | | |
| | | | |
| 3 | Royalties or licenses | __X__None | |
| | | | |

| | | | |
|---|---|---|---|
| 4 | Consulting fees | __X__None | |
| | | | |
| | | | |
| 5 | Payment or honoraria for lectures, presentations, speakers bureaus, manuscript writing or educational events | Berlin Chemie Pvt. Ltd. | Honorarium for speaking at conference |
| | | Merck KGaA | Honorarium for speaking at conference |
| | | | |
| 6 | Payment for expert testimony | __X__None | |
| | | | |
| | | | |
| 7 | Support for attending meetings and/or travel | __X__None | |
| | | | |
| | | | |
| 8 | Patents planned, issued or pending | _X___None | |
| | | | |
| | | | |
| 9 | Participation on a Data Safety Monitoring Board or Advisory Board | _X__None | |
| | | | |
| | | | |
| 10 | Leadership or fiduciary role in other board, society, committee or advocacy group, paid or unpaid | _X__None | |
| | | | |
| | | | |
| 11 | Stock or stock options | __X_None | |
| | | | |
| | | | |
| 12 | Receipt of equipment, materials, drugs, medical writing, gifts or other services | _X__None | |
| | | | |
| | | | |
| 13 | Other financial or non-financial interests | _X__None | |
| | | | |
| | | | |

**Please place an "X" next to the following statement to indicate your agreement:**

**_X_ I certify that I have answered every question and have not altered the wording of any of the questions on this form.**

# ICMJE DISCLOSURE FORM

**Date: 4th June 2026**
**Your Name: Mia Holley**
**Manuscript Title: Weight-based levothyroxine dosing and long-term cardiovascular, skeletal, and mortality outcomes in older adults: an emulated target trial using UK primary care data**
**Manuscript number (if known): thelancetprimarycare-D-26-00085R1**

**In the interest of transparency, we ask you to disclose all relationships/activities/interests listed below that are related to the content of your manuscript. "Related" means any relation with for-profit or not-for-profit third parties whose interests may be affected by the content of the manuscript. Disclosure represents a commitment to transparency and does not necessarily indicate a bias. If you are in doubt about whether to list a relationship/activity/interest, it is preferable that you do so.**

**The following questions apply to the author's relationships/activities/interests as they relate to the current manuscript only.**

**The author's relationships/activities/interests should be defined broadly. For example, if your manuscript pertains to the epidemiology of hypertension, you should declare all relationships with manufacturers of antihypertensive medication, even if that medication is not mentioned in the manuscript.**

**In item #1 below, report all support for the work reported in this manuscript without time limit. For all other items, the time frame for disclosure is the past 36 months.**

| | | Name all entities with whom you have this relationship or indicate none (add rows as needed) | Specifications/Comments (e.g., if payments were made to you or to your institution) |
|---|---|---|---|
| **Time frame: Since the initial planning of the work** | | | |
| 1 | All support for the present manuscript (e.g., funding, provision of study materials, medical writing, article processing charges, etc.) **No time limit for this item.** | __X__None | |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |
| **Time frame: past 36 months** | | | |
| 2 | Grants or contracts from any entity (if not indicated in item #1 above). | __X__None | |
| | | | |
| | | | |
| 3 | Royalties or licenses | __X__None | |
| | | | |
| | | | |
| 4 | Consulting fees | __X__None | |
| | | | |

| | | | |
|---|---|---|---|
| 5 | Payment or honoraria for lectures, presentations, speakers bureaus, manuscript writing or educational events | __X__None | |
| | | | |
| | | | |
| 6 | Payment for expert testimony | _X___None | |
| | | | |
| | | | |
| 7 | Support for attending meetings and/or travel | __X__None | |
| | | | |
| | | | |
| 8 | Patents planned, issued or pending | __X__None | |
| | | | |
| | | | |
| 9 | Participation on a Data Safety Monitoring Board or Advisory Board | __X__None | |
| | | | |
| | | | |
| 10 | Leadership or fiduciary role in other board, society, committee or advocacy group, paid or unpaid | __X__None | |
| | | | |
| | | | |
| 11 | Stock or stock options | __X__None | |
| | | | |
| | | | |
| 12 | Receipt of equipment, materials, drugs, medical writing, gifts or other services | __X__None | |
| | | | |
| | | | |
| 13 | Other financial or non-financial interests | __X__None | |
| | | | |
| | | | |

**Please place an "X" next to the following statement to indicate your agreement:**

**__X_ I certify that I have answered every question and have not altered the wording of any of the questions on this form.**

Dear Editors,
We thank the reviewers and editors for their comments on our manuscript. We have revised the manuscript accordingly and provide point-by-point responses below:

Reviewer 4: I appreciate that the authors performed the analyses I suggested including to exclude patients with TSH checked after the date of the prescription, and introducing a lag. These results are presented in the supplements and the authors choose to discuss these as being "similar" to the main results. However, I disagree with this presentation for two reasons.
First, including post-treatment TSH is a methodological flaw and so table 2 and the relevant analysis about high and low TSH at "baseline" should be reported as pretreatment if it is going to be relevant to clinical decision-making. If it is in fact the case that those in the high dose group continue to have lower pre-dose TSH levels than those in the low dose group (we don't know because you do not provide an updated table)
[Editorial note: please provide the updated table 2 in the appendix reported as pretreatment for comparison]

**We agree that pretreatment measurements are clinically most relevant. We have therefore added a supplementary baseline table restricted to participants with pretreatment TSH measurements only (Appendix p19), allowing direct comparison with the primary analysis.**

2) I would start to wonder about missing data, and that perhaps your high dose group is actually being titrated later compared to the others - which makes the lag analysis even more important. And the lag analysis is not "similar". Figure S4 should not be in supplements. In particular, with a lag introduced, the findings for cardiovascular disease are no longer statistically significant.
[Editorial note: please make sure this is clearly acknowledged in the main text, results, and the potential effects of missing data further discussed in the limitations]

**To address this concern, we now explicitly state in the Results and Discussion that the association between higher LT4 dose and cardiovascular outcomes was attenuated and no longer statistically significant after introducing a lag period. We agree that this finding raises the possibility that early events and healthcare utilisation patterns may partly explain the primary association and have discussed this as a limitation.**

3) Finally, I am concerned about downplaying the missing information for these subgroups - I doubt that it would affect the main findings, but it is a variable that is missing and that usually correlates with important attributes of patient's health care access and response. The authors response notes that race/ethnicity data was not available and said that this was addressed in the discussion on limitations, but I do not see it where is should most appropriately be discussed, in the section on residual confounding. Instead, it is folded in to a later list on data not available in THIN (line 514), all of which (cause of death, health care utilization and practice variation) is discussed in

detail except this. My suggestion is to add it to the paragraph below, which is where I think the response to reviewers suggested it was going to go: "Several limitations should be considered. As with all observational analyses, residual 469 confounding cannot be excluded, particularly confounding by indication, whereby patients 470 prescribed higher initial doses may differ systematically in ways not fully captured by 471 measured covariates. Although inverse probability of treatment weighting was used to 472 balance observed baseline characteristics, unmeasured factors such as disease severity, 473 frailty, malabsorption, adherence, or clinician prescribing preferences may have influenced 474 both dose selection and outcomes." [Editorial note: please expand the limitations as suggested]

**Following the reviewer's suggestion, the Limitations section now discusses the absence of ethnicity data within the paragraph on residual confounding.**

1) Reviewer 5: Thank you for addressing my previous comments. My outstanding comments relate to: 1. your defence of relying on TSH measurement as the main indicator for hypothyroidism, using ref 11 as justification for this. Ref 11 is a CKS that is not available to readers outside the UK so I couldn't check your statement that 2 separate measurements of TSH can be used to diagnose hypothyroidism. The related and more comprehensive NICE Guidance does not suggest this (https://www.nice.org.uk/guidance/ng145/chapter/Recommendations#managing-primary-hypothyroidism). It specifically states that if TSH is raised a T4 measurement should be conducted on the same sample - maybe this isn't available in the UK but it seems strange that it would feature in a NICE Guidance if not. [Editorial note: please make sure this is clarified further, see also Editorial point below]

**We thank the reviewer for highlighting this. We have removed the implication that TSH alone is sufficient for diagnosis under NICE guidance. The manuscript now clarifies that NICE recommends assessment of fT4 when TSH is elevated, but that fT4 was incompletely recorded in THIN and therefore could not be incorporated consistently into cohort definition or adjustment.**

2) My comment on the limitation of not examining the modification of treatment over the follow up period, being the norm in clinical practice was addressed by a sensitivity analysis redefining time zero at the second LT4 prescription. I don't think this really addresses the issue though you do state the following as a limitation: 'subsequent titration commonly occurs in clinical practice', which does partially address it.
The use of the term intention-to-treat in this context seems a bit misleading to me given that, unlike in randomisation you haven't controlled for unknown confounders. [Editorial note: please avoid using this terminology and simply define the populations assessed]

**In response to this comment, we have removed the term “intention-to-treat” and now define the cohort according to LT4 dose at initiation.**

Your abstract conclusion is appropriately cautious given the limitations.

Reviewer 6: Thank you for the thoughtful revision. The responses on E-values, multiple testing, the mediator framing, and the LBW spline-versus-binary all read well. However, there are still a few small but essential details missing.

1) The new ARDs are helpful, but please specify the time horizon and provide confidence intervals. Adding them to the abstract would also be consistent with the rest of the revision. [Editorial note: please note this is required for all estimates presented across the paper]

**We now report 10-year absolute risk differences with corresponding 95% confidence intervals throughout the manuscript, including in the abstract.**

2) My suggestion was to consider including calendar year in the propensity score model. The Discussion now acknowledges temporal trends, which is appreciated, but the analytic suggestion was not taken up. Could you either incorporate it or briefly explain why not? It can also be put into the regression model as a covariate to check it doesn't change, ie, doesn't need to be in PS exposure model, just the outcome model.

**We performed a sensitivity analysis additionally adjusting for calendar year of LT4 initiation. Inclusion of calendar year did not materially alter the estimated associations (Appendix p17), therefore the primary model is unchanged.**

3) The text states overlap was assessed, but a figure or summary statistics (PS density plot, or distribution of stabilised weights and effective sample size) would let me/readers see this directly. Given the 76/24 split and 1st/99th percentile truncation, this is important for this paper, as well as aligning with TARGET. [Editorial note: please make sure this added]

**Accordingly, propensity score overlap plots have been added to the Supplementary Material.**

EDITORS' POINTS: Please note that the list of editorial comments has been tailored to your manuscript, so we expect a response to all of the points.
1) Please note that we are a gold open access journal. You can find information about our article processing charges (APCs) at https://www.thelancet.com/open-access. Please note that as part of our launch, we are reducing the APC to 50% for papers submitted by 30 June, 2026, which would apply if your manuscript is finally considered suitable for acceptance.
2) Generally, please ensure that all results are presented in a balanced, unbiased way, including clearly stating where findings are non-significant. Note that we often edit manuscripts post-acceptance to ensure the unbiased presentation of results, but it is preferable for authors to ensure this is the case at revision stage.

**The manuscript has been reviewed throughout to ensure balanced and unbiased reporting of all findings. Non-significant results are now explicitly stated where**

**relevant, including the attenuation of the cardiovascular association in the lag-adjusted analysis.**

3) Your revised paper should have fewer than 3500 words and a maximum of 30 references. The abstract should be structured (background, methods, findings, interpretation, funding) and should be less than 300 words long. However, we support full reporting of study methods and results, so please ensure that all changes are made in response to the editorial points even if this means that the paper goes over these word limits (Introduction and Discussion sections can be made more concise if necessary).

**Where necessary, text has been condensed while ensuring that all requested details are retained.**

4) Please check with your co-authors, and confirm that all names are spelt correctly, and affiliation details for each author are listed correctly (including department, institution, city, state [if applicable], and country). We cannot guarantee that we will be able to correct names and affiliations after publication of your article.

**We have confirmed with all co-authors that author names and affiliations are correct.**

5) Please ensure that you include full first names for all authors and please supply (after author names on the title page) one preferred degree per author and indicate in the authorship if any authors are full professors (ie, Prof X, MD). Please note that we do not indicate assistant or associate professors.

**Full first names and degree designations have been provided for all authors.**

6) If accepted, only 5 non-text items (figures or tables) can be accommodated; additional material can be provided for in an appendix (see below for formatting instructions).

**The figures and tables have been revised accordingly.**

7) Table 1. This table is informative, but please make sure the Methods elaborate on the emulated trial decisions as requested below.

**Additional detail on the target trial specification and emulation decisions has now been incorporated into the Methods section.**

8) Table 2. The table should include all selected patients, not only those in the cardiovascular cohort. Please amend with a pooled cohort across outcomes, or alternatively, present columns for the three key cohorts (ie, one per outcome; this is our preference). Please provide full categories (ie, male vs female, if sex was considered a binary variable; smoker vs non smoker; hypertension (yes/no); etc). Please present a

breakdown of the ethnic/racial backgrounds of the participants in the baseline characteristics table if possible. If not available as you suggested, please add a footnote to the table stating why this was not available.

**Table 2 has been revised to present baseline characteristics separately for the cardiovascular, skeletal, and mortality cohorts, as requested. Full categorical breakdowns are now provided for all binary and categorical variables. A separate ethnicity category has now been included in Table 2 to provide a breakdown of participants' ethnic backgrounds, including the proportion of missing ethnicity data.**

9) Figure 1. Patient selection. Please amend with a pooled cohort across outcomes, or alternatively, present different flows for the three key cohorts analysed (ie, one per outcome). Probably the later would be most helpful to readers given how the figures are presented.

**Figure 1 has been revised to present the cardiovascular, skeletal, and mortality cohorts separately**.

10) Figure 2. Forest plot – the x-axis should be on a log scale if risk ratios, odd ratios, or prevalence are presented. Please also provide number of events and number of patients in each group. Examples for how to present forest plots/these data can be found here https://www.thelancet.com/journals/lanonc/article/PIIS1470-2045(22)00793-8/fulltext and here https://www.thelancet.com/journals/lanonc/article/PIIS1470-2045(22)00283-2/fulltext

**We have revised all forest plots accordingly.**

11) Figures 3, 4, 5. Please label these as panels (A, B, C) of the same figure 3, since they all present survival for the different outcomes in the 2 selected dose groups. For these kaplan-Meier curves: Please add number at risk and the number of patients censored in each group for each time point on your K-M curves. Please ensure both are cumulative and please use the format “number at risk (number censored)”. Please change the scale on the y axis to 0–100.

**Figures 3–5 have been combined and are now presented as panels A–C of Figure 3. Number at risk and cumulative number censored are now displayed for each group. The y-axis has also been rescaled.**

12) FIGURES: Please ensure that you provide your figures in editable formats. For trial profiles (clinical trials) and study selection diagrams (systematic reviews and meta-analyses), Word files (.doc or .docx) made of boxes with editable text are preferred. For any statistical images such as histograms, survival or time-to-event curves, line graphs, scatter graphs, and forest plots you should provide editable vector files (ie, the original artwork generated by the statistical package used to make the image, typically by using “Export” or “Print to file” commands); our preferred formats for these files are .eps, .pdf,

or .ai. Photographic images must be provided at a minimum of 300 dpi at 107 mm wide. We cannot guarantee accurate reproduction of images without these files. For more information, see download.thelancet.com/flatcontentassets/authors/artwork-guidelines.pdf

**All figures have been supplied in editable vector formats in accordance with journal requirements.**

13) TABLES: Tables should be supplied in a separate Word file (not Excel or fdf/pdf). Each row of data should be in a separate line. Please ensure that rows and columns are not tabbed; data should be entered in cell form.

**All tables have been supplied as editable Word tables in accordance with journal requirements.**

14) REQUIRED CHECKLISTS: please confirm that your study conforms to the relevant guidelines by completing and returning the checklist: TARGET for reporting of observational studies emulating a target trial — https://www.equator-network.org/reporting-guidelines/22462/ and https://target-guideline.org/.

**The TARGET checklist has been completed and submitted in the Supplementary Material.**

15) We adhere to the CRISP reporting guidelines to improve the reporting of research relevant to primary care. Please complete and return the CRISP checklist and implement the necessary changes to the main text of the paper to make sure your paper is reported in a way that is accessible to the diverse primary care community. You can find the checklist at https://www.equator-network.org/reporting-guidelines/improving-the-reporting-of-primary-care-research-consensus-reporting-items-for-studies-in-primary-care-the-crisp-statement/

**The CRISP checklist has been completed and submitted.**

16) The Lancet Primary Care endorses the SAGER guidelines for reporting of sex and gender information in study design, data analyses, results and interpretation of findings: https://www.equator-network.org/reporting-guidelines/sager-guidelines/. For all study types, we encourage correct use of the terms sex (when reporting biological factors) and gender (when reporting identity, psychosocial, or cultural factors). Where possible, please report the sex and/or gender of study participants, and describe the methods used to determine sex and gender. Separate reporting of data by demographic variables, such as age and sex, facilitates pooling of data for subgroups across studies and should be routine, unless inappropriate. Please also discuss the influence or association of variables, such as sex and/or gender, on your findings, where appropriate, and the limitations of the data.

**The manuscript has been revised in accordance with SAGER guidance. Sex was derived from routinely recorded electronic health records and this has been clarified in the Methods.**

17) I wonder if you might consider the possibility of presenting duplicate versions of current figures 3, 4, 5 in the appendix, with the data disaggregated by sex. We have recently launched an initiative to encourage authors to consider the SAGER guidelines when reporting their data (ultimately the goal is to encourage considerations of sex/gender at the planning stage of trials and prospective cohort studies), and given the topic, there is a possibility that this might reveal interesting differences (but would also be interesting if it does not).

**Sex-disaggregated Kaplan–Meier curves have been added to the appendix (Appendix p31) and are referenced in the manuscript.**

18) The Lancet Group has developed guidance on reporting race and ethnicity to advance equity (https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(24)01081-X/fulltext). We encourage researchers to include people from minoritised racial or ethnic populations as participants, and to plan to report and analyse data by race, ethnicity, or both. When reporting on race or ethnicity in the Methods, please define the definitions, categories, or conceptual framework used and how these were assigned. Please also state if the terms used for data collection were not your own (eg, they are the existing terms used in an established database). If race or ethnicity data were not collected, please briefly give the reason/s. In the Discussion, please discuss the representativeness of the study population. Race and ethnicity are sociocultural constructs, not biological traits. Thus, when discussing data in relation to race or ethnicity, please discuss the potential limitations of these data and the possible role of unmeasured confounders, such as socioeconomic and other structural drivers. Consider a strengths-based approach (how findings might promote health) rather than a deficit discourse focused on problems. For research focused on groups that have historically been marginalised, how was community engagement prioritised in the research process? (Eg for Indigenous health papers, a possible tool is the CONSIDER statement.)

**Race and ethnicity data were not routinely available within THIN and therefore could not be analysed. We have clarified this in the Methods and expanded the Discussion to acknowledge the implications for representativeness.**

1. SUMMARY (ABSTRACT): must include:
* Background: Ending with a sentence indicating the aim of this study. Briefly justify here the use of a target emulation trial approach.
* Methods: Identify the study as an attempt to emulate a trial and described the observational data used for emulation. Please add the cutoff dates applied to the datasets used.
* Methods: summarise the specified target trial and key assumptions, justifying the treatment dose cutoffs used (>1·0 vs ≤1·0 mcg/kg/day)

* Methods: A brief summary of the main patient/cohort characteristics (including age limit, disease status, treatments, etc).
* Findings: please summarise key characteristics, report the sex/gender and race/ethnicity breakdown of the participants if possible.
* Findings: Efficacy data for the three key endpoints considered.
* Interpretation: please do not just restate your findings. What do they mean, clinically? What are their implications in primary care?

**The Abstract has been revised in accordance with the editor's recommendations.**

1. INTRODUCTION: Please ensure the introduction states the rationale for using an emulated trial design. Summarise the casual question of the emulated trial. State the objectives of the study related to benefits and harms.

**The Introduction has been revised to further justify the use of a target trial emulation design, define the causal question being addressed, and clarify the study objectives relating to both potential benefits and harms of weight-based levothyroxine dosing.**

2. MAIN TEXT: The main text should be structured as follows: Introduction, Methods, Results, Discussion. Note that subheadings should only be used within the Methods section. The Methods section should be structured as: Study design and participants, Procedures, Outcomes, Statistical analysis, Role of the funding source (see below).
3. Methods, Study design and participants. If applicable, please ensure that the following items are included
* Describe the dataset used: original purpose, type, the geographic locations, setting and time-period. If relevant, describe how the data were linked or pooled. Provide the eligibility criteria, and the sources and methods of selection of participants.

* Please follow the TARGET checklist to ensure you include all relevant details comparing a target trial specification and the target trial emulation, extending the information provided in table 1.
* Move here the information given on ethical considerations please. IRB or ethical approval, including the name of the ethical review broad and approval number. If multiple approvals were obtained this multi-site trial, please name the main approving board and approval number here and add all others to the acknowledgments or appendix.
* Nature of consent sought. Please specify if you had a consent waiver given you used an available dataset.
* A link to the protocol, if you had one. And any protocol amendments that occurred to affecting trial recruitment or conduct during the study. Please provide the date of amendment and the protocol number. Any other amendments should be described in the appendix and cited here.

**The Methods have been restructured in accordance with these requirements. Additional information has been included regarding the THIN database,**

**participant selection, target trial specification, ethical approval, consent waiver, and protocol availability (with a link provided to institutional website).**

1. METHODS: Procedures. Please ensure the following items are included:
* For the dataset:
  * State what data were collected, by whom, and when.
  * State the method used to collect sex/gender data (eg, self-report, genetic testing). If self-reported by study participants, indicate what options were provided.
  * State the method used to collect race/ethnicity data (eg, self-report, census, registry data).
* For each variable of interest, give sources of data and details of methods of assessment (measurement).
* Describe here the assignment procedures, follow-up.
* Define all outcomes, exposures, predictors, potential confounders, and effect modifiers.
* Describe any efforts to address potential sources of bias.
* If AI technology was used in any part of the study methods, please describe: (1) how it was used (with sufficient detail to enable replication); (2) the name of the tool; (3) the version of the tool; and (4) what prompts were used, if applicable.

**The Procedures section has been revised in accordance.**

1. Methods: Outcomes. This section is ok as it is.

2. Methods: Statistical analysis. As applicable, please ensure the following items are included:
* Describe here the outcomes, casual contrasts, identifying assumptions and data analysis plan for the emulated trial.
* Definitions of population assessed for outcomes. Please follow the statistical reviewer advice against using ITT in this content given lack of randomisation.
* Statistical methods for analysis of the outcomes.
* Describe all statistical methods, including those used to control for confounding
* Describe any methods used to examine subgroups and interactions
* Describe how missing data were addressed
* Describe any sensitivity analyses, etc.
* Software package (and version number) used for statistical analyses.

**The Statistical Analysis section has been expanded to describe the causal contrast of interest, identifying assumptions, confounding control methods, subgroup analyses, handling of missing data, sensitivity analyses, and statistical software used. The term "intention-to-treat" has been removed throughout.**

1. The Lancet journals are very supportive of protocol-based research and encourage authors to post the protocol document on a publicly accessible website; a margin link to the website will then be put in the paper. Would you like to do this for your protocol? If so, please provide the link in the Methods section of the main text. Please note that if

you do wish to do this, the weblink must be permanent. Alternatively, the protocol can be included in your appendix if you wish - please indicate this in your responses.

**The study protocol has been made publicly available and the link has been added to the Methods section.**

2. Lancet style is to have a 'Role of the funding source' at the end of the methods. The following points need to be addressed in the "Role of the funding source" statement:
* The role of the funders in the study design.
* The role of the funders in the collection, analysis, or interpretation of the data.
* The role of the funders in the writing of the report.
* If the funding source had no role then this should be stated (ie, "The funder of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report").

**A 'Role of the Funding Source' statement has been added to the end of the Methods section.**

1. RESULTS: General reporting guidance
* Lancet style is to provide p values to two significant figures, unless p<0.0001 (note number of decimal places). The exception is certain genetics studies, in which smaller p values can be reported exactly using scientific notation.
* It is Lancet style to give actual numbers (numerator and denominator) together with all percentages, throughout the text and in tables etc. Percentages should be rounded to whole numbers unless the study population is larger than 1000 individuals.
1. Results: Please explicitly state the number of patients included in analyses. For any patients/individuals excluded, please provide reasons why.
    * Please give median follow-up (and IQR) for the analyses presented.
    * Please ensure number of patients/individuals who had an event are stated for each treatment group (eg, progression event or death for PFS, death for OS, etc.)
    * If appropriate, for the main outcome measures, please include a result for each group, plus a point estimate (eg, RR, HR) with a measure of precision (eg, 95% CI) for the absolute difference between groups, in both the abstract Findings section and the main Results section of the paper.
    * Data presented with descriptive statistics should be given as mean (SD), or as median (IQR) if not following a normal distribution, to indicate their variability.
    * All estimates should be provided with 95% CIs (not SEM) to indicate their precision. Please also give 95% CI for hazard ratios/odds ratios.
    * Give unadjusted estimates and, if applicable, confounder adjusted estimates and their precisions (95% CIs), making sure that the confounders adjusted for are defined clearly and justified in the manuscript.
    * For continuous variables, as appropriate, please ensure that category boundaries are categorised
    * If appropriate, please consider translated estimates of relative risk into absolute risk for a meaningful time period.

* Please add 95% CI to all time-to-event data and other data derived from Kaplan-Meier analyses.

**The Results section has been revised to comply with these recommendations.**

2. DISCUSSION: Please ensure that the section on limitations of the study includes:
* The fact that the study includes a highly selected group of patients that might not represent levothyroxine users and comprised 5% of patients identified in the database
* The fact that many patients with subclinical hypothyroidism were included because only THS levels were considered in the patient selection (no free T3/T4) and discuss how this makes the study reflect real world practice

**The limitations section has been expanded to acknowledge that the study cohort represents a selected subset of levothyroxine users and may not be fully representative of all individuals receiving treatment. We have also discussed the inclusion of patients identified using TSH measurements alone and how this reflects routine primary care practice.**

3. STATEMENTS AND FORMS: Author contribution statement
* All authors are required to provide a signed author contribution statement form, available to download at https://www.thelancet.com/pb-assets/Lancet/authors/tlpc-author-signatures.pdf(completed forms should be uploaded with your revised manuscript if not already submitted).
* The Lancet's journals require that more than one author has verified the underlying data in all research articles. Please state which author(s) have accessed and verified the data, and which author(s) were responsible for the decision to submit the manuscript. For research articles that are the result of an academic and commercial partnership, at least one of the authors named as having accessed and verified data must be from the academic team. The contributors statement should state who those authors are.
* Please note, that if an author is added or deleted from the author list, we need a signed statement from the other authors to agree to the amendment and the current order of the authors in the list (including the author who was removed).

**The Author Contributions statement has been completed.**

1. Conflict-of-interest statements for individual authors:
* All authors should complete and return an ICMJE conflict of interest form, available to download at https://www.thelancet.com/forms?section=icmje-coi
* A conflict of interest exists if authors or their institutions have financial or personal relationships with other people or organisations that could inappropriately influence (bias) their actions. Financial relationships are easily identifiable, but conflicts can also occur because of personal relationships, academic competition, or intellectual passion. A conflict can be actual or potential, and full disclosure to The Editor is the safest course. Failure to disclose conflicts might lead to publication of an Erratum or even to retraction. All submissions to The Lancet Primary Care must include disclosure of all

relationships that could be viewed as presenting a potential conflict of interest (see Lancet 2001; 358: 854-56 and Lancet 2003; 361: 8-9). The Editor may use such information as a basis for editorial decisions, and will publish such disclosures if they are believed to be important to readers in judging the manuscript.
* At the end of the text, under a subheading "Declaration of interest", all authors must disclose any financial and personal relationships with other people or organisations that could inappropriately influence (bias) their work. Examples of financial conflicts include employment, consultancies, stock ownership, honoraria, paid expert testimony, patents or patent applications, and travel grants, all within 3 years of beginning the work submitted. If there are no conflicts of interest, authors should state that none exist ("We declare no competing interests."). Authors should be referred to by their initials in this section. The statement should match the information on the supplied ICMJE forms.
* AI software: Was generative AI software used in your research or in the preparation of your manuscript? If so, we require a statement in the acknowledgement, based on Elsevier's policies: https://www.elsevier.com/about/policies-and-standards/generative-ai-policies-for-journals. Please add such a statement if relevant.

**ICMJE conflict of interest forms have been completed by all authors and the Declaration of Interests section has been updated accordingly. Generative AI was not used in the conduct of the study.**

1. APPENDIX: Please supply the web appendix as a single PDF file, with the pages paginated – when you refer to an item in the appendix, please refer to the page number on which it appears, not the table or section (eg, appendix p 1). Please ensure that all data presented in the appendix is cited in the manuscript.
It is not Lancet policy to edit or style supplementary material for the web; however, this material will be hosted on our website as a pdf conversion of the author-supplied file. Please style your supplementary material as per the guidelines below. Please note that we will be unable to correct any errors in the online appendix following publication; as such, please check carefully when submitting. Please do not include a cover page with details of the paper, as we add a cover page to the appendix file before publication with this information. Required formatting of data in the appendix:
* SI units are required, if appropriate
* Numbers in text and tables should always be provided if % is shown.
* Means should be accompanied by SDs, and medians by interquartile range.
* Exact p values should be provided, unless p<0·0001
* For drug names, recommended international nomenclature (rINN) is required
* References should be in Vancouver style (eg, Smith A, Jones, B, Clements S. Clinical transplantation of tissue-engineered airway. Lancet 2008; 372: 1201-09. Hourigan P. Ankle injuries. In: Sports medicine. Chan D, ed. London: Elsevier, 2008: 230-47.) They should be numbered in order of mention in appendix and numbered separately from references in the full paper
* For figures in the appendix, all images must have a minimum resolution of 300 dpi at a width of 107 mm.

**The appendix has been revised and formatted as a single paginated PDF in accordance with journal requirements. All appendix material is cited within the manuscript where appropriate.**

We thank the reviewers and editors for their time and thoughtful feedback, which have helped improve the manuscript. All changes are highlighted in the revised version.
Kind regards,
Dr Mia Holley
On behalf of all authors

# Weight-based levothyroxine dosing and long-term cardiovascular, skeletal, and mortality outcomes in older adults: an emulated target trial using UK primary care data.

Mia Holley, PhD[1*], Prof Scott Wilkes, PhD[1], Ellen Dorothea Moss, PhD[2], Ian Maxwell, MSc[1], Rosie Dew, PhD[1], Salman Razvi, MD[3].

1. School of Medicine, University of Sunderland, Sunderland, UK
2. Population Health Sciences Institute, Newcastle University, Newcastle upon Tyne, UK
3. Translational and Clinical Research Institute, Newcastle University, Newcastle upon Tyne, UK

* Corresponding author: mia.holley@sunderland.ac.uk

## Research in context

### Evidence before this study

PubMed was searched from January 1, 2000, to October 1, 2025, using combinations of terms for levothyroxine ("levothyroxine", "thyroxine", "LT4"), weight-based dosing ("dose per body weight", "weight-based dose", "mcg/kg", "µg/kg"), hypothyroidism, ageing ("older adults", "elderly", "aged"), and outcomes ("cardiovascular", "fracture", "osteoporosis", "bone", "mortality"). No language restrictions were applied. No studies were identified that examined the associations between initial levothyroxine dose per body weight and cardiovascular, skeletal, or mortality outcomes in adults aged 50 years or older. Existing evidence has focused primarily on the risks of overtreatment, defined by suppressed thyroid-stimulating hormone levels, rather than the prescribed dose per body weight. Current guidance from the United Kingdom National Institute for Health and Care Excellence (NICE) recommends fixed low starting doses of levothyroxine (25-50 mcg0mcg/day) for adults aged 65 years and older, irrespective of body weight. In contrast, younger adults are initiated on weight-based doses. The Baltimore Longitudinal Study of Ageing reported a mean requirement of approximately 1·09 mcgmcg/kg/day to maintain euthyroidism in adults aged 65 years and older, suggesting that lower weight-adjusted levothyroxine doses may be appropriate; however, no large-scale study has directly examined their association with long-term outcomes.

### Added value of this study

Using primary care data from more than 15,000 older adults with hypothyroidism in the United Kingdom, this study emulated a target trial to assess the causal effectassociation of between initial levothyroxine dose per body weight on and cardiovascular, bone, and mortality outcomes. After adjustment for baseline confounding, initiating levothyroxine at doses greater than 1·0 mcg0mcg/kg/day, whether based on total or lean body weight, was associated with significantly higher risks of adverse cardiovascular events, and skeletal events and all-cause mortality.

### Implications of all the available evidence

This evidence suggests that applying standard 1·6 mcg/kg/day formulas dosing may overestimate levothyroxine requirements in adults over 50 years old, while fixed low starting doses may under-treat heavier or leaner older adults. Weight-adjusted, age-sensitive dosing, targeting approximately 1·0 mcg0mcg/kg/day, may be associated with fewer adverse outcomes. could reduce adverse outcomes. These findings support revising clinical guidance to incorporate weight-based levothyroxine prescribing for people aged 65 years and older.

## Abstract

**Background:** Levothyroxine (LT4) is widely prescribed in older adults with hypothyroidism, yet evidence guiding safe initial weight-based dosing is limited. Both under- and over-replacement may adversely affect cardiovascular health, bone integrity, and survival. Given the impracticality of conducting a long-term randomised trial in this population, this study aimed to emulate a target trial to assess associations between initial LT4 dose per body weight and long-term clinical outcomes in older adults, using routinely collected primary care data.

**Methods:** An emulated target trial was conducted using The Health Improvement Network, a UK primary care database, including data from 2006 to 2021. Eligible participants were adults aged >50 years diagnosed with hypothyroidism initiating LT4 during this period. The target trial compared initiation of LT4 at >1·0 versus ≤1·0mcg/kg/day using total and lean body weight definitions. ~~An emulated target trial was conducted to examine associations between initial LT4 dose per body weight and long-term clinical outcomes in older adults. Eligible participants were adults aged >50 years with a recorded diagnosis of hypothyroidism who initiated LT4 therapy between 2006 and 2021 in UK primary care. Data were analysed from The Health Improvement Network (THIN), a UK primary care database. Initial LT4 dose at treatment start was categorised as >1·0 or ≤1·0 mcg/kg/day using total and lean body weight.~~ Inverse probability of treatment weighting adjusted for baseline confounders. Associations with cardiovascular events, bone outcomes, and all-cause mortality were estimated using weighted Cox proportional hazards ~~models~~models, and 10-year absolute risk differences were calculated. Participants were followed for up to 10 years from LT4 initiation until outcome occurrence, death, or study end.

**Findings:** Among 15,463 patients ~~in the final cohort~~(mean age 67 years, 78% female), 11,762 (76%) initiated LT4 at ≤~~1.~~1·~~0 mcg~~0mcg/kg/day and 3,701 (24%) at >~~1.~~1·~~0 mcg~~0mcg/kg/day ~~based on~~(~~-~~total body weight). Unweighted event rates were higher in the higher-dose group for cardiovascular events (19·7% vs 17·7%), bone outcomes (13·1% vs 9·1%), and all-cause mortality (18·5% vs 17·5%). ~~Unweighted event rates showed cardiovascular events in 19.2% of the higher-dose group versus 17.3% of the lower-dose group; bone outcomes in 12.6% versus 9.0%, and all-cause mortality in 20.1% versus 17.7%.~~ Higher LT4 dosing ~~based on total body weight were~~ was associated with increased risk of cardiovascular events (hazard ratio (HR) 1·18, 95% CI 1·08–1·29), bone outcomes (1·29,


Formatted: Font color: Text 1

Formatted: Font color: Text 1

Formatted: Font color: Text 1

Formatted: Font color: Text 1

Formatted: Font color: Text 1

1·16–1·43), and all-cause mortality (1·17, 1·09–1·25). The weighted 10-year absolute risk difference between higher and lower dose groups was 1·3% (95% CI 0·3–3·0) for cardiovascular events, 3·0% (1·6–4·5) for bone outcomes, and 3·4% (1·7–5·1) for all-cause mortality. Similar associations were observed using lean body weight-based dosing.

**Interpretation:** ~~In this observational study of adults over 50 years, initiation of LT4 at doses exceeding 1.0 mcg/kg/day was associated with higher rates of cardiovascular events, bone outcomes, and all-cause mortality.~~ These findings suggest that higher initial weight-based LT4 dosing may be associated with avoidable long-term harm, supporting cautious initiation and titration in older adults in primary care. These findings describe associations and should be interpreted cautiously.

**Funding**: This study was funded by the National Institute for Health and Care Research (NIHR) Applied Research Collaboration (ARC) North East and North Cumbria (NENC) (NIHR200173).

## Introduction

Hypothyroidism is a common chronic endocrine disorder characterised by insufficient thyroid hormone production, particularly thyroxine (T4).[1] Its prevalence increases with age and is notably higher in women, affecting up to 21% of those over 74 years old.[2] This higher prevalence in older adults likely reflects age-related rises in thyroid-stimulating hormone (TSH) levels,[3] a greater susceptibility to autoimmune thyroid disease, and prior treatments that impair thyroid function, such as thyroid surgery and radioactive iodine therapy.[4]

Levothyroxine (LT4), a synthetic form of T4, remains the standard treatment and is among the most widely prescribed medications globally.[5] In the United Kingdom (UK), prescribing rates are highest in people aged over 80 years, where more than 10% receive LT4.[6] The treatment goal is to restore and maintain euthyroidism, typically assessed by TSH normalisation.[7]

Older adults, however, are particularly vulnerable to the effects of suboptimal LT4 dosing. Both under- and over-replacement, indicated by persistently abnormal TSH levels outside the population reference range, have been associated with adverse cardiovascular and bone outcomes.[8,9] Standard weight-based dosing (1·~~6 mcg~~6mcg/kg/day) is derived from studies in

younger adults and may overestimate LT4 needs in older individuals, who have lower lean body mass and heightened peripheral sensitivity to thyroid hormones.[10] Reflecting this, the UK's National Institute for Health and Care Excellence (NICE) recommends weight-based dosing (1·~~6 mcg~~6mcg/kg/day) in adults under 65 years without cardiovascular disease, but substantially lower fixed starting doses (20-5~~0 mcg~~0mcg/day) with titration in those aged 65 years and older or with cardiovascular disease.[11] The American Thyroid Association similarly advises individualised dosing based on both age and weight.[12] In addition, a cautious approach to treatment in older adults is often advocated, particularly where biochemical abnormalities are mild.[13,14] Despite these differences, there remains limited evidence directly examining the relationship between initial LT4 dose per body weight and long-term clinical outcomes in ageing populations.

Long-term randomised trials comparing initial LT4 dosing strategies in older adults are unlikely to be feasible. Observational data structured within a target trial emulation framework therefore provide an opportunity to address this evidence gap ~~examine the association between initial LT4 dosing strategies and long-term clinical outcomes~~ using real-world clinical practice data. By explicitly defining eligibility criteria, treatment strategies, follow-up, and outcomes, target trial emulation can strengthen causal inference from observational data and reduce biases commonly encountered in observational studies.~~Within this framework, the causal question of interest was whether initiating LT4 at higher versus lower weight-based doses is associated with differences in long-term cardiovascular and skeletal outcomes, and mortality in routine clinical practice.~~

~~To address this gap,~~Accordingly, a target trial was emulated using UK primary care data to examine the association between initial LT4 dose per body weight and major clinical outcomes in adults with primary hypothyroidism. Individuals aged over 50 years were included, reflecting evidence from the TEARS study demonstrating age-related increases in TSH in older adults.[3] Participants were stratified into two age groups: 51–64 years, for whom standard weight-based dosing (1·~~6 mcg~~6mcg/kg/day) is generally recommended, and ≥65 years, for whom NICE advocates lower fixed starting doses (25–5~~0 mcg~~0mcg/day).[11] The causal question of interest was whether initiating LT4 at higher versus lower weight-based doses is associated with differences in long-term cardiovascular and skeletal outcomes, and mortality in routine clinical practice, with the objective of estimating the comparative benefits and harms of these dosing strategies in older adults with hypothyroidism.

Initial LT4 dose was expressed per unit of total body weight (TBW) and estimated lean body weight (LBW), allowing comparison across two clinically relevant metrics. Outcomes included (1) cardiovascular events (composite of angina, myocardial infarction, heart failure, stroke, peripheral vascular disease, atrial fibrillation, and revascularisation), (2) bone outcomes (composite of fragility fractures and osteoporosis), and (3) all-cause mortality. Higher versus lower dosing strategies were compared using a threshold of 1·~~0~~ ~~mcg~~0mcg/kg/day, informed by data from the Baltimore Longitudinal Study of Aging, which reported a mean LT4 requirement of 1·09 ~~mcg~~mcg/kg/day in adults aged ≥65 years old.[12] ~~This study aimed to estimate associations between initial LT4 dosing strategies and long-term clinical outcomes in older adults.~~

## Methods

### ~~Data Source~~Study design and participants

Data for this study were obtained from The Health Improvement Network (THIN), a UK primary care database containing anonymised electronic health records from over 850 general practices using the Vision software system, representing approximately 6% of the UK population. THIN includes longitudinal information on demographics, diagnoses, prescriptions, and laboratory test results recorded in routine primary care and is broadly representative of the UK population in terms of age, sex distribution, major disease prevalence, and mortality rates.[15] Data from January 1, 2006, to December 31, 2021, were available for analysis. THIN captures data recorded by general practices and does not include linked secondary care or hospital admission data for this study. The database has been validated and widely used in epidemiological research examining LT4 therapy, thyroid function, cardiovascular disease, fractures, and mortality.[9,16]

~~Patients aged over 50 years with at least one recorded TSH measurement exceeding 4·0 mU/L between January 1, 2006, and January 1, 2022, were extracted from THIN. Data cleaning included removal of duplicate records, verification of measurement units, checks for chronological consistency, and assessment of plausibility for key variables. Baseline TSH values <0·1 or >20 mU/L and body weight <35 or >200 kg were excluded as extreme values. Records with missing baseline information or with inconsistent follow-up data were excluded.~~

~~Study Design~~

This retrospective observational study emulated a target trial to estimate the association between initial LT4 dose and long-term cardiovascular, skeletal, and mortality outcomes in adults with ~~primary~~ hypothyroidism. The design followed the target trial framework proposed by Hernán and Robins,[17] which uses observational data to reproduce key features of a hypothetical randomised controlled trial, including explicit specification of eligibility criteria, treatment strategies, assignment procedures, and follow-up. This study follows both the Strengthening the Reporting of Observational Studies in Epidemiology guidelines for Observational Studies Reporting Propensity Score Analysis and the Transparent Reporting of Observational Studies Emulating a Target Trial checklists (~~Table S1 and S2~~Appendix pages 1-7).[18,19] The primary causal contrast of interest ~~was the intention-to-treat effect, comparing initiation of LT4 at doses >1·0 versus ≤1·0 mcg/kg/day~~compared participants according to the initial prescribed LT4 dose at treatment initiation. Participants were classified according to their initial prescribed dose at treatment initiation~~n (time zero)~~, irrespective of subsequent dose modifications. An overview of the target trial specification and its emulation is presented in Table 1.

~~Eligibility Criteria~~

Patients aged over 50 years with at least one recorded TSH measurement exceeding 4·0 mU/L between January 1, 2006, and January 1, 2022, were extracted from THIN. Eligibility was assessed at the time of LT4 initiation (time zero).

~~*Inclusion Criteria*~~

Participants were eligible if they met all of the following:

- Aged over 50 years at the time of LT4 initiation.
- Recorded diagnosis of primary hypothyroidism.
- Initiated LT4 therapy between January 1, 2006 and December 31, 2021.
- At least one serum TSH measurement recorded within three months before or after LT4 initiation.
- At least one body weight measurement within three months before or after LT4 initiation.

Formatted: Normal

Formatted: Normal

Formatted: Line spacing: 1.5 lines

- At least one height measurement (*lean body mass analyses only*).
- At least six months of follow-up data available after LT4 initiation. Six months of follow-up data refers to the availability of routine primary care records for at least six months after LT4 initiation, ensuring sufficient observation time for outcome events.

~~Exclusion Criteria~~

Participants were excluded if they met any of the following:

- History of thyroid cancer, hyperthyroidism, or pituitary disease.
- Previous thyroid surgery or radioiodine therapy.
- Use of antithyroid drugs, liothyronine, or medications known to substantially affect thyroid function (including lithium and amiodarone) during the study period.
- Pre-existing cardiovascular or bone disease relevant to the outcome under assessment.

Each of the three outcomes (cardiovascular disease, bone health, and all-cause mortality) were analysed separately, using bespoke patient cohorts.

Diagnostic, prescription, and procedure codes used to define inclusion and exclusion criteria were derived from the International Classification of Diseases, Tenth Revision (ICD-10), dictionary of medicines and devices (dm+d), and Office of Population Censuses and Surveys Classification of Interventions and Procedures, 4th revision (OPCS-4) (Appendix page 8~~Table S3~~).

Data cleaning included removal of duplicate records, verification of measurement units, checks for chronological consistency, and assessment of plausibility for key variables. TSH <0·1 or >20mU/L and body weight <35 or >200kg were excluded as extreme values. Records with missing baseline information or with inconsistent follow-up data were excluded.

Formal sample size calculations were not performed, as the study aimed to include the full eligible population defined by the target trial protocol to provide precise observational estimates rather than to test a single prespecified hypothesis.[17]

This study received ethical approval from the University of Sunderland Research Ethics Group (Reference: 011081). Scientific approval for the use of data from THIN was granted by the THIN Scientific Review Committee (Protocol Number 22-003). Individual patient

Formatted: Font: Italic

Formatted: Font: Not Italic

Formatted: Font: Not Bold

consent was waived because anonymised routinely collected electronic health record data were used.

The study protocol was prospectively developed and is available at https://sure.sunderland.ac.uk/id/eprint/19274/. No protocol amendments affecting eligibility criteria, treatment assignment, follow-up procedures, or study conduct occurred during the study period.

**~~Treatment Strategy~~Procedures**

THIN data are collected prospectively during routine primary care consultations by participating UK general practitioners and affiliated healthcare professionals using the Vision electronic health record system. Demographic data, diagnoses, prescriptions, laboratory test results, and consultation records entered during routine clinical care were available for analysis. Data on race and ethnicity were not consistently available within THIN and therefore could not be analysed.

Baseline demographic variables, including age and sex assigned at birth, were obtained from routinely recorded primary care electronic health records within THIN. Clinical covariates, including hypertension, Charlson Comorbidity Index, smoking status, medication use, serum TSH measurements, and body weight, were identified using diagnostic records, prescribing data, and laboratory test entries recorded during routine clinical care. Prescription data were recorded by general practitioners at the point of prescribing, and laboratory results were electronically transferred into the medical record from affiliated laboratories (Appendix pages 9-10). Cardiovascular, bone, and mortality outcomes were identified using ICD-10 and OPCS-4 codes recorded in THIN (Appendix page 11).

To evaluate the impact of initial LT4 dose according to body composition, exposure was defined using two weight-based approaches based on the initial prescribed dose at LT4 initiation:

~~1.~~ Total Body Weight (TBW) dose (mcg/day/kg): initial LT4 dose (mcg/day) divided by TBW (kg).

**Formatted:** Normal (Web)

**Formatted:** Line spacing: 1.5 lines

**Formatted:** Normal

**Formatted:** Line spacing: 1.5 lines

**Formatted:** Normal

**Formatted:** List Paragraph, Bulleted + Level: 1 + Aligned at: 0.25" + Indent at: 0.5"

~~2.•~~ Lean Body Weight (LBW) dose (mcg/day/kg): initial LT4 dose (mcg/day) divided by estimated LBW (kg).

LBW was calculated using the Boer formula,[20] which incorporates sex assigned at birth, height, and weight:

- Males: LBW (kg) = (0·407 × weight (kg)) + (0·267 × height (cm)) − 19·2
- Females: LBW (kg) = (0·252 × weight (kg)) + (0·473 × height (cm)) − 48·3

For each dosing metric, participants were categorised into two groups:

1. **Higher Dose Group**: >1·~~0 mcg~~0mcg/kg/day
2. **Lower Dose Group**: ≤1·~~0 mcg~~0mcg/kg/day

These classifications were applied separately for TBW and LBW, with analyses conducted independently for each approach. ~~Participants were included in only one dosing definition per analysis.~~

TSH and body weight were defined as the closest measurements recorded within three months before or after LT4 initiation. This window was selected to maximise data completeness and reflects routine UK practice.[11] When multiple measurements were available, the value closest to LT4 initiation was used. TSH was used as the primary biochemical marker as it is the most consistently recorded thyroid function test in routine UK clinical practice and is the primary parameter used to guide LT4 dose titration. Free T4 measurements were inconsistently available and therefore not suitable for inclusion in all analyses. Reported free T4 values are based on complete-case data only.

Potential confounders, defined as variables plausibly associated with both LT4 dose and study outcomes, were identified *a priori* using a directed acyclic graph specified in the study protocol and informed by clinical reasoning and prior evidence. The final adjustment set comprised baseline age, sex assigned at birth, Charlson Comorbidity Index, hypertension, serum TSH, smoking status, and use of proton-pump inhibitors, oestrogen, and testosterone.[21,22]

~~Follow-up~~

Participants were followed from the date of LT4 initiation ~~(time zero)~~ until the earliest of:

Formatted: Line spacing: 1.5 lines

Formatted: Normal

Formatted: List Paragraph, Bulleted + Level: 1 + Aligned at: 0.25" + Indent at: 0.5"

Formatted: Line spacing: 1.5 lines

Formatted: Normal

Formatted: List Paragraph, Numbered + Level: 1 + Numbering Style: 1, 2, 3, … + Start at: 1 + Alignment: Left + Aligned at: 0.25" + Indent at: 0.5"

Formatted: Line spacing: 1.5 lines

Formatted: Normal

1. Occurrence of the relevant outcome event (cardiovascular, bone, or mortality),
2. Death (for non-mortality outcomes), or
3. End of the study period, or
4.3. Administrative censoring.

Follow-up continued for up to ten years after LT4 initiation, with eligible participants contributing follow-up time between January 1, 2006, and December 31, 2021.

Several measures were undertaken to reduce potential sources of bias. Eligibility criteria, treatment strategies, and follow-up definitions were specified in advance according to a target trial framework to minimise selection and immortal time biases. Time zero was consistently defined as the date of LT4 initiation. Extreme biologically implausible values for TSH and body weight were excluded.

**Outcomes**

Outcomes were identified using diagnostic codes and clinical entries recorded in THIN. Diagnoses were defined using ICD-10 codes, supplemented by OPCS-4 codes where necessary. A comprehensive list of codes used to define outcomes is provided in Table S4.

*Primary Outcomes**:***

1. **Cardiovascular**: First occurrence of a composite outcome including angina, myocardial infarction, stroke, peripheral vascular disease, heart failure, atrial fibrillation, or any revascularisation procedures (e.g., stenting or bypass grafting) recorded during follow-up.
2. **Bone**: First recorded diagnosis of osteoporosis or a fragility fracture during follow-up.
3. **All-cause mortality**: Death recorded during follow-up. Data on cause-specific mortality were unavailable.

*Secondary Outcomes:*

1. **Healthcare utilisation:** Total number of primary care encounters (including face-to-face, telephone, or online consultations) per patient during follow-up.
2. **Thyroid function control:** Proportion of participants whose latest thyroid function test during follow-up showed an abnormal TSH level (<0·4 or >4·0-mU/L).

**Causal Contrast****Statistical analysis**

Formatted: List Paragraph, Numbered + Level: 1 + Numbering Style: 1, 2, 3, … + Start at: 1 + Alignment: Left + Aligned at: 0.25" + Indent at: 0.5"

Formatted: Line spacing: 1.5 lines

Formatted: Normal

Formatted: Adjust space between Latin and Asian text, Adjust space between Asian text and numbers

The primary causal contrast of interest was ~~the intention-to-treat effect,~~ comparing higher versus lower LT4 dose groups based on dose at treatment initiation. Because treatment allocation was not randomised and longitudinal adherence data were incomplete, analyses estimated the observational analogue of the effect of treatment initiation according to baseline prescribed dose group.~~A per-protocol analysis was not undertaken due to incomplete longitudinal data on TSH and prescription dispensing, which limited assessment of adherence or dose adjustments over time.~~

~~Data Analysis Plan~~

Baseline characteristics of participants in the higher- and lower-dose LT4 groups were summarised using descriptive statistics. Continuous variables are reported as mean (standard deviation) or median (interquartile range (IQR)), and categorical variables as counts and percentages. Covariate balance before and after weighting was assessed using standardised mean differences (SMDs). Baseline data were complete for included participants; therefore, no multiple imputation procedures were required.~~Baseline covariates included demographic factors (age, sex assigned at birth), clinical characteristics (Charlson Index, hypertension), baseline serum TSH, lifestyle factors (smoking status), and medication use (proton pump inhibitors and hormonal therapies).~~

~~Potential confounders, defined as variables plausibly associated with both LT4 dose and study outcomes, were identified *a priori* using a directed acyclic graph specified in the study protocol,[21] and informed by clinical reasoning and prior evidence. The final adjustment set comprised baseline age, sex assigned at birth, Charlson Comorbidity Index, hypertension, serum TSH, smoking status, and use of proton-pump inhibitors, oestrogen, and testosterone. These variables influence both LT4 prescribing practices and the study outcomes.[21,22] We calculated E-values to assess the robustness of observed associations to potential unmeasured confounding.~~

~~Baseline TSH and body weight were defined as the closest measurements recorded within three months before or after LT4 initiation. This window was selected to maximise data completeness and reflects routine UK practice, where NICE guidance recommends confirming hypothyroidism with at least two TSH measurements approximately three months apart.[11] Some patients may have started LT4 before a TSH was formally recorded in the electronic record; allowing measurements within this window ensures these patients are~~

Formatted: Normal

Formatted: Space Before: 0 pt, After: 0 pt

Formatted: Normal

Formatted: Space Before: 0 pt, After: 0 pt

~~included while capturing real-world prescribing patterns. When multiple measurements were available, the value closest to LT4 initiation was used. Baseline data were complete for all participants.~~

To emulate randomisation and adjust for measured confounding, inverse probability of treatment weighting (IPTW) was applied to reduce confounding by indication and other baseline imbalances based on measured covariates. Propensity scores estimating the probability of receiving a higher LT4 dose were derived from logistic regression models~~-~~.

~~including the covariates listed above.~~ Unstabilised weights were calculated as the inverse of the propensity score for participants in the higher-dose group and the inverse of one minus the propensity score for those in the lower-dose group, with extreme weights truncated at the 1st and 99th percentiles. Propensity score distributions were examined to assess overlap between treatment groups, ensuring adequate positivity (Appendix page 12). This created a weighted pseudo-population with balanced baseline covariates between treatment groups. E-values were calculated to assess the robustness of observed associations to potential unmeasured confounding.~~Covariate definitions and coding are provided in Table S5 and S6.~~

Formatted: Space Before: 0 pt, After: 0 pt

Weighted Cox proportional hazards models estimated hazard ratios (HRs) and 95% confidence intervals (CIs) for each outcome. Time zero was defined as the date of LT4 initiation, and participants were followed until outcome occurrence, death, censoring, or study end. Proportional hazards assumptions were assessed using Schoenfeld residuals. Weighted Kaplan–Meier curves were used to illustrate cumulative incidence by LT4 dose group.[23] Absolute risk differences were estimated at 10 years.

To assess potential non-linear dose–response relationships, IPTW-adjusted Cox models incorporating natural cubic splines with three degrees of freedom were fitted. Healthcare utilisation was analysed using IPTW-weighted negative binomial regression, reporting incidence rate ratios (IRRs) with 95% CIs.

Formatted: Space Before: 0 pt, After: 0 pt

Four sensitivity analyses were conducted. First, statin use was additionally included in the inverse probability of treatment weighting model. Second, the calendar year was incorporated into the outcome model. Third, analyses were restricted to participants with TSH measurements recorded prior to LT4 initiation. Fourth, time zero was redefined as the initiation of the second LT4 prescription among participants receiving a repeat

prescription.~~Sensitivity analyses additionally adjusted for statin use. Two further sensitivity analyses were performed: first, restricting the cohort to participants with TSH measurements prior to levothyroxine initiation only; and second, redefining time zero at the initiation of the second levothyroxine prescription among participants receiving a repeat prescription.~~ Pre-specified subgroup analyses were conducted by sex assigned at birth, age group (51–64 years and ≥65 years), and ~~baseline~~ TSH levels (≤10·0 and >10·0 mU/L)~~ levels~~. All analyses were conducted in R version 4.5.1, using the MatchIt package.[24]

~~All analyses were conducted in R version 4.5.1.[24]~~

~~Ethical Considerations~~

~~This study received ethical approval from the University of Sunderland Research Ethics Group (Reference: 011081). Scientific approval for the use of data from THIN was granted by the THIN Scientific Review Committee (Protocol Number 22-003).~~

**Role of the ~~f~~Funding ~~s~~Source**

The funder of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report. ~~The funder was not involved in the study design, data collection, data analysis, data interpretation, report writing, or the decision to submit the paper for publication.~~

**Results**

*Cardiovascular outcomes*

A total of 1~~5,463~~ 9,027 adults ~~aged >50 years, with diagnosed primary hypothyroidism, who initiated LT4 therapy,~~ were ~~included~~ eligible ~~in~~ for ~~the primary cardiovascular~~ analysis (Figure 1). Of those 19,027 patients, 15,463 patients were included in the primary cardiovascular analysis. Based on TBW, 11,762 participants (76%) initiated LT4 at a dose of ≤1·~~0 mcg~~0mcg/kg/day and 3,701 (24%) at a dose of >1·~~0 mcg~~0mcg/kg/day. After applying IPTW ~~to adjust for baseline confounders~~, the pseudo-population ~~was weighted to~~ represented 15,430 (51%) and 14,899 (49%) participants in the low and high dose groups, respectively.

Formatted: Normal, Line spacing: single

Formatted: Normal

Formatted: Font: Not Bold, Italic

Participants ~~were followed for up to 10 years from LT4 initiation, with~~had a median follow-up of 1~~0·~~0·0 years (~~interquartile range~~IQR 8·~~·~~6–10·~~·~~0~~), ending). at the occurrence of the outcome of interest, death, administrative censoring, or the end of the study period (31 December 2021).~~

**Figure 1 Flow diagram illustrating cohort selection, inclusion and exclusion criteria, and allocation to ~~levothyroxine dose~~outcome groups ~~for the cohort included in the cardiovascular outcomes analysis~~.**

~~Baseline characteristics before and after weighting are summarised in Table 2, based on TBW calculations.~~ After weighting, all covariates were well balanced between dose groups, with standardised mean differences <0·1 (~~Table S7~~Appendix page 13). The weighted mean age was 67 years in both treatment groups, with similar distributions of sex assigned at birth (22% versus 21% male), smoking status (43% in both), comorbidity burden, and concomitant medication use (Table 2). ~~Baseline covariate data used for propensity score estimation were complete, with no missing values for included participants.~~

Median ~~baseline~~ TSH was lower in the high dose group (5·7mU/L [2·6–10·2]~~ μIU/mL~~) compared with the low-dose group (6·8mU/L [5·2–9·8]~~ μIU/mL~~), and this difference persisted at the end of follow-up (2·3mU/L [1·2–4·6] versus 2·8mU/L [1·6–4·2]~~ μIU/mL~~). Weighted mean free T4 concentrations were higher in the high-dose group both at baseline (15·2 versus 13·1-pmol/L) and at the end of follow-up (17·3 versus 15·7-pmol/L).

~~*Cardiovascular Outcomes*~~

In analyses based on TBW, initiating LT4 at a higher dose was associated with a statistically significant increased risk of cardiovascular events compared with lower dose therapy (aHR 1·18, 95% ~~CI:~~CI 1·08–1·29) (Figure 2A). Consistent with these adjusted estimates, unweighted event rates were higher in the high-dose group (19·7% vs 17·7%). Similar findings were observed when dosing was defined by LBW (aHR 1·15, 95% ~~CI:~~CI 1·06–1·24) (Figure 2A). ~~Consistent with these adjusted estimates, unadjusted event rates were higher in the high-dose group, with cardiovascular events occurring in 711 (19·2%) participants receiving high-dose LT4 compared with 2,035 (17·3%) of those receiving low-dose therapy.~~

**Formatted:** Font color: Text 1

Subgroup analyses generally demonstrated consistent trends, although not all estimates reached statistical significance. The association was particularly evident among adults aged ≥65 years for both TBW (aHR 1·18, 95% ~~CI:~~CI 1·08–1·29) and LBW (aHR 1·15, 95% ~~CI:~~CI 1·06–1·24) based dosing.~~.~~ E-values for the cardiovascular outcomes ranged from 1·34 to 2·13, indicating that a moderately strong unmeasured confounder would be required to fully explain the observed associations (Appendix page 14). In sensitivity analyses, the association with cardiovascular events was attenuated and no longer statistically significant in the lag-adjusted analysis (Appendix pages 15-28).~~Sensitivity analyses produced comparable results (Figures S1-S4). E-values for the cardiovascular outcomes ranged from 1.24 to 1.65, indicating that a moderately strong unmeasured confounder would be required to fully explain the observed associations (Table S8).~~

**Figure 2 Adjusted hazard ratios ~~(aHR)~~ with 95% confidence intervals (CI) and p-values for Cardiovascular, Bone, and All-Cause Mortality Events by Levothyroxine Dose per Total Body Weight and Lean Body Weight. (A) Cardiovascular events, (B) Bone outcomes, (C) All-cause mortality.**

Kaplan–Meier curves demonstrated a higher cumulative incidence of cardiovascular events among participants initiated on higher LT4 doses (>1·~~0 mcg~~0mcg/kg/day, TBW) compared with lower doses (≤1·~~0 mcg~~0mcg/kg/day, TBW) (Figure 3A). Similar patterns were observed when dosing was defined using LBW ~~(Figure S5)~~and split by sex assigned at birth (Appendix pages 29-30). The ~~absolute risk~~weighted 10-year absolute risk difference between higher and lower LT4 dose groups was ~~3.8~~1·3% (95% CI -0·3-3·0) for the total body weight–based exposure definition and 0·7~~2.8~~% (95% CI -0·7-2·0) for the lean body weight–based definition. Spline analyses indicate increasing cardiovascular risk with TBW-based LT4 dosing, whereas for LBW-based dosing, risk appeared to increase modestly at doses between approximately ~~1.~~1·3–3·0~~-~~mcg/kg/day (~~Figure S6 and Figure S7~~Appendix pages 31-32).

**Figure 3 Kaplan-Meier Survival Curve for Cardiovascular Events, Bone Outcomes and All-Cause Mortality According to ~~by~~ Levothyroxine Dose~~,~~ Based on Total Body Weight. (A) Cardiovascular events, (B) Bone outcomes, (C) All-cause mortality.**

The red line represents the higher-dose group (>1·~~0 mcg~~0mcg/kg/day), and the black line represents the lower-dose group (≤1·~~0 mcg~~0mcg/kg/day). Shaded areas represent 95% confidence intervals. ~~Numbers at risk are shown below.~~

*Bone outcomes~~Bone Outcomes~~*

Higher initial doses of LT4 (>1·~~0 mcg~~0mcg/kg/day) were associated with a statistically significant increased risk of bone outcomes compared with lower doses (≤1·~~0 mcg~~0mcg/kg/day). For TBW-based dosing, the aHR was 1·29 (95% ~~CI:~~CI 1·16–1·43), and for LBW-based dosing, the aHR was 1·16 (95% ~~CI:~~CI 1·06–1·28) (Figure 2B). Similarly, unweighted event rates were higher in the higher-dose group (13·1% vs 9·1%).~~Unadjusted bone events occurred in 9·0% of the low-dose group compared with 12·6% of the high-dose group, consistent with the adjusted findings.~~ Subgroup analyses showed broadly consistent trends across age, sex assigned at birth, and ~~baseline~~ TSH categories, although confidence intervals overlapped ~~one another~~ and not all estimates reached statistical significance. E-values ranged from 1·49 to 2·79, indicating that a moderately strong unmeasured confounder would be required to fully explain the observed associations (Appendix page 14). Sensitivity analyses produced comparable results (Appendix pages 15-28). ~~Sensitivity analyses yielded similar results (Figures S1-S4). E-values ranged from 1.49 to 2.79, indicating that a moderately strong unmeasured confounder would be required to fully explain the observed associations (Table S8).~~

Kaplan–Meier curves demonstrated a higher cumulative incidence of bone events among participants initiated on higher LT4 doses (>1·0~~-~~mcg/kg/day, TBW) compared with lower doses (≤1·0~~-~~mcg/kg/day, TBW), although the confidence intervals partly overlapped (Figure 3B~~4~~). Comparable trends were observed for LBW-based dosing and when split by sex assigned at birth (~~Figure S8~~Appendix pages 33-34). The ~~absolute risk~~weighted 10-year absolute risk difference was 3~~2~~·~~-~~0~~475~~% (95% CI 1·58-4·49) for total body weight–based dosing and 1·~~-~~74% (95% CI 0·51-2·78) for lean body weight–based dosing. Spline analyses suggested increasing bone risk above approximately 1·~~6 mcg~~6mcg/kg/day for TBW-based dosing and above 2·~~0 mcg~~0mcg/kg/day for LBW-based dosing (~~Figure S9 and Figure S10~~Appendix pages 35-36).

~~**Figure 4 Kaplan-Meier Survival Curve for Bone Events by Levothyroxine Dose, Based on Total Body Weight.**~~

~~The red line represents the higher-dose group (>1·0 mcg/kg/day), and the black line represents the lower-dose group (≤1·0 mcg/kg/day). Shaded areas represent 95% confidence intervals. Numbers at risk are shown below.~~

*Mortality outcomes*

Formatted: Font: Italic

~~*All-Cause Mortality*~~

Higher initial LT4 doses were associated with a statistically significant increased risk of all-cause mortality compared with lower doses. Using TBW-based dosing, the aHR was 1·17 (95% ~~CI:~~CI 1·09–1·25), with a similar association observed using LBW-based dosing (aHR 1·12, 95% ~~CI:~~CI 1·05–1·19). Unadjusted event rates ~~showed higher all-cause mortality in the high-dose LT4 group (20·1%) than in the low-dose group (17·7%).~~ were consistent (Figure 2C). Subgroup analyses demonstrated broadly consistent patterns across age, sex assigned at birth, and ~~baseline~~ TSH categories, although statistical significance varied between strata. E-values for mortality outcomes ranged from 1·28 to 2·21(Appendix pages 14). Sensitivity analyses produced comparable results (~~Figures S1-S~~Appendix pages 15-28~~4~~). ~~E-values for mortality outcomes ranged from 1.28 to 2.21, suggesting that a modest-to-moderate unmeasured confounder would be required to fully explain the observed associations (Table S8).~~

Kaplan–Meier curves indicated a modestly higher cumulative incidence of all-cause mortality among participants initiated on higher LT4 doses (>1·~~0 mcg~~0mcg/kg/day, TBW) compared with lower doses (≤1·~~0 mcg~~0mcg/kg/day, TBW), with overlapping confidence intervals (Figure 3C~~5~~). Similar trends were observed for LBW-based dosing and when split by sex assigned at birth (~~Figure S11~~Appendix pages 37-38). The ~~absolute risk~~weighted 10-year absolute risk difference was 3·43% (95% CI 1·74-5·13) ~~3.67%~~ for total body weight–based dosing and 2~~2~~·~~.2~~174~~%~~ (95% CI 0·76-3·47) for lean body weight–based dosing. Spline analysis demonstrated a gradual increase in mortality risk with TBW-based LT4 dosing above approximately 1·0 mcg/kg/day (~~Figure S12~~Appendix page 39). For LBW-based dosing, the relationship was less clear, with risk appearing to increase modestly at doses between approximately ~~1.~~1·3–3·0 mcg/kg/day (Appendix page 40). ~~(Figure S13).~~

~~**Figure 5 Kaplan-Meier Survival Curve for All-Cause Mortality Events by Levothyroxine Dose, Based on Total Body Weight.**~~

~~The red line represents the higher-dose group (>1·0 mcg/kg/day), and the black line represents the lower-dose group (≤1·0 mcg/kg/day). Shaded areas represent 95% confidence intervals. Numbers at risk are shown below.~~

*Secondary outcomes*~~*Secondary Outcomes*~~

Older adults with hypothyroidism who were initiated on higher LT4 doses had greater healthcare utilisation compared with those started on lower doses. The higher dose group had a statistically significantly higher number of ~~GP~~ visits compared with the lower dose group (incidence rate ratio ~~IRR~~ 1·09, 95% CI 1·06–1·13). Median ~~GP~~ visits were 246 (IQR 152–378) in the lower dose group and 272 (IQR 164–422) in the higher dose group.

At study end, abnormal TSH levels (<0·4 or >4·0 mU/L~~μIU/mL~~) were recorded in 27% (4,166/15,430) of participants in the low-dose group and 30% (4,470/14,899) of the high-dose group. Within the low-dose group, 22% (917/4,166) had low, and 78% (3,249/4,166) had elevated TSH values, compared with 43% (1,935/4,470) and 57% (2,535/4,470), respectively, in the high-dose group. While the overall prevalence of abnormal TSH did not differ between groups, the distribution of low versus high TSH values differed between dosing groups, with low TSH observed more frequently among participants initiated on higher LT4 doses.

## Discussion

~~*Summary of the Main Findings*~~

In this emulated target trial of adults aged over 50 years with primary hypothyroidism, initiation of LT4 at doses greater than 1·~~0 mcg~~0mcg/kg/day, whether based on TBW or LBW, was associated with higher risks of cardiovascular events, bone outcomes, and all-cause mortality. Notably, stratified analyses suggested that the potential harm of higher initial LT4 doses was more pronounced in adults 65 years and over, whereas outcomes in the 50-64 years age group were largely non-significant, highlighting age-related differences in sensitivity to LT4 dosing. Participants initiated on higher LT4 doses also had a greater

number of primary care consultations, suggesting an association between higher initial dosing and increased healthcare utilisation. Together, these findings indicate that higher initial weight-based LT4 dosing in older adults is associated with less favourable long-term outcomes.

For LBW-based dosing, binary analyses comparing doses above versus below 1·0 mcg/kg/day suggested higher risks of cardiovascular events and mortality. In contrast, spline analyses treating LBW as a continuous variable showed hazard ratios mostly near 1, with only a modest increase in risk at doses between approximately 1·3–3·0 mcg/kg/day. This pattern may reflect a threshold effect around 1·0 mcg/kg/day, which the binary analysis captures, or the limited number of participants at very low or very high LBW doses, which can flatten the spline curve. TBW-based dosing consistently showed increasing risk across the relevant dose range in both binary and spline analyses.

~~*Comparison with Other Literature*~~

These findings are consistent with previous observational studies linking LT4 overtreatment to increased cardiovascular morbidity and mortality in older adults.[8] For example, prior research has reported higher rates of atrial fibrillation among individuals receiving LT4 doses exceeding 75 ~~mcg~~mcg/day.[25] Similarly, the increased risk of osteoporosis and fractures observed in this study aligns with earlier findings of adverse skeletal outcomes among older adults treated with higher LT4 doses.[26–28] Previous literature also suggests that multiple LT4 dose adjustments may be associated with increased healthcare utilisation, including more frequent monitoring, medication adjustments, and management of complications.[29]

Together, this literature supports that LT4 requirements derived from studies in younger adults, typically approximately at 1·~~6 mcg~~6mcg/kg/day, may not be appropriate for older populations, who have lower lean body mass and increased hormone sensitivity. However, the present study examines the initial prescribed dose rather than achieved biochemical ~~control, and~~control and therefore complements rather than replaces evidence based on TSH-defined overtreatment.

~~*Strengths and Limitations*~~

Target trial emulation seeks to approximate the causal question that would be addressed in a hypothetical ~~aims to provide the same results that would have been found if an~~ RCT ~~had been run to answer a given question~~. This is achieved by strictly pre-defining eligibility, treatment, and follow-up so that it matches as closely as possible to a ~~hypothetical (~~target~~)~~ RCT, which minimises the biases usually found in observational studies. To emulate random assignment to treatment arms, we used ~~inverse probability of treatment weighting~~IPTW.

This study's strengths include its use of a large population-based UK primary care cohort with long-term follow-up and the application of a target trial emulation framework to reduce bias through clearly defined eligibility, treatment strategies, and follow-up.[17] To emulate a randomised trial and improve comparability between treatment groups, ~~inverse probability of treatment weighting~~IPTW was applied. This method adjusts the contribution of each participant so that baseline characteristics are balanced across groups, effectively creating a 'pseudo-population' whose size reflects the re-weighted contributions of participants. In addition, the evaluation of dosing based on both TBW and LBW provides a novel insight into practical alternatives for safer LT4 prescribing in older adults.

Several limitations should be considered. As with all observational analyses, residual confounding cannot be excluded, particularly confounding by indication, whereby patients prescribed higher initial doses may differ systematically in ways not fully captured by measured covariates. Although ~~inverse probability of treatment weighting~~IPTW was used to balance observed baseline characteristics, unmeasured factors such as disease severity, frailty, malabsorption, adherence, or clinician prescribing preferences may have influenced both dose selection and outcomes. Race and ethnicity data were not available within THIN, and, therefore, could not be accounted for in the analyses, despite their potential association with healthcare access, treatment patterns, and health outcomes. In addition, the study population was highly selected, comprising approximately 5% of patients identified in the database who met eligibility criteria; this reflects the application of strict inclusion and exclusion criteria required for the emulated target trial design, but may limit generalisability to all LT4 users in routine clinical practice.

Longitudinal data on dose adjustments, adherence, and serial TSH measurements were incomplete, precluding per-protocol analyses. Analysis therefore focused on initial LT4 dose

as a proxy for intended treatment strategy, ~~analogous to an intention-to-treat approach~~based on exposure at treatment initiation, recognising that subsequent titration commonly occurs in clinical practice. For cardiovascular outcomes, the association observed in the primary analysis was attenuated and no longer statistically significant when time zero was redefined at the second LT4 prescription. This finding suggests that part of the observed cardiovascular association could reflect early treatment changes, treatment persistence, or residual confounding. In contrast, associations with bone outcomes and all-cause mortality remained broadly consistent across sensitivity analyses~~Redefining time zero at the second LT4 prescription, as a proxy for treatment persistence, yielded similar results, supporting the robustness of the intention-to-treat framework despite potential early discontinuation or dose changes~~.

Estimated lean body weight may be inaccurate in frail older adults, potentially causing misclassification.[30] Excluding patients without recorded weights could introduce selection bias, since weights are more often measured in those with multiple comorbidities or in individuals who are underweight or overweight. Categorising doses using a ~~1.~~1·0 mcg/kg/day threshold may group widely different doses together while separating similar doses across the cut-off, potentially obscuring dose–response relationships. Sensitivity analyses using natural cubic splines to model dose continuously produced similar associations, supporting the robustness of the findings.

Some baseline measurements, including TSH, may have been recorded shortly after LT4 initiation, which could influence the observed associations. ~~Baseline~~ TSH and body weight were defined as the closest measurements within three months before or after treatment, reflecting real-world UK primary care practice where some values may be entered electronically post-initiation. Consequently, the lower TSH and higher ~~free thyroxine~~free T4 observed in the high-dose group may partly reflect early treatment effects rather than true pre-treatment thyroid function. ~~Free thyroxine~~ Free T4 and free triiodothyronine were measured inconsistently and therefore not included in weighting. However, in sensitivity analyses restricted to participants with TSH measurements recorded prior to LT4 initiation, effect estimates were materially unchanged, suggesting that inclusion of post-initiation measurements did not substantially bias the findings.

Diagnosis was based on a coded record of primary hypothyroidism, and thyroid-stimulating hormone (TSH) was included as a baseline variable without systematic free T4 or free triiodothyronine measurements, as these were not routinely recorded. This approach captures both overt and subclinical disease. ~~Diagnosis was based on a coded record of primary hypothyroidism. This approach included both overt and subclinical hypothyroidism, introducing some heterogeneity.~~ Nevertheless, initial ~~levothyroxine~~ LT4 dosing is generally recommended uniformly for elevated TSH, which aligns with the study's focus on initial dosing strategies. Patients with post-surgical or post-ablative hypothyroidism were excluded to maintain a clinically homogeneous cohort, as these individuals typically require immediate full replacement dosing and follow different treatment pathways.

THIN captures only UK primary care data and does not reliably include information on ethnicity, family history, consultation type, or secondary care. The c~~C~~ause of death was not available, precluding assessment of cause-specific mortality and limiting interpretation of mortality findings. Death represents a competing risk for cardiovascular and bone outcomes, and this limitation should be considered when interpreting the results. Healthcare utilisation was included as a secondary outcome and could also act as a mediator, potentially influencing dose titration and earlier detection of complications. The analyses conduced involved multiple outcomes, exposure definitions, and subgroup comparisons, and certain estimates, particularly in subgroups, may reflect chance rather than true associations. Residual confounding from unmeasured factors, as well as from variation between practices and care pathways, cannot be excluded. Prescribing practices for ~~levothyroxine~~ LT4 may have changed over the study period (2006–2021), and evolving clinical guidance and clinician behaviour could have introduced residual confounding. Diagnostic coding inaccuracies and the restriction to primary care may limit generalisability to other settings or healthcare systems.

~~*Implications for Practice and Research*~~

These findings highlight potential risks associated with higher initial LT4 dosing in adults over 50 years, particularly in those aged 65 years and over, and underscore the need for careful dose selection and monitoring in this population. The results describe associations observed in routine clinical practice and should not be interpreted as prescriptive dosing

recommendations. Clinical decisions should continue to balance symptom control, biochemical targets, comorbidities, and patient preferences.

While weight-based LT4 dosing may provide a more tailored approach, its implementation in routine primary care requires consideration of feasibility. Accurate weight measurement is essential, and calculation tools could be integrated into electronic prescribing systems or clinical decision support prompts to help facilitate safe and practical dosing in everyday practice.

Future research should aim to evaluate alternative LT4 initiation strategies in older adults, ideally through pragmatic randomised trials or prospective cohort studies incorporating adherence monitoring, serial thyroid function testing, cardiovascular and skeletal endpoints, and patient-reported outcomes. Further work is also needed to clarify optimal TSH targets and the influence of frailty, multimorbidity, and polypharmacy on LT4 requirements in ageing populations.

## Data Sharing

The data that supports the findings of this study are available from THIN, a wholly owned subsidiary of Cegedim SA, which owns the proprietary rights to THIN data. Restrictions apply to the availability of these data, which were used under license for the current study and are not publicly available. ~~Data are, however, available from the authors upon reasonable request and with the permission of THIN.~~

## ~~Competing~~ Declaration of i~~I~~nterest~~s~~

~~Salman Razvi~~SR has received speaker fees from Meck KGaA and Berlin Chemie AG, makers of Levothyroxine. All other authors declare that there were no conflicts of interest.

## Author contributions

SR developed the study concept. SW, SR, MH, RD, IM, and EDM designed the study. MH conducted all analyses, with guidance from EDM, SR, RD, IM, and SW. MH drafted the manuscript with feedback from SW, SR, RD, IM, and EDM. All authors read and approved the final manuscript and agree to be accountable for all aspects of the work.

**Acknowledgements**

This study is funded by the National Institute for Health and Care Research (NIHR) Applied Research Collaboration (ARC) North East and North Cumbria (NENC) (NIHR200173). The views expressed are those of the authors and not necessarily those of the NIHR or the Department of Health and Social Care.

**Abbreviations**

| | |
|---|---|
| aHR | Adjusted hazard ratio |
| CI | Confidence interval |
| dm+d | Dictionary of medicines and devices |
| ICD-10 | International Classification of Diseases, Tenth Revision |
| IPTW | Inverse probability of treatment weighting |
| IQR | Interquartile range |
| LBW | Lean body weight |
| LT4 | Levothyroxine |
| OPCS-4 | Office of Population Censuses and Surveys Classification of Interventions and Procedures, 4th revision |
| RCT | Randomised controlled trial |
| SMD | Standardised mean differences |
| TBW | Total body weight |
| THIN | The Health Improvement Network |
| TSH | Thyroid-stimulating hormone |
| T4 | Thyroxine |
| UK | United Kingdom |

**References:**

1. Cooper DS, Biondi B. Subclinical thyroid disease. Lancet 2012;379(9821):1142–1154; doi: 10.1016/S0140-6736(11)60276-6.

2. Canaris GJ, Manowitz NR, Mayor G, et al. The Colorado thyroid disease prevalence study. Arch Intern Med 2000;160(4):526–534; doi: 10.1001/archinte.160.4.526.

3. Vadiveloo T, Donnan PT, Murphy MJ, et al. Age- and gender-specific TSH reference intervals in people with no obvious thyroid disease in Tayside, Scotland: the Thyroid Epidemiology, Audit, and Research Study (TEARS). J Clin Endocrinol Metab 2013;98(3):1147–1153; doi: 10.1210/jc.2012-3191.

4. Surks MI, Hollowell JG. Age-specific distribution of serum thyrotropin and antithyroid antibodies in the US population: implications for the prevalence of subclinical hypothyroidism. J Clin Endocrinol Metab 2007;92(12):4575–4582; doi: 10.1210/jc.2007-1499.

5. Fournier J, Barret L, Khouri C, et al. The evidence base of the 10 most prescribed drugs in England, France, and the United States: a scoping review. Journal of Clinical Epidemiology 2024;174:111478; doi: 10.1016/j.jclinepi.2024.111478.

6. Leng O, Razvi S. Hypothyroidism in the older population. Thyroid Res 2019;12:2; doi: 10.1186/s13044-019-0063-3.

7. Centanni M, Duntas L, Feldt-Rasmussen U, et al. ETA guidelines for the use of levothyroxine sodium preparations in monotherapy to optimize the treatment of hypothyroidism. European Thyroid Journal 2025;14(4); doi: 10.1530/ETJ-25-0123.

8. Feldt-Rasmussen U, Effraimidis G, Bliddal S, et al. Risks of suboptimal and excessive thyroid hormone replacement across ages. J Endocrinol Invest 2024;47(5):1083–1090; doi: 10.1007/s40618-023-02229-7.

9. Thayakaran R, Adderley NJ, Sainsbury C, et al. Thyroid replacement therapy, thyroid stimulating hormone concentrations, and long term health outcomes in patients with hypothyroidism: longitudinal study. BMJ 2019;366:l4892; doi: 10.1136/bmj.l4892.

10. Santini F, Pinchera A, Marsili A, et al. Lean body mass is a major determinant of levothyroxine dosage in the treatment of thyroid diseases. J Clin Endocrinol Metab 2005;90(1):124–127; doi: 10.1210/jc.2004-1306.

11. Anonymous. Levothyroxine | Prescribing Information | Hypothyroidism | CKS | NICE. n.d. Available from: https://cks.nice.org.uk/topics/hypothyroidism/prescribing-information/levothyroxine/ [Last accessed: 8/5/2025].

12. Jonklaas J, Bianco AC, Bauer AJ, et al. Guidelines for the treatment of hypothyroidism: prepared by the american thyroid association task force on thyroid hormone replacement. Thyroid 2014;24(12):1670–1751; doi: 10.1089/thy.2014.0028.

13. Holley M, Razvi S, Farooq MS, et al. Cardiovascular and bone health outcomes in older people with subclinical hypothyroidism treated with levothyroxine: a systematic review and meta-analysis. Syst Rev 2024;13(1):123; doi: 10.1186/s13643-024-02548-7.

14. Holley M, Razvi S, Maxwell I, et al. Assessing the Cardiovascular Effects of Levothyroxine Use in an Ageing UK Population with Subclinical Hypothyroidism: Emulated Target Trial (ACEL-UK-ETT). Thyroid; 2026.; doi: 10.1177/10507256261426576.

15. Blak B, Thompson M, Dattani H, et al. Generalisability of The Health Improvement Network (THIN) database: demographics, chronic disease prevalence and mortality rates. Informatics in primary care 2011;19:251–5; doi: 10.14236/jhi.v19i4.820.

16. Holley M, Razvi S, Maxwell I, et al. Assessing the Cardiovascular Effects of Levothyroxine Use in an Ageing United Kingdom Population (ACEL-UK): Cohort

794 Study. The Journal of Clinical Endocrinology & Metabolism 2025;dgaf208; doi:
795 10.1210/clinem/dgaf208.

796 17. Hernán MA, Robins JM. Using Big Data to Emulate a Target Trial When a Randomized
797 Trial Is Not Available. American Journal of Epidemiology 2016;183(8):758–764; doi:
798 10.1093/aje/kwv254.

799 18. Hernán MA, Wang W, Leaf DE. Target Trial Emulation: A Framework for Causal
800 Inference From Observational Data. JAMA 2022;328(24):2446–2447; doi:
801 10.1001/jama.2022.21383.

802 19. Yao XI, Wang X, Speicher PJ, et al. Reporting and Guidelines in Propensity Score
803 Analysis: A Systematic Review of Cancer and Cancer Surgical Studies. J Natl Cancer
804 Inst 2017;109(8):djw323; doi: 10.1093/jnci/djw323.

805 20. Boer P. Estimated lean body mass as an index for normalization of body fluid volumes in
806 humans. American Journal of Physiology-Renal Physiology 1984; doi:
807 10.1152/ajprenal.1984.247.4.F632.

808 21. Flynn RW, Bonellie SR, Jung RT, et al. Serum thyroid-stimulating hormone
809 concentration and morbidity from cardiovascular disease and fractures in patients on
810 long-term thyroxine therapy. J Clin Endocrinol Metab 2010;95(1):186–193; doi:
811 10.1210/jc.2009-1625.

812 22. Vita R, Saraceno G, Trimarchi F, et al. Switching levothyroxine from the tablet to the
813 oral solution formulation corrects the impaired absorption of levothyroxine induced by
814 proton-pump inhibitors. J Clin Endocrinol Metab 2014;99(12):4481–4486; doi:
815 10.1210/jc.2014-2684.

816 23. Denz R, Klaaßen-Mielke R, Timmesfeld N. A comparison of different methods to adjust
817 survival curves for confounders. Statistics in Medicine 2023;42(10):1461–1479; doi:
818 10.1002/sim.9681.

819 24. R Core Team. R: A Language and Environment for Statistical Computing. 2024.

820 25. Gong IY, Atzema CL, Lega IC, et al. Levothyroxine dose and risk of atrial fibrillation: A
821 nested case-control study. American Heart Journal 2021; doi: 10.1016/j.ahj.2020.09.016.

822 26. Ko Y-J, Kim JY, Lee J, et al. Levothyroxine dose and fracture risk according to the
823 osteoporosis status in elderly women. J Prev Med Public Health 2014;47(1):36–46; doi:
824 10.3961/jpmph.2014.47.1.36.

825 27. Leese GP, Flynn RV. Levothyroxine dose and fractures in older adults. BMJ
826 2011;342:d2250; doi: 10.1136/bmj.d2250.

827 28. Turner MR, Camacho X, Fischer HD, et al. Levothyroxine dose and risk of fractures in
828 older adults: nested case-control study. BMJ 2011;342:d2238; doi: 10.1136/bmj.d2238.

829 29. Ernst FR, Barr P, Elmor R, et al. The Economic Impact of Levothyroxine Dose
830 Adjustments: the CONTROL HE Study. Clin Drug Investig 2017;37(1):71–83; doi:
831 10.1007/s40261-016-0462-3.

30. Mitchell SJ, Kirkpatrick CMJ, Le Couteur DG, et al. Estimation of lean body weight in older community-dwelling men. Br J Clin Pharmacol 2010;69(2):118–127; doi: 10.1111/j.1365-2125.2009.03586.x.

Formatted: Bibliography

CRISP checklist