Towards personalized TSH reference ranges: A genetic and population-based approach in three independent cohorts

Short title: Genetically determined TSH reference ranges

Aleksander Kuś [1,2,3]*, Rosalie B.T.M. Sterenborg [1,2,4]*, Eirin B. Haug [5], Tessel E. Galesloot [6], W. Edward Visser [1], Johannes W.A. Smit [4], Tomasz Bednarczuk [3], Robin P. Peeters [1], Bjørn O. Åsvold [5,7,8], Alexander Teumer [9,10,11,12], Marco Medici [1,2,4]

1. Academic Center for Thyroid Diseases, Department of Internal Medicine, Erasmus Medical Center, Rotterdam, The Netherlands
2. Department of Epidemiology, Erasmus Medical Center, Rotterdam, The Netherlands
3. Department of Internal Medicine and Endocrinology, Medical University of Warsaw, Warsaw, Poland
4. Department of Internal Medicine, Radboud University Medical Center, Nijmegen, The Netherlands
5. HUNT Center for Molecular and Clinical Epidemiology, Department of Public Health and Nursing, NTNU, Norwegian University of Science and Technology, Trondheim, Norway
6. Department for Health Evidence, Radboud University Medical Center, Nijmegen, The Netherlands
7. Department of Endocrinology, Clinic of Medicine, St. Olavs Hospital, Trondheim University Hospital, Trondheim, Norway
8. HUNT Research Centre, Department of Public Health and Nursing, NTNU, Norwegian University of Science and Technology, Levanger, Norway
9. Institute for Community Medicine, University Medicine Greifswald, Greifswald, Germany
10. DZHK (German Center for Cardiovascular Research), partner site Greifswald, Greifswald, Germany
11. Department of Population Medicine and Lifestyle Diseases Prevention, Medical University of Bialystok, Bialystok, Poland
12. Department of Psychiatry and Psychotherapy, University Medicine Greifswald, Greifswald, Germany

*** contributed equally**

Corresponding author:

Marco Medici MD, PhD, MSc

Room Ee502, Academic Center for Thyroid Diseases, Department of Internal Medicine and Department of Epidemiology, Erasmus Medical Center, Dr Molewaterplein 50, 3015 GE, Rotterdam, The Netherlands

Email: m.medici@erasmusmc.nl

Tel: +31-10-7040570

Abstract

Background: Serum thyroid-stimulating hormone (TSH) measurement is the diagnostic cornerstone for primary thyroid dysfunction. There is high inter-individual, but limited intra-individual variation in TSH concentrations, largely due to genetic factors. The currently used wide population-based reference intervals may lead to inappropriate management decisions.

Methods: A polygenic score (PGS) including 59 genetic variants was used to calculate genetically-determined TSH reference ranges in a thyroid disease-free cohort (N=6,834). Its effect on reclassification of diagnoses was investigated when compared to using population-based reference ranges. Next, results were validated in a second independent population-based thyroid disease-free cohort (N=3,800). Potential clinical implications were assessed in a third independent population-based cohort including individuals without thyroid disease (N=26,321) as well as individuals on levothyroxine (LT4) treatment (N=1,132).

Results: PGS was a much stronger predictor of individual TSH concentrations than FT4 (total variance in TSH concentrations explained 9.2-11.1% vs. 2.4-2.7%, respectively) or any other non-genetic factor (total variance in TSH concentrations explained 0.2-1.8%). Genetically-determined TSH reference ranges differed significantly between PGS quartiles in all cohorts, while the differences in FT4 concentrations were absent or only minor. Up to 24.7-30.1% of

individuals, previously classified as having subclinical hypo- and hyperthyroidism when using population-based TSH reference ranges, were reclassified as euthyroid when genetically-determined TSH reference ranges were applied. Individuals in the higher PGS quartiles had a higher probability of being prescribed LT4 treatment compared to individuals from the lower PGS quartiles (3.3% in Q1 vs. 5.2% in Q4, $P_{for\ trend}=1.7x10^{-8}$).

Conclusions: Individual genetic profiles have potential to personalize TSH reference ranges, with large effects on reclassification of diagnosis and LT4 prescriptions. As the currently used PGS can only predict approximately 10% of inter-individual variation in TSH concentrations, it should be further improved when more genetic variants determining TSH concentrations are identified in future studies.

Key words: thyroid, TSH, reference range, hypothalamus-pituitary-thyroid axis setpoint, polygenic score, single nucleotide polymorphism, genetics

Introduction

Thyroid dysfunction is among the most common disorders worldwide, affecting 5-15% of the general population [1]. Due to the highly pleiotropic effects of thyroid hormones (THs), both hypothyroidism and hyperthyroidism are associated with various adverse health outcomes and mortality [2,3]. Thyroid function is narrowly regulated by the hypothalamus-pituitary-thyroid (HPT) axis, in which thyroid-stimulating hormone (TSH) plays a key regulatory role. As TSH is generally the most sensitive indicator of thyroid function, the diagnosis of thyroid dysfunction is primarily based on measurements of TSH concentrations [2,3]. Ever since the introduction of TSH testing in daily clinical practice in the 1970's, reference ranges have been based on the 2.5th and 97.5th percentiles of observed values in a reference population of presumably healthy individuals [4]. However, serum TSH and TH concentrations in healthy individuals show substantial inter-individual variation leading to wide population-based reference ranges, while the intra-individual variation is much smaller, suggesting that every individual has its own unique HPT-axis setpoint, *i.e.* a specific TSH concentration corresponding to an optimal function of the thyroid gland [5,6]. Consequently, a TSH concentration within the population-based reference range does not exclude mild thyroid dysfunction, as this level might be abnormal for the respective individual. *Vice versa*, individuals with TSH concentrations outside the population-based reference range likely form a heterogeneous group both including individuals with abnormal TSH concentrations due to mild thyroid disease, as well as non-diseased individuals with a unique HPT-axis setpoint at the extremes of the distribution [7]. While currently used wide population-based TSH reference ranges enable easily the diagnosis of overt primary thyroid dysfunction, narrower personalized reference ranges seem crucial for an accurate diagnosis of mild thyroid dysfunction, as applying wide population-based reference ranges to an individual patient may lead to incorrect diagnoses and related over- and undertreatment.

Multiple large-scale observational studies reported associations between mild thyroid dysfunction (*i.e.*, subclinical hypo- and hyperthyroidism) and an increased risk of various adverse health outcomes, such as cardiovascular diseases, depression and mortality [8-16]. However, well-powered randomized clinical trials are lacking [17] with international guidelines still being inconclusive whether subclinical thyroid dysfunction should be treated or not [18-20]. An individualized approach to treatment of subclinical thyroid dysfunction is often advised, implying a clear need for personalized TSH reference ranges [20].

Unfortunately, except for childhood and pregnancy, TSH reference ranges used in clinical practice do not take any individual patient characteristics into account. Several environmental and individual factors have been implied to influence TSH concentrations, including iodine intake, age, sex, body-mass index (BMI), drugs and tobacco smoking [21]. Nonetheless, twin studies have demonstrated that genetic factors are the major determinants of thyroid function in the general population, being responsible for up to an estimated 65% of the inter-individual variation in TSH and TH concentrations [22]. Over the last 20 years, candidate gene and genome-wide association studies (GWAS) have identified dozens of genetic variants regulating variation in reference-range TSH concentrations [23]. However, to date no attempts have been made to use these genetic variants to personalize TSH reference ranges, nor has this been done for other (endocrine) laboratory measurements.

Therefore, in this study we used a polygenic score (PGS) to calculate genetically-determined TSH reference ranges in a thyroid disease-free cohort and assessed its effects on reclassification of diagnoses when compared to using population-based reference ranges. Next, results were validated in a second independent population-based cohort, after which potential clinical implications were assessed in a third independent population-based cohort including individuals without thyroid disease as well as individuals on levothyroxine (LT4) treatment.

Materials and methods

Participants

The first part of this study was performed in two Dutch population-based cohorts: the Rotterdam Study was used for discovery analyses and the Nijmegen Biomedical Study was used for validation. Design and objectives of these two studies have been previously described in detail elsewhere [24,25]. The Rotterdam Study has been approved by the Medical Ethics Committee of the Erasmus Medical Center (registration number MEC 02.1015). The Nijmegen Biomedical Study has been approved by the Radboud University Medical Center Institutional Review Board (registration number CMO 2001/055). All participants signed informed consent for participation and the use of data in research. Participants with available thyroid function tests including TSH, FT4, thyroid peroxidase antibody (TPOAb) measurements and genotyping data were identified for analyses. Clinical characteristics were collected regarding age, sex, BMI and smoking status. Individuals aged < 18 years, with reported thyroid disease, thyroid surgery (if available), using thyroid medications (*i.e.,* levothyroxine, thiamazole, carbimazole, or propylthiouracil), or with TPOAb positivity according to the assay manufacturer's cut-off value, were excluded from all analyses. In total, 6,834 participants from the Rotterdam Study and 3,800 participants from the Nijmegen Biomedical Study were eligible for analyses (**Figure 1**).

The second part of this study was performed in the Trøndelag Health Study (HUNT) cohort, which did not participate in the GWAS on reference-range TSH concentrations by Teumer *et al.* [26]. The Trøndelag Health Study is a longitudinal, repeatedly surveyed, population-based health study conducted in the Nord-Trøndelag region, Norway, since 1984 [27,28]. Participation in the Trøndelag Health Study is based on informed consent and the study has been approved by the Norwegian Data Protection Authority and the Regional Committee for Medical and Health Research Ethics in Central Norway (registration number 2015/584). In this study, we

included participants from the HUNT2 survey cohort with available information on TSH concentrations and genotyping data, that were selected for TSH measurements (i.e. all females born in 1955 or earlier, a random selection of 50% of males born in 1955 or earlier and a random selection of 5% of males and females born in 1956 or later; please see: https://hunt-db.medisin.ntnu.no/hunt-db/variable/7238). All participants were aged 18 years or older. We used self-reported information on LT4 use to identify individuals on LT4 treatment. Individuals with reported thyroid surgery, radioiodine treatment and past or present use of carbimazole were excluded from the analyses to ensure the primary diagnosis of non-iatrogenic hypothyroidism in individuals on LT4 treatment. In total, 1,132 individuals on LT4 treatment and 26,321 individuals without thyroid disease were included in the study. FT4 or TPOAb measurements were only available in small subsamples and therefore not used for the analyses in HUNT [29].

The research was completed in accordance with the Declaration of Helsinki as revised in 2013.

Thyroid function measurements

In the Rotterdam Study, non-fasting serum samples collected from the participants were stored at -80C and thyroid function tests (TSH, FT4, TPOAb) were later performed using the same electrochemiluminescence immunoassay (Roche, Mannheim, Germany) in all participants; TPOAb concentrations greater than 35 kU/mL were regarded as positive, according to the assay manufacturer's recommendations. In the Nijmegen Biomedical Study, the same parameters (TSH, FT4, TPOAb) were measured in non-fasting serum samples using an immunoluminometric assay on a random-access analyzer for TSH (Architect; Abbott Diagnostics Division), a luminescence enzyme immunoassay on a random-access assay system for FT4 (Vitros ECI; Ortho Clinical Diagnostics), and a fluorescence immunoenzymometric assay (AxSYM Anti-TPO; Abbott Diagnostics Division) for TPOAb.

TPOAb concentrations greater than 12 kU/mL were regarded as positive, according to the assay manufacturer's recommendations. TSH measurements in the Trøndelag Health Study have been described previously [29]; TSH was measured using DELFIA hTSH Ultra from Wallac Oy (Turku, Finland).

After excluding individuals with reported thyroid disease, taking thyroid medications and/or TPOAb-positivity (in the Rotterdam Study and the Nijmegen Biomedical Study cohorts), cohort-specific population-based reference ranges for TSH in the Rotterdam Study, the Nijmegen Biomedical Study and the Trøndelag Health Study, as well as cohort-specific population-based reference ranges for FT4 in the Rotterdam Study and the Nijmegen Biomedical Study, were constructed using the 2.5^{th} and the 97.5^{th} percentiles for each trait. For TSH, population-based reference ranges were 0.43-5.11 mU/L in the Rotterdam Study, 0.35-3.82 mU/L in the Nijmegen Biomedical Study, and 0.51-5.20 mU/L in the Trøndelag Health Study. Population-based reference ranges for FT4 were 11.97-20.12 pmol/L in the Rotterdam Study and 9.84-18.10 pmol/L in the Nijmegen Biomedical Study. In total, 6,501, 3,613 and 25,042 individuals in the Rotterdam Study, the Nijmegen Biomedical Study, and the Trøndelag Health Study, respectively, had TSH concentrations within the population-based reference range.

Genotyping

Genotyping procedures in all three cohorts have been described in the **Supplementary Materials and Methods**, and in detail elsewhere [24,25,27,28].

Polygenic Score

Sixty-one single nucleotide polymorphisms (SNPs) associated with reference-range TSH concentrations at a genome-wide significance level (p-value $<5x10^{-8}$) were identified based on the results of a large-scale meta-analysis of GWAS on reference-range thyroid function by

Teumer *et al.* [26]. SNPs unavailable for the analyses in the studied cohorts were replaced by a proxy variant (r^2>0.8 in the 1000 Genomes Project European population) whenever possible (**Supplementary Table 1**). Two SNPs, rs200574439 (nearest gene *NKX2-3*) and rs8176645 (nearest gene *ABO*) were left out of the final analysis due to unavailability in the Rotterdam Study cohort with no available proxies. In total, 59 independent SNPs (**Supplementary Table 1**) were used in all cohorts to calculate a weighted polygenic score (*PGS*) for TSH concentrations for every individual, defined as a weighted sum of the number (dosage, d_i) of risk alleles of the analyzed SNPs, with weights (w_i) for each SNP corresponding to the beta estimates from the regression analysis on reference-range TSH concentrations, derived from the summary statistics of the GWAS by Teumer *et al.* [26], as illustrated bellow:

$$PGS = w_1 d_1 + \ldots + w_i d_i$$

The total score was then rescaled to a range between 0 and 100 by dividing the difference between individual and cohort-specific minimal PGS by the difference between a cohort-specific maximal and minimal PGS, and multiplying the product by 100.

Statistical analyses

After exclusion of individuals aged <18 years, with reported thyroid disease or surgery, thyroid medication use and/or TPOAb-positivity, all participants within the Rotterdam Study and the Nijmegen Biomedical Study cohorts were divided into four equal groups (quartiles: Q1, Q2, Q3 and Q4) by their PGS. Next, PGS-quartile-specific reference ranges for TSH concentrations in the Rotterdam Study and the Nijmegen Biomedical Study were calculated, defined as the 2.5th and the 97.5th percentiles. To distinguish whether differences in genetically-determined TSH concentrations reflect HPT-axis setpoint effects or thyroid disease, linear regression analyses were performed evaluating the association between the PGS and reference-range TSH and FT4 concentrations after inverse normal transformation

(TSH_int and FT4_int, respectively). The Mann–Whitney U test was used to directly compare median TSH and FT4 concentrations between subsequent PGS quartiles in each cohort. Moreover, we used a linear regression analysis to assess the relationship between log-transformed TSH and FT4 concentrations in individuals from each PGS quartile to further assess the effects of the genetically-determined TSH concentrations on the HPT-axis setpoint. A linear regression analysis was also used to investigate associations between PGS and TPOAb concentrations below the positivity cut-off level in the Rotterdam Study and the Nijmegen Biomedical Study cohorts. Subsequently, we investigated the effects of applying PGS-quartile-specific, instead of population-based TSH reference ranges on the reclassification of individual thyroid status (i.e. the diagnosis of (subclinical) hypothyroidism, euthyroidism or (subclinical) hyperthyroidism). Finally, we evaluated the impact of genetically-determined TSH concentrations on treatment decisions by assessing the number of individuals on LT4 treatment in each PGS quartile in the Trøndelag Health Study cohort. The Cochran-Armitage test for trend was used to determine whether there was a significant difference in proportion of individuals on LT4 treatment between PGS quartiles.

Results

A flow diagram of study participants is shown in **Figure 1** and clinical characteristics of the study cohorts are provided in **Table 1**.

Genetically-determined TSH reference ranges in two independent populations

After excluding individuals with age <18 years, TPOAb-positivity, reported thyroid disease or surgery, and/or taking thyroid medications, the Rotterdam Study cohort was stratified into four quartiles based on the PGS. Next, PGS-quartile-specific TSH reference ranges (2.5^{th} – 97.5^{th} percentiles) were calculated. As illustrated in **Figure 2A**, PGS-quartile-specific TSH reference ranges differed from the population-based reference ranges, while there were also evident differences between PGS quartiles (0.28-3.98 mU/L in Q1 vs. 0.81-6.16 mU/L in Q4). To exclude cohort-specific effects and assay differences, these analyses were repeated in an independent cohort (the Nijmegen Biomedical Study) with a different TSH assay, which showed similar results (PGS-quartile-specific TSH reference ranges of 0.20-2.94 mU/L in Q1 vs. 0.53-4.48 mU/L in Q4; **Figure 2B**).

Relationship between genetically-determined TSH concentrations and the HPT-axis setpoint

Next, we verified the associations between the PGS, TSH and FT4 concentrations. Simple linear regression models indicated strong associations between the PGS and TSH concentrations within the population-based reference range in the Rotterdam Study (β=0.020 SD, $P=2x10^{-168}$) and the Nijmegen Biomedical Study (β=0.016 SD, $P=9x10^{-78}$), while associations between the PGS and FT4 concentrations were much weaker in both cohorts (β= -0.004 SD, $P=5x10^{-8}$ and β=-0.005 SD, $P=1x10^{-6}$ in the Rotterdam Study and the Nijmegen Biomedical Study, respectively). Direct comparisons of individuals from subsequent PGS

quartiles in both cohorts showed significant differences in median TSH concentrations, while the differences in median FT4 concentrations were absent or minor (**Table 2**).

Furthermore, multiple linear regression models showed that the PGS is by far the strongest predictor of individual TSH concentrations in both cohorts, a much stronger predictor than FT4 concentrations (total variance in TSH concentrations explained 9.2-11.1% vs. 2.4-2.7%, respectively) or other non-genetic factors including age, sex, BMI or smoking status (total variance in TSH concentrations explained 0.2-1.8%; **Table 3**).

Subsequently, while all TPOAb-positive individuals were already excluded in our study, we for completeness also tested for associations with TPOAb concentrations below the positivity cut-off level to rule out that increasing TSH reference range upper limits in subsequent PGS quartiles were driven by early stages of autoimmune hypothyroidism. These analyses showed no associations between the PGS and TPOAb concentrations in either the Rotterdam Study (P=0.56) or the Nijmegen Biomedical Study cohort (P=0.55).

Finally, as the effects of the PGS were much more pronounced on TSH compared to FT4, we verified this observation by linear regression analyses investigating the relationships between log-transformed TSH (TSH_log) and FT4 concentrations in individuals stratified by PGS quartiles (**Supplementary Table 2**). These results are illustrated in **Figure 3A** for the Rotterdam Study, indicating that at the same FT4 concentration, individuals in a higher PGS quartile had a higher TSH concentration. These findings were also replicated in the Nijmegen Biomedical Study cohort (**Figure 3B**). Together with all previous findings, this confirms a HPT-axis setpoint effect, *i.e.* an upward shift in TSH concentrations with increasing PGS quartiles at similar FT4 concentrations.

Diagnostic consequences (reclassification of diagnoses)

Given the observed effects of PGS quartiles on TSH reference ranges, we next assessed the diagnostic consequences of applying such personalized reference ranges. The application of

PGS-quartile-specific instead of population-based reference ranges led to a significant reclassification of thyroid status in the Rotterdam Study and the Nijmegen Biomedical Study cohorts: 24.7% and 24.5% of individuals classified as having (subclinical) hypothyroidism when applying population-based TSH reference ranges in the Rotterdam Study and the Nijmegen Biomedical Study, respectively, were reclassified as euthyroid (**Figure 4** and **Table 4**). Similarly, 30.1% of individuals previously classified as having (subclinical) hyperthyroidism when applying population-based TSH reference ranges in the Rotterdam Study and the Nijmegen Biomedical Study were reclassified as euthyroid (**Figure 4** and **Table 4**). A comparable number (but a smaller proportion) of individuals classified as euthyroid when using the population-based TSH reference ranges were reclassified as having (subclinical) hypothyroidism (0.6% and 0.7% in the Rotterdam Study and the Nijmegen Biomedical Study cohorts, respectively) and (subclinical) hyperthyroidism (0.8% and 0.7% in the Rotterdam Study and the Nijmegen Biomedical Study cohorts, respectively) when applying PGS-quartile-specific TSH reference ranges (**Figure 4** and **Table 4**). No sex-differences in terms of reclassification of diagnosis after applying genetically-determined reference ranges for TSH levels were observed in the Rotterdam Study cohort, while in the Nijmegen Biomedical Study cohort more females than males were reclassified from being euthyroid to having (subclinical) hypothyroidism (0.4% vs. 1.0%, P=0.03, **Supplementary Table 3**), and more males than females were reclassified from having (subclinical) hyperthyroidism to being euthyroid (41.8% vs. 20.0%, P=0.02, **Supplementary Table 3**).

Clinical consequences

Finally, we assessed the potential clinical impact of using genetically-determined TSH reference ranges in a third independent cohort of participants from the Trøndelag Health Study with available information on LT4 use. First, we verified the effects of the PGS on TSH reference ranges and diagnostic reclassification in the Trøndelag Health Study in thyroid

disease-free individuals, which showed similar results as in the Rotterdam Study and the Nijmegen Biomedical Study (**Supplementary Figure 1**, **Supplementary Tables 4 & 5**). Next, we evaluated the effects of genetically-determined TSH concentrations on treatment decisions in a large group of LT4 users (N=1,132) by assessing the number of patients on LT4 treatment in each PGS quartile. As illustrated in **Figure 5,** individuals from the higher PGS quartiles had a higher probability of being prescribed LT4 treatment compared to individuals from the lower PGS quartiles (3.3% in Q1 vs. 5.2% in Q4, $P_{for\ trend} = 1.7 \times 10^{-8}$).

Discussion

This is the first study using a PGS to predict genetically-determined TSH reference ranges. We showed substantial differences between genetically-determined and population-based reference ranges, with consistent findings across all three independent populations. For example, we showed in the Rotterdam Study that TSH concentrations of 6.0 mU/L can be regarded as normal in 25% of the individuals from the general population with the highest polygenic scores (*i.e.*, PGS Q4), whereas a TSH of 4.2 mU/L would already be regarded as abnormal for individuals in PGS Q1. If genetic testing were to become routinely available in practice, calculating a PGS to establish a genetically-determined reference range for an individual patient would inform a clinician if the observed TSH concentration is adequate or not for this specific patient, enabling personalized treatment decisions. Importantly, these analyses were carried out in TPOAb-negative individuals, while sensitivity analyses showed that there were neither any associations with TPOAb concentrations below the TPOAb-positivity cut-off. Furthermore, the effects of PGS on TSH concentrations were not accompanied by a proportional difference in FT4 concentrations. Indeed, regression analyses showed that the PGS is by far the strongest predictor of individual TSH concentrations, compared to FT4 concentrations or other non-genetic factors including age, sex, BMI or smoking status. Finally, further analyses illustrated that for an identical FT4 concentration, TSH concentrations increased with increasing PGS quartiles. These findings are also in line with the results of the GWAS on reference-range TSH and FT4 concentrations, which revealed a very limited genetic overlap between these two traits [26]. Taken together, all of these findings strongly point to a genetically-determined HPT-axis setpoint effect, and not to an enrichment of disease causing genetic variants in higher PGS quartiles. This is obviously a key finding when considering its use in clinical practice.

Application of the genetically-determined TSH reference ranges led to a reclassification of the thyroid status to euthyroidism in up to 25-30% of the individuals that are diagnosed with (subclinical) hypo- and hyperthyroidism when using the population-based TSH reference ranges. This finding could add to the understanding of the large heterogeneity in clinical presentation and treatment efficacy observed among patients diagnosed with subclinical thyroid dysfunction.

Improved diagnosis of thyroid dysfunction has become even more important nowadays, since TSH is one of the most frequently ordered tests in everyday clinical practice [30]. This is because thyroid dysfunction is often accompanied by non-specific complaints which are common in the general population, such as tiredness and weight changes, for which TSH testing is part of the diagnostic work-up [31-33]. Whereas LT4 is among the most commonly prescribed drugs [34], a large community-based study in the UK by Taylor *et al.* [35], as well as several other studies [36-38], showed that the TSH threshold to treat subclinical hypothyroidism has lowered, with most of the LT4 prescribed patients having only a mildly elevated TSH concentration (5-6 mU/L) at the time of the index prescription. While our analyses suggest that individuals in higher PGS quartiles have higher TSH concentrations due to a setpoint effect instead of having thyroid disease, we also show that there is a significant overrepresentation of individuals with higher PGS among the LT4 users. Specifically, individuals in Q4 of the PGS in the Trøndelag Health Study (HUNT) cohort had a 5.2% probability of being prescribed LT4 compared to 3.3% for individuals in Q1. This suggests that the genetically-determined higher TSH concentrations in these individuals might have incorrectly led to LT4 initiation. This is worrisome as LT4 is seldom stopped once initiated and many LT4 users have a suppressed TSH with an increased risk of cardiovascular complications and fractures [39-41].

Furthermore, in the context of our findings it is also noteworthy that 5-10% of patients diagnosed with hypothyroidism have persistent complaints despite biochemical euthyroidism

on LT4 treatment, which is a large unresolved knowledge gap in endocrinology [42]. Part of this could well be explained by initial incorrect diagnoses or suboptimal treatment due to the use of the wide population-based TSH reference ranges. The use of genetically-determined, narrow TSH reference ranges could allow for a better classification of thyroid status, and thereby more tailored therapies as well as prevention of unnecessary therapies.

A limitation of our study is that only single TSH and FT4 measurements were performed, and we cannot exclude that some were affected by (transient) interfering factors. As the currently used PGS can only predict approximately 10% of inter-individual variation in TSH concentrations, further improvements should be made when more genetic variants determining TSH concentrations are identified in future studies. Furthermore, while the majority of the inter-individual variation in TSH concentrations is determined by genetic factors, also some non-genetic factors (*e.g.* age, sex, BMI) could have a (modest) contribution to the personalized TSH reference range. However, despite the differences in TSH concentrations between the analyzed cohorts (potentially attributed to differences in clinical characteristics and assays used), our findings were consistent across all three independent populations. Nevertheless, as we only included individuals from European ancestries, our findings cannot be directly extrapolated to other ancestries. However, our study can serve as a blueprint for similar studies in other populations, after genetic factors determining TSH levels in non-European populations are established in dedicated GWAS. In mixed populations consisting of individuals with diverse ancestries, using a multiethnic PGS might be required [43,44].

In conclusion, this is the first study using individual genetic profiles to personalize TSH reference ranges. Our findings were consistent across three large independent cohorts, with large effects on diagnosis reclassification and LT4 prescription behavior. Future studies

should investigate whether addition of more genetic variants and non-genetic factors could further refine its predictive ability.

Acknowledgements

The Rotterdam Study is funded by Erasmus Medical Center and Erasmus University, Rotterdam, Netherlands Organization for the Health Research and Development (ZonMw), the Research Institute for Diseases in the Elderly (RIDE), the Ministry of Education, Culture and Science, the Ministry for Health, Welfare and Sports, the European Commission (DG XII), and the Municipality of Rotterdam. The authors are grateful to the study participants, the staff from the Rotterdam Study and the participating general practitioners and pharmacists. The Rotterdam Study has been approved by the Medical Ethics Committee of the Erasmus MC (registration number MEC 02.1015) and by the Dutch Ministry of Health, Welfare and Sport (Population Screening Act WBO, license number 1071272-159521-PG). The Rotterdam Study has been entered in the Netherlands National Trial Register (NTR; http://www.trialregister.nl) and in the WHO International Clinical Trials Registry Platform (ICTRP; http://www.who.int/ictrp/network/primary/en/) under shared catalogue number NTR6831. All participants provided written informed consent to participate in the study and to have their information obtained from treating physicians. The generation and management of GWAS genotype data for the Rotterdam Study (RS I, RS II, RS III) was executed by the Human Genotyping Facility of the Genetic Laboratory of the Department of Internal Medicine, Erasmus MC, Rotterdam, The Netherlands. The GWAS datasets are supported by the Netherlands Organization of Scientific Research NWO Investments (nr. 175.010.2005.011, 911-03-012), the Genetic Laboratory of the Department of Internal Medicine, Erasmus MC, the Research Institute for Diseases in the Elderly (014-93- 015; RIDE2), the Netherlands Genomics Initiative (NGI)/Netherlands Organisation for Scientific Research (NWO) Netherlands Consortium for Healthy Aging (NCHA), project nr. 050-060-

810. We thank Pascal Arp, Mila Jhamai, Marijn Verkerk, Lizbeth Herrera and Marjolein Peters, MSc, and Carolina Medina-Gomez, MSc, for their help in creating the GWAS database, and Karol Estrada, PhD, Yurii Aulchenko, PhD, and Carolina Medina-Gomez, MSc, for the creation and analysis of imputed data. This study makes use of sequence reference data generated by the UK10K Consortium, derived from samples from the ALSPAC and TwinsUK datasets. A full list of the investigators who contributed to the generation of the data is available from www.UK10K.org. Funding for UK10K was provided by the Wellcome Trust under award WT091310. We thank Jie Huang at the Wellcome Trust's Sanger Institute, at Hinxton, U.K. for the creation of imputed data, with the support of Marijn Verkerk, Carolina Medina-Gomez, MSc, and Anis Abuseiris and their input for the analysis setup. This study makes use of an extended dataset of RS II and RS III samples based on Illumina Omni 2.5 and 5.0 GWAS genotype data, imputed to 1KG using the two-phase imputation method. This dataset was funded by the Genetic Laboratory of the Department of Internal Medicine, the department of Forensic Molecular Biology, and the department of Dermatology, Erasmus MC, Rotterdam, The Netherlands. We thank Linda Broer, PhD, for the creation of imputed data, with the support of Marijn Verkerk and Carolina Medina-Gomez, MSc, for the analysis setup. We would like to thank Karol Estrada PhD, Fernando Rivadeneira PhD, and Anis Abuseiris (Erasmus MC Rotterdam, The Netherlands), for their help in creating GRIMP, and we thank BigGRID for access to their grid computing resources .

The Nijmegen Biomedical Study is a population-based survey conducted at the Department for Health Evidence and the Department of Laboratory Medicine of the Radboud university medical center. Principal investigators of the Nijmegen Biomedical Study are L.A.L.M. Kiemeney, A.L.M. Verbeek, D.W. Swinkels en B. Franke.

The Trøndelag Health Study (HUNT) is a collaboration between HUNT Research Centre (Faculty of Medicine and Health Sciences, NTNU, Norwegian University of Science and Technology), Trøndelag County Council, Central Norway Regional Health Authority, and the

Norwegian Institute of Public Health. The genotyping in the HUNT Study was financed by the National Institutes of Health (NIH), University of Michigan, The Research Council of Norway, and Central Norway Regional Health Authority and the Faculty of Medicine and Health Sciences, NTNU. The genotype quality control and imputation have been conducted by the K.G. Jebsen Center for Genetic Epidemiology, Department of Public Health and Nursing, Faculty of Medicine and Health Sciences, NTNU.

Authors Contribution Statement

Aleksander Kuś: conceptualization (lead), methodology (lead), statistical analysis (lead); interpretation of the results (equal), writing of the original draft (lead); Rosalie B.T.M. Sterenborg: conceptualization (lead), methodology (lead), statistical analysis (lead); interpretation of the results (equal), writing of the original draft (supporting); Eirin B. Haug: conceptualization (supporting), statistical analysis (lead); interpretation of the results (equal), review and editing of the original draft (equal); Tessel E. Galesloot: conceptualization (supporting), methodology (supporting), statistical analysis (supporting); interpretation of the results (equal), review and editing of the original draft (equal); W. Edward Visser: conceptualization (supporting), interpretation of the results (equal), review and editing of the original draft (equal); Johannes W.A. Smit: conceptualization (supporting), interpretation of the results (equal), review and editing of the original draft (equal); Tomasz Bednarczuk: conceptualization (supporting), interpretation of the results (equal), review and editing of the original draft (equal); Robin P. Peeters: conceptualization (supporting), interpretation of the results (equal), review and editing of the original draft (equal); Bjørn O. Åsvold: conceptualization (supporting), interpretation of the results (equal), review and editing of the original draft (equal); Alexander Teumer: conceptualization (supporting), methodology (supporting), statistical analysis (supporting); interpretation of the results (equal), review and editing of the original draft (equal); Marco Medici: conceptualization (lead), methodology

(lead), interpretation of the results (equal), writing of the original draft (supporting), overall supervision (lead).

Author Disclosure Statement

No competing financial interests exist for any of the authors.

Funding Statement

Aleksander Kuś was supported by the Exchange in Endocrinology Expertise (3E) program of the European Union of Medical Specialists (UEMS), Section and Board of Endocrinology. Marco Medici was supported by funding from the European and American Thyroid Associations, the Erasmus University Rotterdam, and the Dutch Organization for Scientific Research (NWO). Rosalie B.T.M. Sterenborg, Eirin B. Haug, Tessel E. Galesloot, W. Edward Visser, Johannes W.A. Smit, Tomasz Bednarczuk, Robin P. Peeters, Bjørn O. Åsvold, and Alexander Teumer have no funding information to declare.

References

1. Taylor PN, Albrecht D, Scholz A, et al. Global epidemiology of hyperthyroidism and hypothyroidism. Nature Reviews Endocrinology 2018;14(5):301-316, doi:10.1038/nrendo.2018.18
2. Chaker L, Bianco AC, Jonklaas J, et al. Hypothyroidism. Lancet 2017;390(10101):1550-1562, doi:10.1016/S0140-6736(17)30703-1
3. De Leo S, Lee SY, Braverman LE. Hyperthyroidism. Lancet 2016;388(10047):906-918, doi:10.1016/S0140-6736(16)00278-6

4. Razvi S, Bhana S, Mrabeti S. Challenges in Interpreting Thyroid Stimulating Hormone Results in the Diagnosis of Thyroid Dysfunction. Journal of Thyroid Research 2019;2019:4106816, doi:10.1155/2019/4106816
5. Andersen S, Pedersen KM, Bruun NH, et al. Narrow individual variations in serum T(4) and T(3) in normal subjects: a clue to the understanding of subclinical thyroid disease. The Journal of Clinical Endocrinology and Metabolism 2002;87(3):1068-1072, doi:10.1210/jcem.87.3.8165
6. Browning MC, Ford RP, Callaghan SJ, et al. Intra- and interindividual biological variation of five analytes used in assessing thyroid function: implications for necessary standards of performance and the interpretation of results. Clinical Chemistry 1986;32(6):962-966
7. Cooper DS, Biondi B. Subclinical thyroid disease. Lancet 2012;379(9821):1142-1154, doi:10.1016/S0140-6736(11)60276-6
8. Rodondi N, den Elzen WP, Bauer DC, et al. Subclinical hypothyroidism and the risk of coronary heart disease and mortality. JAMA 2010;304(12):1365-1374, doi:10.1001/jama.2010.1361
9. Chaker L, Baumgartner C, den Elzen WP, et al. Subclinical Hypothyroidism and the Risk of Stroke Events and Fatal Stroke: An Individual Participant Data Analysis. The Journal of Clinical Endocrinology and Metabolism 2015;100(6):2181-2191, doi:10.1210/jc.2015-1438
10. Rieben C, Segna D, da Costa BR, et al. Subclinical Thyroid Dysfunction and the Risk of Cognitive Decline: a Meta-Analysis of Prospective Cohort Studies. The Journal of Clinical Endocrinology and Metabolism 2016;101(12):4945-4954, doi:10.1210/jc.2016-2129
11. Loh HH, Lim LL, Yee A, et al. Association between subclinical hypothyroidism and depression: an updated systematic review and meta-analysis. BMC Psychiatry 2019;19(1):12, doi:10.1186/s12888-018-2006-2
12. Journy NMY, Bernier MO, Doody MM, et al. Hyperthyroidism, Hypothyroidism, and Cause-Specific Mortality in a Large Cohort of Women. Thyroid 2017;27(8):1001-1010, doi:10.1089/thy.2017.0063
13. Biondi B, Cappola AR, Cooper DS. Subclinical Hypothyroidism: A Review. JAMA 2019;322(2):153-160, doi:10.1001/jama.2019.9052

14. Collet TH, Gussekloo J, Bauer DC, et al. Subclinical hyperthyroidism and the risk of coronary heart disease and mortality. Archives of Internal Medicine 2012;172(10):799-809, doi:10.1001/archinternmed.2012.402
15. Blum MR, Bauer DC, Collet TH, et al. Subclinical thyroid dysfunction and fracture risk: a meta-analysis. JAMA 2015;313(20):2055-2065, doi:10.1001/jama.2015.5161
16. Hong JW, Noh JH, Kim DJ. Association between subclinical thyroid dysfunction and depressive symptoms in the Korean adult population: The 2014 Korea National Health and Nutrition Examination Survey. PloS One 2018;13(8):e0202258, doi:10.1371/journal.pone.0202258
17. Stott DJ, Rodondi N, Kearney PM, et al. Thyroid Hormone Therapy for Older Adults with Subclinical Hypothyroidism. The New England Journal of Medicine 2017;376(26):2534-2544, doi:10.1056/NEJMoa1603825
18. Pearce SH, Brabant G, Duntas LH, et al. 2013 ETA Guideline: Management of Subclinical Hypothyroidism. European Thyroid Journal 2013;2(4):215-228, doi:10.1159/000356507
19. 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
20. Sawka AM, Cappola AR, Peeters RP, et al. Patient Context and Thyrotropin Levels Are Important When Considering Treatment of Subclinical Hypothyroidism. Thyroid 2019;29(10):1359-1363, doi:10.1089/thy.2019.0494
21. Chaker L, Korevaar TI, Medici M, et al. Thyroid Function Characteristics and Determinants: The Rotterdam Study. Thyroid 2016;26(9):1195-1204, doi:10.1089/thy.2016.0133
22. Hansen PS, Brix TH, Sorensen TI, et al. Major genetic influence on the regulation of the pituitary-thyroid axis: a study of healthy Danish twins. The Journal of Clinical Endocrinology and Metabolism 2004;89(3):1181-1187, doi:10.1210/jc.2003-031641
23. Kus A, Chaker L, Teumer A, et al. The Genetic Basis of Thyroid Function: Novel Findings and New Approaches. The Journal of Clinical Endocrinology and Metabolism 2020;105(6):1707-1721, doi:10.1210/clinem/dgz225

24. Ikram MA, Brusselle GGO, Murad SD, et al. The Rotterdam Study: 2018 update on objectives, design and main results. European Journal of Epidemiology 2017;32(9):807-850, doi:10.1007/s10654-017-0321-4
25. Galesloot TE, Vermeulen SH, Swinkels DW, et al. Cohort Profile: The Nijmegen Biomedical Study (NBS). International Journal of Epidemiology 2017;46(4):1099-1100j, doi:10.1093/ije/dyw268
26. Teumer A, Chaker L, Groeneweg S, et al. Genome-wide analyses identify a role for SLC17A4 and AADAT in thyroid hormone regulation. Nature Communications 2018;9(1):4455, doi:10.1038/s41467-018-06356-1
27. Asvold BO, Langhammer A, Rehn TA, et al. Cohort Profile Update: The HUNT Study, Norway. International Journal of Epidemiology 2023;52(1):e80-e91, doi:10.1093/ije/dyac095
28. Brumpton BM, Graham S, Surakka I, et al. The HUNT study: A population-based cohort for genetic research. Cell Genomics 2022;2(10):100193, doi:10.1016/j.xgen.2022.100193
29. Bjoro T, Holmen J, Kruger O, et al. Prevalence of thyroid disease, thyroid dysfunction and thyroid peroxidase antibodies in a large, unselected population. The Health Study of Nord-Trondelag (HUNT). European Journal of Endocrinology 2000;143(5):639-647, doi:10.1530/eje.0.1430639
30. Javaid U, Kennedy D, Addison C, et al. Frequency, determinants and costs of thyroid function testing in a laboratory serving a large population. European Journal of Endocrinology 2022;186(5):553-560, doi:10.1530/EJE-21-1172
31. Kroenke K, Wood DR, Mangelsdorff AD, et al. Chronic fatigue in primary care. Prevalence, patient characteristics, and outcome. JAMA 1988;260(7):929-934
32. Baker R, Shaw EJ. Diagnosis and management of chronic fatigue syndrome or myalgic encephalomyelitis (or encephalopathy): summary of NICE guidance. BMJ 2007;335(7617):446-448, doi:10.1136/bmj.39302.509005.AE
33. Semlitsch T, Stigler FL, Jeitler K, et al. Management of overweight and obesity in primary care-A systematic overview of international evidence-based guidelines. Obesity Reviews 2019;20(9):1218-1230, doi:10.1111/obr.12889

34. Rodriguez-Gutierrez R, Maraka S, Ospina NS, et al. Levothyroxine overuse: time for an about face? The Lancet Diabetes & Endocrinology 2017;5(4):246-248, doi:10.1016/S2213-8587(16)30276-5
35. Taylor PN, Iqbal A, Minassian C, et al. Falling threshold for treatment of borderline elevated thyrotropin levels-balancing benefits and risks: evidence from a large community-based study. JAMA Internal Medicine 2014;174(1):32-39, doi:10.1001/jamainternmed.2013.11312
36. Medici BB, Nygaard B, la Cour JL, et al. Changes in Prescription Routines for Treating Hypothyroidism Between 2001 and 2015: An Observational Study of 929,684 Primary Care Patients in Copenhagen. Thyroid 2019;29(7):910-919, doi:10.1089/thy.2018.0539
37. Delemer B, Aubert JP, Nys P, et al. An observational study of the initial management of hypothyroidism in France: the ORCHIDEE study. European Journal of Endocrinology 2012;167(6):817-823, doi:10.1530/EJE-11-1041
38. Brito JP, Ross JS, El Kawkgi OM, et al. Levothyroxine Use in the United States, 2008-2018. JAMA Internal Medicine 2021;181(10):1402-1405, doi:10.1001/jamainternmed.2021.2686
39. 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. The Journal of Clinical Endocrinology and Metabolism 2010;95(1):186-193, doi:10.1210/jc.2009-1625
40. Mammen JS, McGready J, Oxman R, et al. Thyroid Hormone Therapy and Risk of Thyrotoxicosis in Community-Resident Older Adults: Findings from the Baltimore Longitudinal Study of Aging. Thyroid 2015;25(9):979-986, doi:10.1089/thy.2015.0180
41. Biondi B, Cappola AR. Subclinical hypothyroidism in older individuals. The Lancet Diabetes & Endocrinology 2022;10(2):129-141, doi:10.1016/S2213-8587(21)00285-0
42. Wiersinga WM. Paradigm shifts in thyroid hormone replacement therapies for hypothyroidism. Nature Rviews Endocrinology 2014;10(3):164-174, doi:10.1038/nrendo.2013.258
43. Marquez-Luna C, Loh PR, South Asian Type 2 Diabetes (SAT2D) Consortium, et al. Multiethnic polygenic risk scores improve risk prediction in diverse populations. Genetic Epidemiology 2017;41(8):811-823, doi:10.1002/gepi.22083

44. Patel AP, Wang M, Ruan Y, et al. A multi-ancestry polygenic risk score improves risk prediction for coronary artery disease. Nature Medicine 2023;29(7):1793-1803, doi:10.1038/s41591-023-02429-x

Figure legends

Figure 1. Flow diagram of study participants in the three study cohorts.

Figure 2. Polygenic Score (PGS) quartile-specific TSH reference ranges in two independent populations. PGS-quartile-specific TSH reference ranges are shown in red, and population-based reference ranges are shown in grey (Rotterdam Study: 0.43-5.11 mU/L; Nijmegen Biomedical Study: 0.35-3.82 mU/L). Solid horizontal lines correspond to median TSH concentrations in each PGS quartile.

Figure 3. The relationships between log-transformed TSH (TSH_log) and FT4 levels, stratified by polygenic score (PGS) quartiles. Each line corresponds to a linear regression analysis in individuals from a specific PGS quartile (Q1-Q4) in the Rotterdam Study (Figure 3A) and Nijmegen Biomedical Study (Figure 3B).

Figure 4. Reclassification of thyroid status after application of genetically-determined instead of population-based TSH reference range. Up to 25-30% of the individuals diagnosed with subclinical hypo- and hyperthyroidism when using the population-based TSH reference ranges have been reclassified as euthyroid based on their genetically-determined TSH reference range.

643 **Figure 5. Number of patients on LT4 treatment by polygenic score (PGS) quartile in the**
644 **Trøndelag Health Study (HUNT) cohort.** The red horizontal line indicates the expected
645 number of patients on LT4 treatment in each PGS quartile (total LT4 users / 4), when there
646 would be no association between the PGS quartile and LT4 use. *P*-value corresponds to the
647 Cochran-Armitage test for trend. Median TSH concentrations increased in subsequent PGS
648 quartiles: Q1: 1.3 mU/L; Q2: 1.5 mU/L; Q3: 1.6 mU/L; Q4: 1.9 mU/L.
649

Table 1. Clinical characteristics of the study cohorts.

Cohort	Rotterdam Study	Nijmegen Biomedical Study	Trøndelag Health Study (HUNT)
Number of individuals	6,834	3,800	26,321
Ethnicity	Caucasian	Caucasian	Caucasian
Iodine status	Sufficient	Sufficient	Sufficient
Sex distribution (males n,%)	3307 (48.4%)	1885 (49.6%)	8882 (33.7%)
Age (years)	65.2 (9.9)	54.8 (18.0)	57.5 (13.1)
BMI (kg/m^2)	27.2 (4.2)	25.1 (4.0)	26.7 (4.2)
Smoking status:			
- current	1,310 (19.2%)	1,277 (33.6%)	7,390 (28.1%)
- former	3,352 (49.0%)	1,655 (43.6%)	7,508 (28.5%)
- never	2,100 (30.7%)	856 (22.5%)	10,856 (41.2%)
- NA	72 (1.1%)	12 (0.3%)	567 (2.2%)
TSH (mU/L)	1.83 (0.43-5.11)	1.33 (0.35-3.82)	1.60 (0.51-5.20)
FT4 (pmol/L)	15.62 (11.97-20.12)	13.3 (9.84-18.10)	NA

Displayed numbers are after exclusion of individuals aged < 18 years, reported thyroid disease, thyroid surgery (if available), thyroid medications (*i.e*. levothyroxine, thiamazole, carbimazole, or propylthiouracil), and/or TPOAb positivity (in the Rotterdam Study and the Nijmegen Biomedical Study). Age and BMI are displayed as mean (SD). TSH and FT4 concentrations are displayed as median (cohort-specific 95% reference range).

Abbreviations: BMI, body mass index; FT4, free thyroxin; NA, not available; TSH, thyroid-stimulating hormone.

Table 2. TSH and FT4 levels in individuals stratified by Polygenic Score (PGS) quartiles. Median TSH and FT4 levels in individuals from each PGS quartile (Q1-Q4) in the Rotterdam Study and the Nijmegen Biomedical Study are shown for comparison.

	Rotterdam Study				**Nijmegen Biomedical Study**			
	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
Median TSH levels [mU/L]	1.44	1.73	2.00	2.31	1.04	1.27	1.43	1.62
Q1 vs Q2	$P = 4.2 \times 10^{-27}$				$P = 5.3 \times 10^{-14}$			
Q2 vs Q3	$P = 1.7 \times 10^{-13}$				$P = 5.9 \times 10^{-9}$			
Q3 vs Q4	$P = 2.7 \times 10^{-20}$				$P = 4.5 \times 10^{-8}$			
Median FT4 levels [pmol/L]	15.84	15.69	15.55	15.41	13.6	13.3	13.2	13.2
Q1 vs Q2	$P = 0.05$				$P = 3.8 \times 10^{-3}$			
Q2 vs Q3	$P = 0.10$				$P = 0.02$			
Q3 vs Q4	$P = 0.02$				$P = 1.00$			

Table 3. Multiple linear regression analyses on TSH concentrations in the Rotterdam Study and Nijmegen Biomedical Study cohorts.

	Rotterdam Study				Nijmegen Biomedical Study			
	Estimate	**Std. Error**	**P-value**	**Explained Variance****	**Estimate**	**Std. Error**	**P-value**	**Explained Variance****
(Intercept)	-0.5828	0.1079	6.83e-08	---	-0.8051	0.1010	2.16e-15	---
PGS	0.0198	0.0007	3.08e-164	0.1111	0.0161	0.0009	2.48e-74	0.0920
FT4*	-0.1122	0.0105	1.81e-26	0.0268	-0.0883	0.0144	9.13e-10	0.0243
TPOAb*	0.0008	0.0116	9.45e-01	0.0002	0.0467	0.0137	6.63e-04	0.0021
Sex (male)	-0.0433	0.0210	3.91e-02	0.0022	0.0733	0.0284	9.99e-03	0.0004
Age (years)	-0.0068	0.0011	2.44e-10	0.0049	-0.0074	0.0008	6.26e-19	0.0182
BMI (kg/m2)	0.0050	0.0025	4.45e-02	0.0021	0.0177	0.0035	4.64e-07	0.0028
Current smoking	-0.2055	0.0302	1.16e-11	0.0071	-0.1918	0.0373	2.82e-07	0.0072
Former smoking	-0.0647	0.0240	6.99e-03		-0.0636	0.0329	5.34e-02	

*effect estimates per 1 standard deviation (SD) change
** explained variance corresponding to r^2 from a linear regression model

Abbreviations: BMI, body mass index; FT4, free thyroxine; PGS, polygenic score; TPOAb, thyroid peroxidase antibody.

Table 4. Reclassification of diagnoses after applying genetically-determined instead of conventional population-based TSH reference ranges.

Diagnostic reclassification groups	Rotterdam Study	Nijmegen Biomedical Study
(Subclinical) hypothyroidism → Euthyroidism	42/170 (24.7%)	23/94 (24.5%)
Euthyroidism → (Subclinical) hypothyroidism	42/6501 (0.6%)	25/3613 (0.7%)
(Subclinical) hyperthyroidism → Euthyroidism	49/163 (30.1%)	28/93 (30.1%)
Euthyroidism → (Subclinical) hyperthyroidism	55/6501 (0.8%)	26/3613 (0.7%)

Figure 1

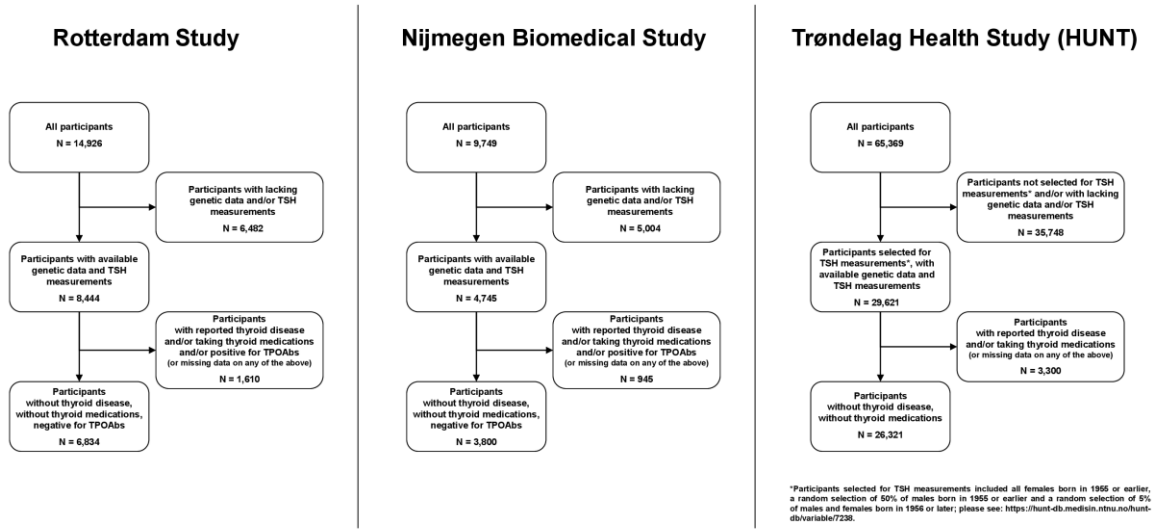


Figure 2

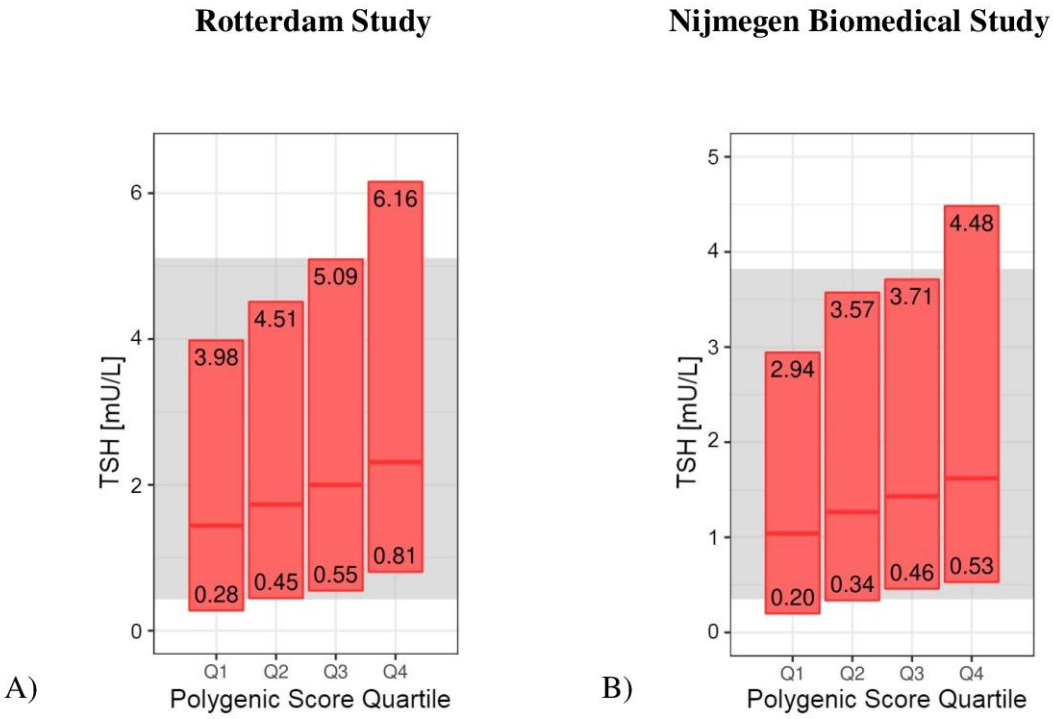


Figure 3

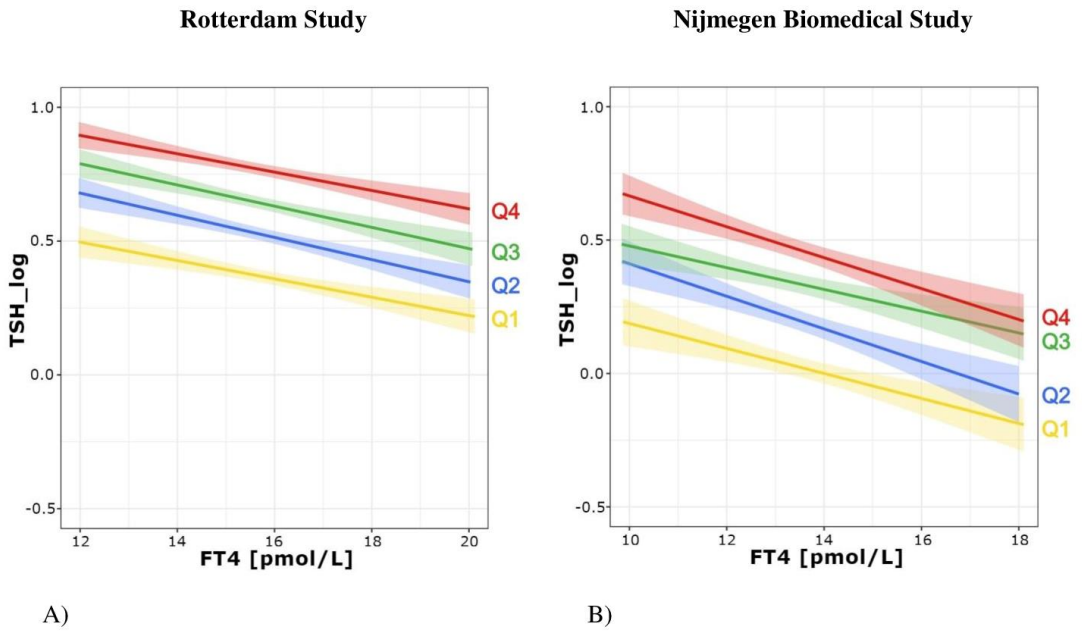


Figure 4

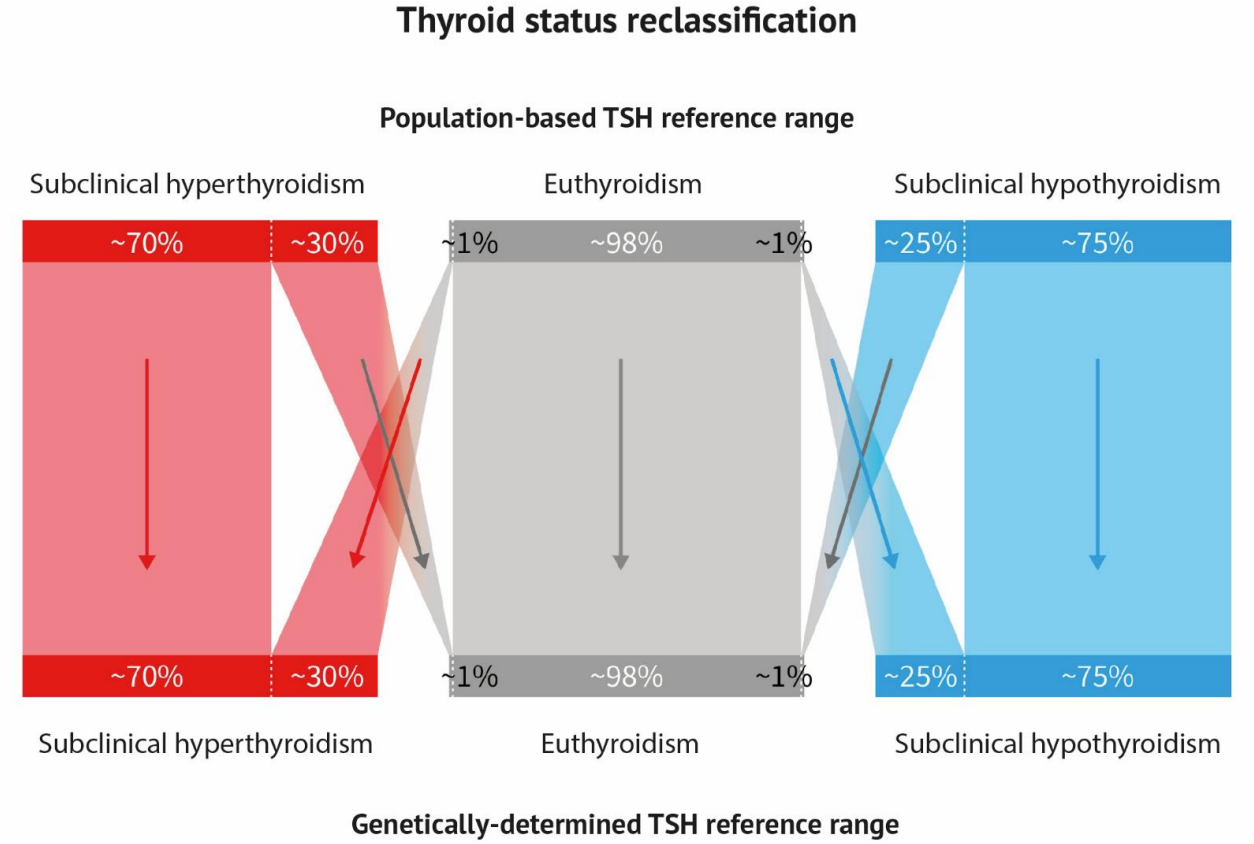
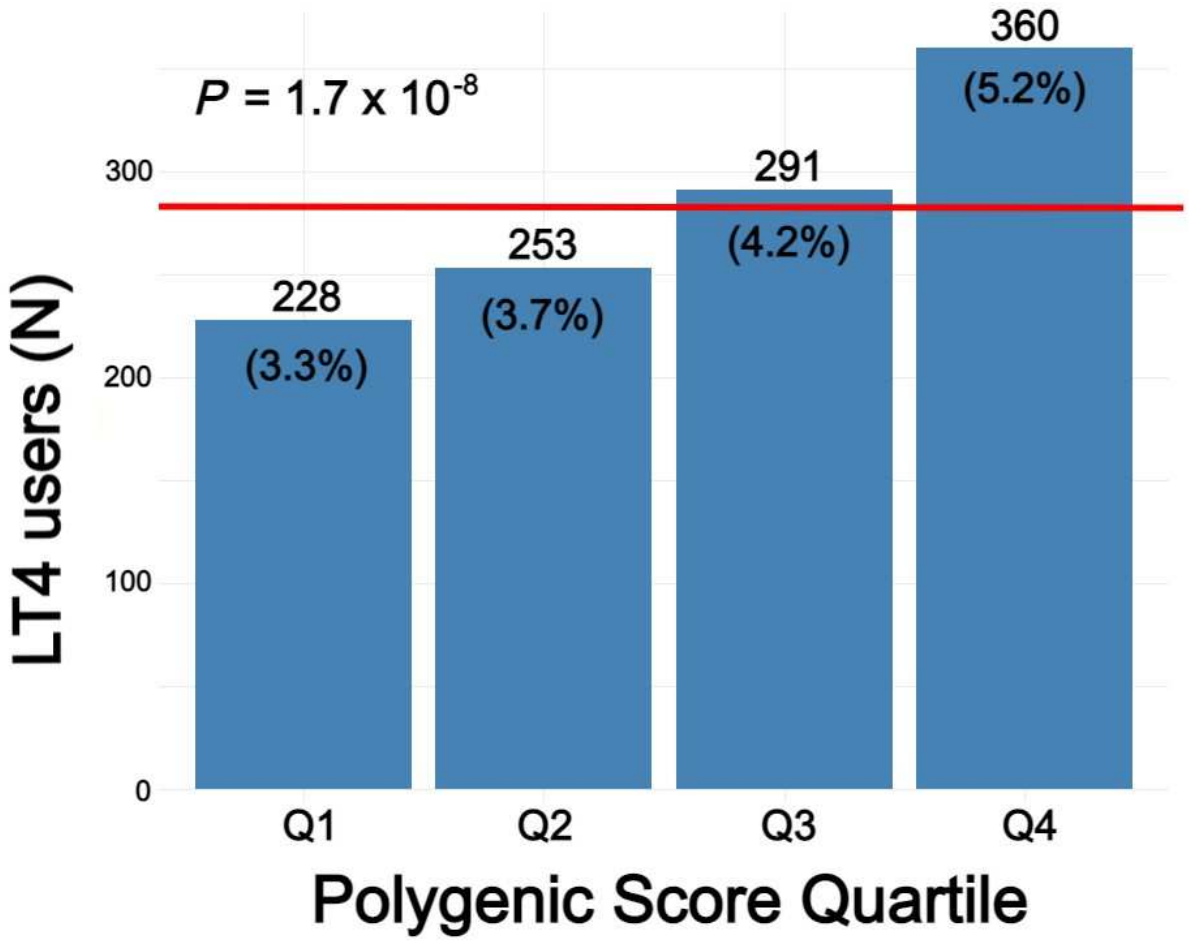


Figure 5


P = 1.7 x 10-8
360
(5.2%)
291
(4.2%)
253
(3.7%)
228
(3.3%)
300
200
100
0
LT4 users (N)
Q1
Q2
Q3
Q4
Polygenic Score Quartile

Supplementary Materials and Methods

Genotyping

Genotyping was conducted using the Illumina 550K and 610K arrays in the Rotterdam Study, and the llumina HumanOmniExpress-12 and -24 BeadChip arrays in the Nijmegen Biomedical Study, as described in detail elsewhere [1,2]. In both cohorts, participants with mismatch between genetically predicted and registered sex were excluded. In the Rotterdam Study, participants with excess autosomal heterozygosity, or recognized as being an outlier with identical-by-state clustering analysis were additionally excluded. SNP dosages were imputed with the reference panel from the 1000 Genomes Project [3] in the Rotterdam Study and combined together with Genome of The Netherlands (GoNL) in the Nijmegen Biomedical Study using MACH [4] and IMPUTE2 [5] software, respectively. In the Trøndelag Health Study, DNA samples were genotyped using Illumina HumanCoreExome v1.0 and 1.1, and imputed using Minimac3 with a merged reference panel of Haplotype Reference Consortium (HRC) [6] and whole genome sequencing data for 2201 samples from the Trøndelag Health Study [7]. All genomic positions were based on build 37 (GrCh37).

References

1. Galesloot TE, Vermeulen SH, Swinkels DW, et al. Cohort Profile: The Nijmegen Biomedical Study (NBS). International journal of epidemiology 2017;46(4):1099-1100j.
2. Ikram MA, Brusselle GGO, Murad SD, et al. The Rotterdam Study: 2018 update on objectives, design and main results. European Journal of Epidemiology 2017;32(9):807-850.
3. Genomes Project Consortium, Abecasis GR, Auton A, et al. An integrated map of genetic variation from 1,092 human genomes. Nature 2012;491(7422):56-65.
4. Li Y, Willer CJ, Ding J, et al. MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. Genetic Epidemiology 2010;34(8):816-834.

5. Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS Genetics 2009;5(6):e1000529.
6. McCarthy S, Das S, Kretzschmar W, et al. A reference panel of 64,976 haplotypes for genotype imputation. Nature Genetics 2016;48(10):1279-83.
7. Brumpton BM, Graham S, Surakka I, et al. The HUNT study: A population-based cohort for genetic research. Cell Genomics 2022;2(10):100193.

Supplementary Tables (STables) Index

Supplementary Table 1. Genetic variants associated with normal-range TSH concentrations included in the Polygenic Score (PGS).

Chromosome	Position	SNP	Gene	Effect_allele	Other_allele	Proxies RS / NBS / HUNT	EAF_GWAS_Teumer	Effect_TSH	StdErr_TSH	Pvalue_TSH	Effect_FT4	StdErr_FT4	P.value_FT4
1	19771438	rs12089835	CAPZB	t	c	---	0.3479	0.0725	0.0065	1.27E-28	-0.0167	0.0068	1.45E-02
1	19843576	rs10917469	CAPZB	a	g	---	0.8439	0.1112	0.0085	3.95E-39	-0.0289	0.0089	1.13E-03
1	19862320	rs74804879	CAPZB	t	c	HUNT: rs11801304, r2=0.7385, D'=1.00 in CEU	0.6486	0.0501	0.0065	1.22E-14	-0.0147	0.0069	3.35E-02
1	61610049	rs334725	NFIA	a	g	---	0.952	0.1737	0.0147	2.45E-32	-0.0577	0.0149	1.07E-04
1	108357391	rs17020122	VAV3	t	c	---	0.0852	0.1044	0.0114	5.32E-20	-0.0174	0.0118	1.42E-01
2	217580413	rs16856540	IGFBP5	t	c	---	0.8381	-0.0549	0.0084	7.81E-11	0.0048	0.0093	6.03E-01
2	217625523	rs13015993	IGFBP5	a	g	---	0.7333	0.0818	0.0069	4.52E-32	-0.0155	0.0077	4.27E-02
2	218236786	rs6724073	DIRC3	t	c	---	0.7406	0.0508	0.0079	1.35E-10	-0.0244	0.0087	5.23E-03
3	12239852	rs1663070	SYN2	t	c	---	0.7417	-0.0463	0.0070	3.49E-11	-0.0010	0.0074	8.89E-01
3	149220109	rs28502438	TM4SF4	t	c	---	0.5668	0.0338	0.0061	3.70E-08	-0.0088	0.0066	1.83E-01
3	185514088	rs13100823	IGF2BP2	t	c	---	0.3061	-0.0406	0.0066	6.76E-10	0.0208	0.0071	3.21E-03
3	193916181	rs59381142	HES1	a	g	---	0.243	-0.0580	0.0076	1.70E-14	0.0074	0.0079	3.53E-01
4	149587905	rs6535624	NR3C2	a	g	---	0.4409	0.0419	0.0062	1.60E-11	-0.0058	0.0067	3.87E-01
4	149665602	rs11732089	NR3C2	t	c	---	0.7961	0.1150	0.0076	1.73E-51	0.0047	0.0082	5.63E-01
5	76439961	rs62362610	PDE8B	c	g	---	0.083	0.0726	0.0118	7.73E-10	-0.0178	0.0125	1.56E-01
5	76488613	rs1119208	PDE8B	t	c	---	0.3666	0.0457	0.0064	6.65E-13	-0.0098	0.0068	1.50E-01
5	76495539	rs139424329	PDE8B	a	g	---	0.014	-0.2000	0.0322	5.14E-10	0.0190	0.0338	5.74E-01
5	76532571	rs2127387	PDE8B	a	g	---	0.4087	0.1435	0.0062	1.10E-117	-0.0318	0.0067	2.38E-06
5	76554807	rs7702192	PDE8B	a	c	---	0.4723	0.0697	0.0061	2.61E-30	-0.0123	0.0065	6.10E-02
5	76652403	rs113974964	PDE8B	t	c	---	0.0466	-0.1237	0.0146	2.06E-17	0.0191	0.0157	2.24E-01
5	76660193	rs139149784	PDE8B	a	g	---	0.0271	0.1556	0.0285	4.97E-08	0.0008	0.0303	9.79E-01
5	76773148	rs182873197	PDE8B	t	c	r2=1.00; D'=1.00 in CEU; HUNT: rs78676901, r2=1.0	0.0511	-0.0799	0.0142	1.71E-08	0.0126	0.0153	4.08E-01
6	31108129	rs1265091	PSORS1C1	t	c	RS: rs1063646, r2=0.96, D'=1.00 in CEU	0.202	0.0571	0.0086	3.20E-11	-0.0187	0.0094	4.69E-02
6	43805362	rs744103	VEGFA/LOC100132354	a	t	---	0.6909	0.0919	0.0069	6.73E-41	-0.0304	0.0073	3.31E-05
6	43905037	rs9381266	VEGFA/LOC100132354	t	c	---	0.7421	0.0726	0.0070	1.84E-25	-0.0207	0.0075	5.55E-03
6	148521292	rs9497965	SASH1	t	c	---	0.4007	0.0444	0.0062	9.81E-13	-0.0065	0.0067	3.27E-01
6	165973757	rs73022105	PDE10A	t	c	---	0.9548	0.1049	0.0155	1.20E-11	-0.0138	0.0167	4.06E-01
6	166047034	rs1079418	PDE10A	a	g	---	0.6877	0.1009	0.0066	8.23E-53	-0.0168	0.0071	1.90E-02
8	23356964	rs56009477	SLC25A37	a	g	---	0.8383	0.0524	0.0084	3.72E-10	-0.0155	0.0090	8.46E-02
8	32433013	rs2439301	NRG1	a	g	---	0.2331	-0.0587	0.0076	8.15E-15	0.0083	0.0081	3.08E-01
8	70365025	rs10957494	SULF1	a	g	---	0.6915	-0.0402	0.0066	1.10E-09	0.0087	0.0071	2.21E-01
8	133771635	rs118039499	TG	a	c	---	0.9765	0.1837	0.0240	1.99E-14	-0.0213	0.0254	4.01E-01
8	133951991	rs2739067	TG	a	g	---	0.598	-0.0415	0.0062	2.43E-11	0.0148	0.0067	2.61E-02
9	4290544	rs10814915	GLIS3	t	c	---	0.4438	0.0421	0.0061	5.06E-12	-0.0265	0.0066	5.62E-05
9	16214340	rs9298749	C9orf92	a	c	---	0.5878	-0.0393	0.0064	8.80E-10	0.0142	0.0069	3.88E-02
10	8682180	rs11255790	GATA3	t	c	---	0.302	-0.0410	0.0066	6.83E-10	0.0066	0.0071	3.58E-01
10	89849519	rs4933466	PTEN	a	g	---	0.6045	0.0395	0.0063	5.13E-10	-0.0149	0.0068	2.80E-02
11	45228686	rs12284404	PRDM11	a	g	---	0.2733	-0.0667	0.0069	2.48E-22	0.0060	0.0074	4.14E-01
11	115045237	rs4445669	CADM1	t	c	NBS: rs11215397, r2=0.89, D'=1.00 in CEU	0.4591	-0.0397	0.0061	5.76E-11	0.0092	0.0065	1.58E-01
13	24782080	rs7329958	SPATA13	t	c	---	0.3482	-0.0439	0.0065	1.13E-11	0.0105	0.0071	1.37E-01
14	36536181	rs398745	MBIP	a	c	---	0.5943	-0.0520	0.0062	3.97E-17	0.0248	0.0066	1.83E-04
14	36713154	rs2254613	MBIP	t	g	---	0.5505	-0.0346	0.0063	3.44E-08	0.0180	0.0068	7.99E-03
14	81490842	rs11159482	TSHR	t	c	---	0.0877	0.0846	0.0129	6.30E-11	-0.0176	0.0140	2.10E-01
14	81594143	rs59334515	TSHR	t	c	---	0.2236	-0.0539	0.0073	1.10E-13	-0.0136	0.0078	8.24E-02
14	81619945	rs12893151	TSHR	a	c	---	0.216	-0.0624	0.0078	1.02E-15	-0.0007	0.0084	9.31E-01
14	93585331	rs8015085	ITPK1	a	g	NBS: rs34162105, r2=0.94, D'=1.00 in CEU	0.2125	0.0671	0.0077	2.45E-18	-0.0348	0.0083	2.53E-05
15	49711185	rs17477923	FAM227B/FGF7	t	c	---	0.7355	0.0826	0.0069	2.57E-33	-0.0357	0.0073	1.15E-06
15	49749735	rs11639111	FAM227B/FGF7	t	c	---	0.4086	0.0450	0.0062	3.60E-13	-0.0132	0.0066	4.72E-02
15	89113877	rs13329353	DET1	t	c	---	0.6775	0.0614	0.0065	5.17E-21	-0.0324	0.0070	3.62E-06
16	4015313	rs1045476	ADCY9	a	g	---	0.1771	0.0490	0.0082	2.36E-09	-0.0029	0.0090	7.45E-01
16	14405428	rs30227	MIR365A	t	c	---	0.6115	-0.0468	0.0063	7.59E-14	0.0114	0.0067	8.72E-02
16	79745487	rs17767491	MAF	a	g	---	0.6784	0.0883	0.0065	3.35E-42	-0.0134	0.0070	5.43E-02
17	44762589	rs77819282	NSF	a	g	RS: rs199437, r2=0.97, D'= 1.00 in CEU	0.2371	0.0452	0.0074	1.13E-09	-0.0029	0.0080	7.15E-01
17	59338574	rs1157994	BCAS3	a	g	---	0.0452	-0.0904	0.0155	5.28E-09	-0.0505	0.0168	2.64E-03
17	70121339	rs1042673	SOX9	a	g	---	0.5214	-0.0546	0.0061	3.57E-19	-0.0028	0.0065	6.72E-01
17	70369758	rs963384	SOX9	t	c	---	0.4644	0.0351	0.0063	2.77E-08	0.0031	0.0068	6.43E-01
19	7222655	rs4804413	INSR	t	c	---	0.4371	0.0532	0.0062	8.64E-18	-0.0145	0.0067	2.98E-02
20	22596879	rs1203944	FOXA2	t	c	---	0.2279	-0.0509	0.0073	2.42E-12	0.0114	0.0078	1.45E-01
23	3612081	rs12390237	PRKX	a	g	---	0.6177	-0.0458	0.0068	1.74E-11	0.0027	0.0072	7.05E-01

Abbreviations: Position, variant genomic position based on build 37 (GrCh37); SNP, single nucleotide polymorphism; Gene, annotated gene; RS, Rotterdam Study; NBS, Nijmegen Biomedical Study; HUNT, Trøndelag Health Study; EAF_GWAS_Teumer, effect allele frequency in the study by T

Stable 2. Linear regression analyses investigating the relationships between log-transformed TSH (TSH_log) and FT4 concentrations in individuals stratified by polygenic score (PGS) quartiles in the Rotterdam Study and the Nijmegen Biomedical Study cohorts.

		Rotterdam Study				Nijmegen Biomedical Study			
		Estimate	**Std. Error**	**t-value**	**P-value**	**Estimate**	**Std. Error**	**t-value**	**P-value**
Q1	Intercept	-0.3486	0.0209	-16.65	2.54E-57	-0.3455	0.0282	-12.27	5.09E-32
	FT4*	-0.1213	0.0239	-5.07	4.48E-07	-0.1139	0.0335	-3.40	6.99E-04
Q2	Intercept	-0.0729	0.0212	-3.44	6.02E-04	-0.0783	0.0285	-2.74	6.21E-03
	FT4*	-0.1623	0.0244	-6.65	4.01E-11	-0.1383	0.0346	-4.00	6.79E-05
Q3	Intercept	0.1036	0.0212	4.90	1.08E-06	0.1320	0.0283	4.66	3.69E-06
	FT4*	-0.1362	0.0248	-5.48	4.86E-08	-0.0585	0.0321	-1.82	6.84E-02
Q4	Intercept	0.3021	0.0207	14.58	3.02E-45	0.2884	0.0281	10.25	2.61E-23
	FT4*	-0.1317	0.0238	-5.53	3.79E-08	-0.1560	0.0314	-4.96	8.46E-07

*effect estimates per 1 standard deviation (SD) change

Supplementary Table 3. A comparison of the diagnosis reclassification after applying genetically-determined TSH reference ranges in males and females.

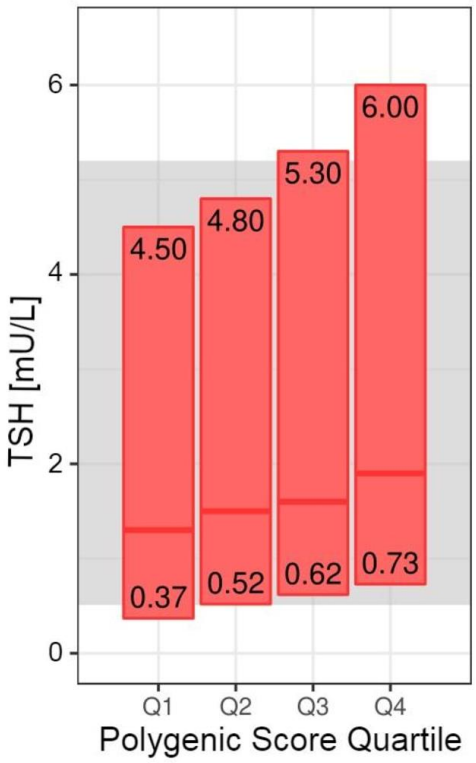
	Rotterdam Study				Nijmegen Biomedical Study			
	all	males	females	P-value	all	males	females	P-value
(Subclinical) hypothyroidism → Euthyroidism	42/170 (24.7%)	15/56 (26.8%)	27/114 (23.7%)	NS	23/94 (24.5%)	9/40 (22.5%)	14/54 (25.9%)	NS
Euthyroidism → (Subclinical) hypothyroidism	42/6501 (0.6%)	19/3183 (0.6%)	23/3318 (0.7%)	NS	25/3613 (0.7%)	7/1802 (0.4%)	18/1811 (1.0%)	0.03
(Subclinical) hyperthyroidism → Euthyroidism	49/163 (30.1%)	20/68 (29.4%)	29/95 (30.5%)	NS	28/93 (30.1%)	18/43 (41.8%)	10/50 (20.0%)	0.02
Euthyroidism → (Subclinical) hyperthyroidism	55/6501 (0.8%)	29/3183 (0.9%)	26/3318 (0.8%)	NS	26/3613 (0.7%)	12/1802 (0.7%)	14/1811 (0.8%)	NS

Supplementary Table 4. A comparison of TSH concentrations in individuals stratified by quartiles of PGS in the Trøndelag Health Study (HUNT) cohort.

	Trøndelag Health Study (HUNT)			
	Q1	**Q2**	**Q3**	**Q4**
Median TSH levels [mU/L]	1.30	1.50	1.60	1.90
Q1 vs Q2	$P = 1.7 \times 10^{-47}$			
Q2 vs Q3	$P = 1.1 \times 10^{-37}$			
Q3 vs Q4	$P = 6.8 \times 10^{-45}$			

Supplementary Table 5. Diagnosis reclassification after applying genetically determined TSH reference ranges in the Trøndelag Health Study (HUNT) cohort.

Individuals reclassified	Trøndelag Health Study (HUNT)
(Subclinical) hypothyroidism → Euthyroidism	83/644 (12.9%)
Euthyroidism → (Subclinical) hypothyroidism	81/25042 (0.3%)
(Subclinical) hyperthyroidism → Euthyroidism	171/635 (26.9%)
Euthyroidism → (Subclinical) hyperthyroidism	183/25042 (0.7%)



Supplementary Figure 1. Polygenic Score (PGS) quartile-specific TSH reference ranges in the Trøndelag Health Study (HUNT) cohort. PGS quartile-specific TSH reference ranges are shown in red, and population-based reference range is shown in grey (0.51-5.20 mU/L). Solid horizontal lines correspond to median TSH concentrations in each PGS quartile.