PROTOCOL AND STATISTICAL ANALYSIS PLAN

SUBCLINICAL HYPERTHYROIDISM, CARDIOVASCULAR DISEASE AND ALL-CAUSE MORTALITY IN PRIMARY CARE

Study design and setting

We will conduct a retrospective cohort study using data from the Dutch General Practitioner (GP) Database from the PHARMO Data Network. The GP Database is a longitudinal database comprising data from electronic patient records registered by GPs. Records include information on diagnoses and symptoms, laboratory test results, and prescriptions. Prescription records include information on the type of product, prescription date, strength, dosage regimen, quantity, and route of administration. Drug prescriptions are coded according to the WHO Anatomical Therapeutic Chemical (ATC) Classification System. Diagnoses and symptoms will be coded according to the International Classification of Primary Care (ICPC). The GP Database covers a catchment area representing 3.2 million residents (~20% of the Dutch population). More information has been published elsewhere (Overbeek JA, Swart KMA, Houben E, Penning-van Beest FJA, Herings RMC. Completeness and Representativeness of the PHARMO General Practitioner (GP) Data: A Comparison with National Statistics. Clin Epidemiol. 2023;15:1-11. https://doi.org/10.2147/CLEP.S389598).

Time frame

Data from January 1, 2012, to December 31, 2021, will be analyzed, encompassing a ten-year period to assess the incidence trends of subclinical hyperthyroidism (SHT) and its cardiovascular complications.

Study population

The study population consists of patients identified through TSH measurements requested in primary care, excluding those with known thyroid diseases.

SHT is defined by a TSH concentration below the lower limit of the reference interval and an FT4 concentration within the reference interval in the same sample or patients with a recorded ICPC code A91.07 (SHT). Method specific reference intervals will be used for each TSH and FT4 measurement.

Exclusion criteria for subclinical hyperthyroidism are: in the 2 years prior to inclusion (1), use of thyroid medication (ATC starting with H03), amiodarone (ATC C01BD01) or ever recorded use of lithium (ATC N05AN01) or (2) mention of ICPC codes: T85 (hyperthyroidism), T86 (hypothyroidism), A91.06 (subclinical hypothyroidism), A91.07 (subclinical hyperthyroidism) or T71 (thyroid malignancy).

The patients included based on ICPC code could include several misclassifications, where the biochemical diagnosis was subclinical hypothyroidism, in that case, only patients with a laboratory confirmed subclinical hyperthyroidism will be included. Furthermore, patients without available data for the study period or who could not be matched to controls will be excluded from the analysis. In addition, subjects younger than 18 years of age at time of inclusion will also be excluded. In order to minimize the chance of thyroid disorders being related to pregnancy we will exclude patients with

subclinical hyperthyroidism which are included three months before to one year after a registered pregnancy (ICPC codes W78 "Desired pregnancy" or W79 "Unwanted pregnancy"). If a pregnancy is registered in the follow-up TSH and FT4 values will also be excluded in the lab dataset three months before to 1 year after a registered pregnancy. We will perform an additional sensitivity analysis, requiring a second suppressed TSH measurement 4 weeks to 6 months after the initial TSH measurement for inclusion as subclinical hyperthyroidism.

As a reference group, euthyroid patients will be included, defined as having TSH and FT4 measurements within the reference interval. Exclusion criteria are TSH or FT4 measurements outside of the reference interval or mention of ICPC codes: T85 (hyperthyroidism), T86 (hypothyroidism), A91.06 (subclinical hypothyroidism), A91.07 (subclinical hyperthyroidism) or T71 (thyroid malignancy) during 5-year follow-up. The euthyroid reference group were matched to cases by age, sex, and GP practice with four intended matched subjects for each SHT subject.

A detailed overview of the patient selection process will be presented in order to illustrated the flow diagram of the inclusion and exclusion criteria applied in this study.

Exploratory and confirmatory factor analysis

Exploratory factor analysis (EFA) was conducted on a randomly selected subset comprising 25% of the database. The EFA helped to define the exact variables and develop robust outcome measures. After finalizing the analysis protocol based on the EFA results, the protocol was registered on the ISRCTN registry to maintain transparency and reproducibility.

Once the analysis protocol is published, the remaining 75% of the database, which had been withheld to prevent bias, will be decoded and made available for research. A confirmatory factor analysis (CFA) will then be performed on this larger dataset to validate the factors identified in the EFA and to confirm the consistency and reliability of the defined outcome measures.

Outcomes

The primary outcomes of interest are cardiovascular events and conditions associated with SHT, specifically: atherosclerotic complications, atrial fibrillation, heart failure and an increased cardiometabolic risk profile. Additionally, all-cause mortality is an outcome of interest. Atherosclerotic complications are defined as ICPC codes K74 (angina pectoris), K74.01 (unstable angina pectoris), K74.02 (stable angina pectoris), K75 (acute myocardial infarction), K76 (chronic ischemic heart disease), K76.01 (coronary sclerosis), K76.02 (earlier myocardial infarction more than 4 weeks ago), K89 (transient cerebral ischemia), K90 (cerebrovascular accident), K90.01 (subarachnoid hemorrhage) , K90.02 (intracerebral hemorrhage) , K90.03 (cerebral infarction) and K92.01 (intermittent claudication). Atrial fibrillation is defined as ICPC code K78 (atrial fibrillation/flutter). Heart failure is defined as ICPC codes K77 (heart failure, K77.01 (acute heart failure), K77.03 (HFpEF) and K77.04 (HFmrEF or HFrEF).

The outcome cardiometabolic risk is defined as a composite variable representing the sum of individual risk factors associated with cardiometabolic disease. Each identified risk factor contributes one point to the total cardiometabolic risk score. The risk factors are:

1. Hypertension: Defined by the presence of ICPC codes K86 or K87 or the use of antihypertensive medications (specifically antihypertensives [ATC C02], diuretics [ATC C03], beta-blocking agents [ATC C07], calcium channel blockers [ATC C08] or agents acting on the renin-angiotensin system [ATC C09]), whichever comes first: 1 point.

2. Hypercholesterolemia: Defined by either LDL cholesterol level >2.6 mmol/L (measurement timepoint to be determined) or use of statins (ATC C10), whichever comes first: 1 point.
3. Kidney Disease: Defined by either: estimated glomerular filtration rate (eGFR) <60 mL/min/1.73 m² or albumin-to-creatinine ratio (ACR) >3 mg/mmol, whichever comes first: 1 point.
4. Diabetes Mellitus: Defined by the presence of ICPC code T90 or the use of diabetes medications (ATC A10), whichever comes first: 1 point.

A defined episode will excluded if a defined ICPC code is already presented in the year before cohort entry date, to increase the likelihood of including only new events. Data will be compared after 2 and 5 years of follow-up.

Statistical analysis

Statistical analyses will be conducted to compare the cardiovascular outcomes and associated factors between patients with subclinical hyperthyroidism (SHT) and the euthyroid reference group. For categorical variables, absolute and relative frequencies will be reported. For continuous variables, means and standard deviations (SD) will be calculated for normally distributed data, while medians and interquartile ranges (IQR) will be used for non-normally distributed data.

The primary outcomes are the incidence of cardiovascular events and conditions, including: atherosclerotic complications, atrial fibrillation, heart failure, cardiometabolic risk profile and all-cause mortality. Incidence rates of cardiovascular events and conditions will be calculated for both SHT and euthyroid controls. Rates are expressed as events per 1000 person-years. The time-to-event will be calculated from the cohort entry date to the first occurrence of the event or censored at the end of available follow-up.

Multivariate-adjusted cox proportional hazards models will be used to estimate the hazard ratios (HRs) and 95% confidence intervals (CIs) for assessing the risk of cardiovascular events and all-cause mortality between the SHT and reference group, after correction for confounding factors. Considered confounders were age, sex and the presence of comorbid conditions in the year before inclusion such as hypertension, diabetes mellitus, hypercholesterolemia and kidney disease (defined as in outcomes).

Additionally, subgroup analyses will be performed by stratifying for different age groups (notably, 18-29, 30-49, 50-69, and 70+ years of age), sex and TSH concentration at inclusion. Further analyses will include subgroups groups with persistent SHT, recovery, or progression to hyperthyroidism. These groups were defined as follows: (1) Progression to overt hyperthyroidism, defined as FT4 levels above the reference range during the follow-up period; (2) Recovery, characterized by TSH levels returning to within the reference range, but never above, at any time during the follow-up; (3) Persisting subclinical hyperthyroidism, defined as persistently suppressed TSH levels with normal FT4 levels throughout the follow-up period. For these groups the TSH and FT4 values in the first four weeks will be excluded, since they are deemed too close to the inclusion date and the Dutch primary care guideline recommends testing after three months. A sensitivity analysis will be conducted to assess the robustness of the findings by excluding events occurring within the first year of follow-up, in order to minimize the potential for reverse causation.

Confidence intervals were set at 95% for all estimates. All statistical analyses were conducted using R version 4.2.2 (2022-10-31 ucrt). The following R packages were used for the analysis: car, dplyr, ggplot2, ggpubr, lubridate, multcomp, survminer, survival, readr, and tidyr.