BMC Medicine

https://doi.org/10.1186/s12916-026-04942-5

Article in Press

Application of age- and sex-specific reference intervals for thyroid-stimulating hormone and free thyroxine in evaluating incidence and trends of thyroid dysfunctions: A population-based cohort study

Received: 7 January 2026

Accepted: 14 May 2026

Published online: 20 May 2026

Cite this article as: Lu J., Tang C., Ge G.M. *et al.* Application of age- and sex-specific reference intervals for thyroid-stimulating hormone and free thyroxine in evaluating incidence and trends of thyroid dysfunctions: A population-based cohort study. *BMC Med* (2026). https://doi.org/10.1186/s12916-026-04942-5

Jiawen Lu, Ching-Man Tang, Grace Meng-Qin Ge, Kenneth King-Yip Cheng, Keith Cheuk-Lun Lee, Kathryn Choon-Beng Tan, Stanley Kam-Ki Lam, Elaine Yun-Ning Cheung, Ching-Lung Cheung & Gloria Hoi-Yee Li

Application of age- and sex-specific reference intervals for thyroid-stimulating hormone and free thyroxine in evaluating incidence and trends of thyroid dysfunctions: A population-based cohort study

Running title: Age- and sex-specific reference intervals for TSH and FT4

Jiawen Lu[1], Ching-Man Tang[1], Grace Meng-Qin Ge[2], Kenneth King-Yip Cheng[1], Keith Cheuk-Lun Lee[1], Kathryn Choon-Beng Tan[3], Stanley Kam-Ki Lam[4,5], Elaine Yun-Ning Cheung[6], Ching-Lung Cheung[2,7], Gloria Hoi-Yee Li[1*].

[1] Department of Health Technology and Informatics, The Hong Kong Polytechnic University, Hung Hom, Hong Kong.

[2] Department of Pharmacology and Pharmacy, The University of Hong Kong, Pokfulam, Hong Kong.

[3] Department of Medicine, The University of Hong Kong, Pokfulam, Hong Kong.

[4] The Nethersole School of Nursing, The Chinese University of Hong Kong, Shatin, Hong Kong.

[5] School of Nursing and Health Sciences, Hong Kong Metropolitan University, Hong Kong.

[6] Hong Kong Institute of Diabetes and Obesity, The Chinese University of Hong Kong, Shatin, Hong Kong.

[7] Hinda and Arthur Marcus Institute for Aging Research, Hebrew SeniorLife, Boston, MA, United States.

*Corresponding author: Gloria Hoi-Yee Li

Email: gloria-hy.li@polyu.edu.hk

Abstract

Background

Diagnosis of thyroid dysfunctions heavily relies on the biochemical tests for thyroid-stimulating hormone (TSH) and free thyroxine (FT4). Conventional laboratory-specific reference intervals (RIs) for TSH and FT4 are uniform across adults and ignore age-, sex-, and time-related variations, which may lead to misdiagnosis of thyroid dysfunction. This study aimed to establish the age- and sex-specific RIs for TSH and FT4, and to evaluate their implications on the incidence and secular trends of thyroid dysfunctions.

Methods

Using a population-based electronic health record database in Hong Kong, we identified individuals without thyroid-related disorders who had valid TSH (N=2,111,661) and FT4 (N=825,522) measurements from 2006 to 2019. After harmonizing data across multiple institutions, we derived age- and sex-specific lower and upper reference limits (LRLs and URLs) for TSH and FT4 using the 2.5th and 97.5th percentiles of the biomarker distribution in seven age groups for both females and males. In comparison with the laboratory-specific RIs, we reclassified the thyroid status of the cohort individuals. We also compared the standardized incidence rate and secular trends in thyroid dysfunctions defined by the age- and sex-specific RIs and the conventional RIs.

Results

TSH and FT4 distributions varied by age, sex, and calendar year. Age- and sex-specific RIs for TSH and FT4 were significantly wider among elderly individuals (≥70 years) compared to the youngest adults (18-29 years). Among those with both TSH and FT4 measurements, 30.78% were classified as having thyroid dysfunctions using laboratory-specific RIs, whereas applying age- and sex-specific RIs reduced this proportion to 18.57%. The use of age- and sex-specific RIs was associated with lower observed incidence rates and revealed estimates of incidence and secular trends that are likely closer to the underlying disease burden.

Conclusions

Age- and sex-specific RIs could mitigate misdiagnosis of thyroid dysfunctions, thereby improving diagnostic accuracy and enabling more precise monitoring of secular trends. The secular trends and joinpoints identified for the age- and sex-specific URLs and LRLs of TSH and FT4 from 2006 to 2019 indicate that regular RI review is needed to ensure their continued clinical relevance. Future validation in geographically and ethnically diverse populations is warranted.

Keywords: Electronic health records; age- and sex-specific reference intervals; thyroid-stimulating hormone; free thyroxine.

Background

Thyroid hormones play a fundamental role in development and metabolic homeostasis throughout the lifespan. Yet, most thyroid dysfunctions manifest

through non-specific symptoms, such as fatigue and unintentional weight fluctuation [1, 2]. Its diagnosis thus relies heavily on biochemical markers, particularly thyroid-stimulating hormone (TSH) and free thyroxine (FT4), which vary across age [3-11], sex [3-7, 9], ethnicity [3, 6, 11], and geographical regions [4, 9]. Studies reported inconsistent findings: while some showed age-related increase in TSH [3-6, 10], others noted reduction in older populations [7, 8], with similarly conflicting results for FT4 [5, 9, 10]. Sex differences in TSH levels are also controversial [3-7, 12]. However, conventional laboratory-specific reference intervals (RIs) of TSH and FT4 are uniform for all and disregard these variations, possibly leading to misdiagnosis. Inappropriately narrow and broad RIs could result in overdiagnosis and underdiagnosis of true thyroid dysfunctions, respectively. Both scenarios give rise to suboptimal treatment. Establishing age- and sex-specific RIs for TSH and FT4 is therefore essential for enhancing the management of thyroid dysfunctions in various populations.

In establishing RIs for biochemical markers, direct methods involve sampling of preselected populations, but such approach is often limited by small sample size and selection bias [13]. With electronic health record (EHR) databases, routine laboratory test records are readily available for analysis, facilitating the use of indirect approach to derive RIs [13]. This is particularly relevant to thyroid function tests, as RIs for TSH established indirectly using such databases did not differ significantly from those derived from rigorously screened individuals without pre-existing thyroid diseases [14].

Despite international efforts to establish age- and sex-specific RIs for TSH and FT4 [5, 11, 15, 16], previous studies have not systematically explored the secular trends in these RIs, nor evaluated their impact on the incidence of thyroid dysfunctions. To address these research gaps, we utilized harmonized data from a representative EHR database in Hong Kong to establish age- and sex-specific RIs for TSH and FT4 from 2006 to 2019. We also evaluated the clinical implications of the age- and sex-specific RIs on reclassification and secular trends in incidence of thyroid dysfunctions (Workflow in ***Fig. 1***).

Methods

Data source

Anonymized medical records were retrieved from the Clinical Data Analysis and Reporting System (CDARS) [17], a population-based EHR database established since 1993 by the Hong Kong Hospital Authority (HA). The HA is the sole public healthcare service provider in Hong Kong through 43 public hospitals and institutions, and 123 outpatient clinics. Thus, the demographics, laboratory tests, diagnoses, procedures, and drug prescription records of all public healthcare service users were captured by CDARS.

Cohort identification for establishment of age- and sex-specific RIs

The cohort identification process is illustrated in ***Fig. 2***. From 1 January 2006 to 31 December 2019, a total of 2,466,036 and 1,162,569 public healthcare service users had TSH and FT4 laboratory test measurements, respectively, at 42 HA-managed hospitals or clinics with on-site thyroid function testing facilities (number of thyroid function tests performed in each institution listed

in ***Additional file 1: Table S1***). The cohort for establishing age- and sex-specific RIs was derived by excluding (i) laboratory test records measured during pregnancy or measured in individuals aged <18; (ii) individuals with pre-existing thyroid-related disorders, thyroid cancer, and/or disorders of the pituitary-hypothalamic axis (International Classification of Diseases, 9th Revision, Clinical Modification diagnosis codes in ***Additional file 1: Table S2***); (iii) individuals who had been prescribed medications for disorders of the pituitary, hypothalamus, and/or thyroid gland (British National Formulary codes in ***Additional file 1: Table S2***); and (iv) individuals who underwent thyroid-related procedures (including radioactive iodine therapy and thyroidectomy).

Harmonization of laboratory test records

The annual distribution of laboratory-specific lower and upper reference limits applied across institutions managed by HA for TSH and FT4 are presented in ***Additional file 2: Fig. S1*** and ***Fig. S2***, respectively. Given that interlaboratory variability among assays of thyroid function tests exists [18], we harmonized the TSH and FT4 laboratory test records using the methodology by Chuang-Stein (***Additional file 2: Supplementary Methods***) [19, 20] to enable combination of laboratory data from different institutions under HA with different laboratory-specific RIs. After harmonization, the most frequently adopted RIs in CDARS (0.27-4.2 mIU/L for TSH and 12-22 pmol/L for FT4) were applied uniformly to all laboratory tests measured in different institutions. These standardized RIs were known as the harmonized

RIs in this study. Thyroid status defined using harmonized test records and harmonized RIs was equivalent to those determined by raw values and laboratory-specific RIs. Relationship of age with harmonized TSH and FT4 was analyzed via sex-stratified generalized additive models.

Establishment of age- and sex-specific RIs

Due to the right-skewed distribution, harmonized TSH levels were log-transformed to approximate normality prior to establishment of RIs [21]. As the harmonization process may drive the TSH values towards 0 or negative numbers, addition of a small constant offset was required before log-transformation. In each calendar year, all TSH and FT4 measurements were stratified by age (18-29, 30-39, 40-49, 50-59, 60-69, 70-84, and ≥85 years) and sex. For individuals with multiple measurements in the same calendar year, only the first measurement was considered. The 2.5^{th} and 97.5^{th} percentiles of harmonized biomarker distributions were defined as the lower reference limits (LRLs) and upper reference limits (URLs), respectively [22]. The log-transformed LRLs and URLs of TSH were eventually back-transformed by exponentiation and subtraction of the offset to their original units (mIU/L) for clinical interpretability. For each age group, comparisons between sex were conducted using Wilcoxon rank-sum test for the skewed TSH values and Student's t-test for normally distributed FT4 levels. The trends in the medians, LRLs and URLs of TSH and FT4 were evaluated by joinpoint regression (Joinpoint Regression Program, version 5.4.0, Statistical Research and Applications Branch, National Cancer Institute), which

estimated the annual percent change (APC) and average annual percent change (AAPC) based on the best model selected by weighted Bayesian Information Criterion [23]. Compared to the youngest age group of 18-29, linear regression analysis was conducted to evaluate the associations of the remaining age groups with URLs, LRLs and widths of RIs, with adjustment for calendar year.

Thyroid status reclassification and incidence of thyroid dysfunctions

Individuals with TSH and FT4 levels measured within seven days were included to define their thyroid status using laboratory-specific, and age- and sex-specific RIs separately for comparison. If an individual had multiple measurements of TSH and FT4, only the first set of measurements was used to define the thyroid status. Euthyroidism was identified by both TSH and FT4 within RIs. Four thyroid dysfunctions included overt hypothyroidism (Ohypo) defined by TSH>URL and FT4<LRL; subclinical hypothyroidism (Schypo) as TSH>URL and FT4 within RI; overt hyperthyroidism (Ohyper) as TSH<LRL and FT4>URL, and subclinical hyperthyroidism (Schyper) as TSH<LRL and FT4 within RI. For each individual, thyroid status was considered reclassified if the status determined by age- and sex-specific RIs differed from that defined using the laboratory-specific RIs. The percentage of reclassified cases was calculated as the proportion in each age-sex sub-group. Annual incidence rates of four thyroid dysfunctions were calculated (detailed methodology described in ***Additional file 2: Supplementary Methods***) [24, 25] and their secular trends were analyzed during 2010-2019

using joinpoint regression [26]. Due to the intrinsic difference between the thyroid function testing platform by Abbott and other manufacturers, we additionally conducted a sensitivity analysis to establish the age- and sex-specific RIs for TSH and FT4 stratified by Abbott and non-Abbott platforms, examining their impact on thyroid status reclassification and incidence of thyroid dysfunctions (detailed methodology described in ***Additional file 2: Supplementary Methods***) [27-29].

Statistical analysis

All analyses were performed using R version 4.4.1. Continuous variables were summarized as mean ± standard deviation or median (interquartile range, IQR), and categorical variables as frequencies (percentages). Statistical significance was defined as a two-sided *P-value <0.05*.

Results

Cohort characteristics in establishment of age- and sex-specific RIs

Two cohorts without prior thyroid-related disorders were derived respectively for individuals with TSH (comprising 2,111,661 individuals with 3,978,396 valid measurements) and FT4 (825,522 individuals with 1,176,088 valid measurements) test records (***Fig. 2***). Individuals in the TSH cohort had a mean age of 57.40±18.78 years, consisting of mainly females (58.2%). In the FT4 cohort, the mean age was 60.01±19.22 years, and 60.0% were female. Age had significant non-linear relationships with TSH and FT4 levels. Particularly, both the median TSH and mean FT4 were significantly different between the two sexes in all age groups (comparison in ***Additional file 1:***

Table S3. Generalized additive model plots showed non-linear associations between age and TSH/FT4 levels, with distinct patterns observed (***Additional file 2: Fig. S3***).

Secular trends in median TSH and FT4 by age groups and sex are presented in ***Additional file 1: Table S4 and Additional file 2: Fig. S4***. In each age group in both sexes, two joinpoints were identified for median TSH during 2006-2019. Whereas at least one joinpoint was identified in median FT4 during 2006-2019 for all age- and sex-stratified sub-groups. We thus established the age- and sex-specific RIs of both TSH and FT4 on annual basis.

Age- and sex-specific RIs of TSH and FT4

Fig. 3 and ***Additional file 1: Table S5*** present the age- and sex-specific RIs of TSH and FT4 by calendar year, established from subgroups with sufficient sample size (TSH: N=4,563-46,956; FT4: N=1,131-16,317). The age- and sex-specific URLs of TSH consistently exceeded the harmonized URL of 4.2 mIU/L in all age groups in females across 2006-2019, except for young males ≤39 in 2016 and/or 2017 (***Fig. 3A***). ***Additional file 1: Table S6*** presents the linear regression analysis of RI values by age group. Compared to the youngest age group of 18-29, the age group of ≥85 years had the highest increase in the TSH URLs in both females (Beta [95% CI]: 1.06 [0.98 to 1.14], P<0.001) and males (1.39 [1.28 to 1.51], P<0.001). TSH LRLs had minimal variation of <0.3 mIU/L across all age groups in both sexes (***Additional file 1: Tables S5***). When compared to the youngest age group of 18-29, significant increases in the RI widths were observed in all age groups in

females and males ≥50 years (P<0.05; ***Additional file 1: Table S6***).

For FT4, URLs in females demonstrated a J-shaped distribution (***Fig. 3B***), with minimum and peak attained at the age group of 40-49 (compared to the youngest age group: -1.62 [-1.82 to -1.41], P<0.001) and ≥85 years (1.91[1.7 to 2.11], P<0.001), respectively (***Additional file 1: Table S6***). Male URLs showed a U-shaped pattern, and the lowest URL was seen in age group of 50-59 (compared to the youngest age group: -1.55[-1.82 to -1.28], P<0.001) (***Fig. 3B and Additional file 1: Table S6***). Inconsistent variations in FT4 LRLs were observed across age groups in females, but they contributed to a significant change in the RI width in age groups ≥40 when compared to the youngest age group (all P<0.001). In males, the LRLs in the age group of 18-29 were significantly higher than all other age groups, contributing to the significant increases in the RI widths among those aged ≥60 years (P<0.05) (***Additional file 1: Table S6***). From 2006 to 2019, secular trends and joinpoints were also identified for the URLs and LRLs of TSH and FT4 (detailed description of the results are provided in ***Additional file 2: Supplementary Results and Fig. S5,*** with annual trend statistics listed in ***Additional file 1: Tables S7-S8***).

Reclassification of thyroid status

A total of 932,621 individuals with TSH and FT4 measurements within seven days were eligible for thyroid status classification using the laboratory-specific and newly established age- and sex-specific RIs. Using laboratory-specific RIs, we classified the entire cohort as follows: euthyroidism (64.02%),

Ohypo (2.63%), Schypo (10.79%), Ohyper (6.52%) and Schyper (10.84%). Reclassification of thyroid status upon application of the newly established age- and sex-specific RIs were visualized by a chord diagram (***Fig. 4***), with substantial proportion of the entire cohort reclassified as euthyroid from both subclinical (Schypo: 4.81%; Schyper: 5.63%) and overt (Ohypo: 0.26%; Ohyper: 1.04%) thyroid dysfunction cases defined by laboratory-specific RIs. ***Additional file 1: Table S9*** presents the reclassification rates of thyroid function statuses when applying laboratory-specific versus age- and sex-specific RIs. Females (16.98%) were more likely reclassified as euthyroid than males (11.21%). More older adults (≥85 years: 23.08% for females, 15.03% for males) were reclassified as euthyroid than the younger adults (18-29 years: 19.98% for females, 9.42% for males). Proportion of individuals classified with each of the four thyroid dysfunctions decreased by up to 4.09% in the entire cohort.

Secular trends in the incidence of thyroid dysfunctions

The standardized incidence rates of the four thyroid dysfunctions defined by laboratory-specific, and age- and sex-specific RIs, together with their secular trends, are presented in ***Table 1*** and ***Fig. 5***. The standardized incidence rates of thyroid dysfunctions were substantially lower when they were defined by age- and sex-specific RIs compared to the conventional laboratory-specific RIs.

With age- and sex-specific RIs, incidence of Ohypo per 100,000 person-years rose from 14.55 in 2010 to 15.56 in 2019 (AAPC: 0.98% [95% CI: 0.2 to 1.76%],

P=0.014). In contrast, no overall secular trend was observed for Ohypo defined using laboratory-specific RIs during 2010-2019 although two joinpoints were identified. Incidence (per 100,000 person-years) of Schypo defined by age- and sex-specific RIs increased from 58.02 in 2010 to 69.45 in 2019 (AAPC: 2.1% [1.42 to 2.83%], *P*<0.001) without any joinpoints identified. Despite the similar increasing trend (AAPC: 2.35% [0.93 to 3.93%], *P*<0.001), Schypo defined by laboratory-specific RIs showed a higher incidence (82.74-122.16 per 100,000 person-years) and two joinpoints were detected.

Standardized incidence of Ohyper (per 100,000 person-years) defined by age- and sex-specific RIs demonstrated an overall increasing trend from 19.36 in 2010 to 33.53 in 2019 (AAPC: 6.59% [5.85 to 7.13%], *P*<0.001), whereas incidence of Ohyper determined by laboratory-specific RIs remained stable (AAPC: -0.07% [-0.94 to 0.8%], *P*=0.839). Although the incidence in Schyper defined by both the RIs demonstrated an increasing overall trend from 2010 to 2019, the magnitude of increase by age- and sex-specific RIs (AAPC: 1.28% [0.57 to 1.99%], *P*<0.001) was milder than that using laboratory-specific RIs (AAPC: 4.2% [2.71 to 5.75%], *P*<0.001). In sex-stratified analyses, age-standardized incidence rates of thyroid dysfunctions defined using age- and sex-specific RIs were consistently lower than those determined by laboratory-specific RIs, except for Schyper in males during 2010-2013 (comparison of secular trends in sex-stratified standardized incidence rates defined by laboratory-specific RI versus age- and sex-specific RI are illustrated in ***Additional file 2: Fig. S6***).

Sensitivity analysis for age- and sex-specific RIs stratified by platforms

Establishing and applying the age- and sex-specific RIs for Abbott and non-Abbott platforms (presented in ***Additional file 1: Tables S10-S11***) resulted in similar thyroid status reclassification patterns (chord diagram in ***Additional file 2: Fig. S7***) and longitudinal incidence trends for hyperthyroidism and hypothyroidism (comparison of secular trend in ***Additional file 2: Fig. S8*** and detailed description in ***Additional file 2: Supplementary Results***).

Discussion

This study established the population-based age- and sex-specific RIs for TSH and FT4 using an indirect method and one of the largest EHR datasets to date. The derived RIs were significantly wider in elderly individuals than in the youngest adults. With these age- and sex-specific RIs, the proportion of individuals classified with thyroid dysfunctions decreased from 30.78% to 18.57%, highlighting that conventional laboratory-specific RIs may overestimate the prevalence and incidence of thyroid dysfunctions, and lead to inaccurate assessment of the secular trends.

Our findings demonstrated that age- and sex-specific RIs for TSH and FT4 were significantly wider in older adults compared to younger individuals. For TSH, the widened RIs were primarily driven by a pronounced increase in URLs, consistent with known age-related changes in thyroid function [7, 15, 30, 31]. For FT4 in females aged ≥70, LRLs were not materially different from the youngest group, while URLs were highest in this age group following

the J-shaped pattern. In males aged ≥70, the LRLs were significantly lower than that of the youngest adults, whereas URLs exhibited a U-shaped pattern, both contributing to wider FT4 RIs in this age group. Yet, the J- or U-shaped patterns of FT4 URLs diverged from previous reports which indicated stable or slightly increased FT4 levels with advancing age [10, 12]. Such discrepancies may be attributed to the different sample sizes and sex distributions in elderly cohorts across studies [31, 32]. Proposed mechanisms explaining these age-related changes include a physiological recalibration of the hypothalamic-pituitary-thyroid axis set-point. Older individuals exhibit a reduced hypothalamic pituitary TSH response to thyroid hormone deficiency, leading to a smaller TSH increase for a comparable reduction in FT4 [33, 34]. It has also been proposed that aging may be associated with reduced TSH bioactivity or altered thyroidal responsiveness to TSH, although the underlying mechanisms remain incompletely understood [10]. In addition, age-related low-grade inflammation may influence peripheral thyroid hormone metabolism. Proinflammatory cytokine signaling has been shown to downregulate type 1 deiodinase, which may impair the conversion of thyroxine into active triiodothyronine [35], while inflammatory states may induce type 3 deiodinase activity, thereby promoting thyroid hormone inactivation [36]. Notably, the TSH LRLs established in our study were comparatively lower than studies conducted in France [15], Netherlands [16] and China [11], likely attributed to the different study designs and methodologies. First, some studies adopted strict exclusion cut-offs (e.g.,

excluding TSH values <0.10 and >10.0 mIU/L [15]) in establishing the RIs, while we prioritized the retention of all records to capture biochemical variation at population-level more comprehensively. Second, the Netherland [16] study established the RIs per manufacturer, implying that all laboratories had the same RIs provided by the manufacturer and no harmonization is required. Third, the Chinese study was indeed a multi-center study with thyroid function measured by assays of different manufacturers [11]. Yet, they directly pooled the data without harmonization, which requires cautious interpretation due to the existence of interlaboratory variation [18].

Despite recent efforts to establish age-specific [16] or age- and sex-specific RIs [5, 11, 15] for TSH and FT4 in various populations using indirect approach, our study further advanced the methodology. First, we harmonized TSH and FT4 data collected from multiple institutions, enabling broader inclusion criteria and minimizing imprecision arising from inter-laboratory variability and manufacturer differences [18]. Although these harmonized, age- and sex-specific reference limits are not directly comparable to raw measurements or international clinical thresholds, raw values can be readily converted to the harmonized scale using our transformation process to enable direct comparison. Second, unlike previous studies that pooled data across different calendar years to establish the new RIs [5, 11, 15], we stratified our cohort by the measurement year and revealed that medians, URLs and LRLs for TSH and FT4 varied from 2006 to 2019, with notable trends and joinpoints.

Attributable to the use of population-based EHR database, our annual cohorts remained sufficiently large (N≥4,563 for TSH and N≥1,131 for FT4) to reliably establish the age- and sex-specific RIs. The secular trends observed for TSH and FT4 demonstrate the temporal variation of thyroid function in the population, providing proof-of-concept evidence that age- and sex-specific RIs should be regularly reviewed. Compared with FT4, TSH is more sensitive to subtle changes in environment and physiology [38], leading to greater temporal variability. For practical implementation, we suggest reviewing age- and sex-specific RIs for FT4 concurrently with TSH at least every five years, or at a maximum interval of every ten years, to ensure ongoing diagnostic accuracy.

By implementing the age- and sex-specific RIs, an additional 14.76% of our cohort was reclassified as euthyroid, markedly reducing the number of thyroid dysfunction diagnoses, particularly among the elderly and females. This change mitigates the risk of overdiagnosis and unnecessary treatment, which aligns with studies conducted in France [15] and the Netherlands [16]. While evidence suggests that overtreating elderly patients with Schypo by levothyroxine may cause a higher risk of osteoporotic fractures [39], this highlights the clinical importance of age- and sex-specific RIs in preventing overtreatment.

Conversely, the upward shift of age- and sex-specific URLs and widening of RIs with age may potentially miss the true Ohypo and Schypo cases and incur risk of underdiagnosis in the elderly. Underdiagnosis and undertreatment of

true hypothyroidism are linked to elevated risk of adverse cardiovascular events [40]. To address this concern, we examined individuals reclassified as euthyroid using age- and sex-specific RIs. Most were originally defined as Schyper or Schypo by laboratory-specific RIs (chord diagram in ***Fig. 4***). Reclassification rate from Schypo to euthyroid increased with age, consistent with the well-documented age-related rightward shift of the TSH distribution [7, 30]. In contrast, reclassification rate from Schyper to euthyroid did not exhibit a clear link with age, aligning with the greater stability of the lower tail of the TSH distribution across age groups. This asymmetry supports the mechanistic interpretation that age- and sex-specific RIs primarily correct for age-related shift in the upper, but not the lower, tail of the TSH reference distribution. We further examined the raw TSH distribution among all 734,734 individuals classified as euthyroid using the age- and sex-specific RIs. Only 11,401 (1.55%) had a raw TSH value in the extreme range conventionally regarded as indicative of overt thyroid dysfunction in routine clinical practice (<0.1 or >10 mIU/L). Of these, 11,232 (98.5%) had originally been classified as either Ohyper (n=5,858) or Schyper (n=5,374), indicating that extreme raw values among the euthyroid individuals reclassified by age- and sex-specific RIs were concentrated at the suppressed rather than the elevated end of the TSH spectrum.

Since the clinical diagnosis of overt thyroid dysfunction requires confirmation by repeat testing and integration of clinical findings rather than reliance on a single biochemical snapshot [41, 42] , such cases would be expected to

undergo further evaluation. Evidence from the Thyroid Hormone Replacement for Untreated Old Adults with Subclinical Hypothyroidism Trial (TRUST) [43] and a pooled analysis including Institute for Evidence-Based Medicine in Old Age (IEMO) 80-plus thyroid trial [44] has further shown that withholding levothyroxine from older adults with Schypo was not associated with adverse cardiovascular or symptomatic outcomes, indicating that the underdiagnosis and undertreatment concern raised by age- and sex-specific RIs is unlikely to translate into clinical harm in the elderly population. Nonetheless, we acknowledge that a small subset of individuals classified as overt thyroid dysfunction under laboratory-specific RIs may fall within age- and sex-specific RIs. We therefore recommend that these RIs be interpreted as an adjunct to, rather than a replacement for, clinical assessment. In clinical implementation, in cases where thyroid function tests were abnormal under either laboratory-specific or age- and sex-specific RIs, prompt clinical review is required. In addition, raw TSH values <0.1 or >10 mIU/L should trigger clinical review irrespective of the thyroid status defined by the age- and sex-specific RIs. This would safeguard against underdiagnoses while allowing age- and sex-specific RIs to inform treatment decisions and reduce overtreatment of physiologically normal variation.

The adoption of age- and sex-specific RIs also resulted in lower standardized incidence rates for the four major thyroid dysfunctions when compared to the use of laboratory-specific RIs during 2010-2019, which may be explained by the consistently higher age- and sex-specific URLs for TSH compared to the

harmonized URL (***Fig. 3***). In particular, the incidence rate of Ohyper (19.36-33.53 per 100,000 person-years) defined by age- and sex-specific RIs in all calendar years were approximately half of that defined by laboratory-specific RIs (65.79-72.88 per 100,000 person-years; ***Table 1***). Application of the age- and sex-specific RIs of TSH and FT4 in clinical setting may therefore reduce over-diagnosis and over-treatment of overt thyroid dysfunctions, particularly Ohyper. Our secular trend analysis using age- and sex-specific RIs further revealed an increasing burden of all thyroid dysfunctions from 2010 to 2019, likely attributable to the rising number of thyroid function tests conducted in the public healthcare institutions during the study period, particularly in the older age groups. Such increase may be explained by the recommendation of thyroid function testing by international guidelines for high-risk groups such as patients with atrial fibrillation (AF) [45, 46] and obesity [47], as well as their elevating incidence over the years [48-50]. The incident Ohyper defined by age- and sex-specific RIs was also observed to increase from 2016 to 2019, which may again be explained by the increased volume–of thyroid function tests, that accompanied the sharp increase in the standardized incidence of AF from 2015 onwards [49]. Adoption of the platform-stratified age- and sex-specific RIs did not lead to materially different trends in thyroid dysfunctions (***Additional file 2: Supplementary Discussion***) [27]. Notably, the overall increasing trend of all thyroid dysfunctions during 2010-2019 and the sharp increase from 2016 to 2019 for the standardized incidence in Ohyper were not observed when laboratory-specific RIs were applied (***Table 1***), which

might delay the possible measures that could be implemented in early monitoring and prevention of Ohyper. On the other hand, the observed significant trends in thyroid dysfunctions classified by laboratory-specific RIs reflect change in clinical paradigm. Following the National Academy of Clinical Biochemistry guidelines issued in 2003 [51], there was a push to lower the TSH URL. However, in 2012, the American Thyroid Association and the American Association of Clinical Endocrinologists co-developed guidelines [52] that endorsed maintaining higher URLs of 4.12 mIU/L. Correspondingly, we revealed that the proportion of high TSH URLs ≥4.37 mIU/L increased from 28.2% in 2012 to 66.4% in 2016, followed by a decline to 44.6% in 2019 (annual distribution of laboratory-specific TSH reference limits shown in ***Additional file 2: Fig. S1***). This change in the proportion of high TSH URLs implied that less individuals were defined as Ohypo or Schypo from 2012 to 2016, while the number increased from 2016 to 2019. Moreover, the proportion of relatively low TSH LRLs ≤0.27 mIU/L was reduced from 71.8% in 2012 to 44.9% in 2015, followed by an increase to 58.5% in 2019. Such shift in low TSH LRLs suggested that individuals categorized as Schyper or Ohyper increased from 2012 to 2015 and subsequently declined from 2015 to 2019.

Despite the aforementioned strengths and clinical implications, this study also had limitations. First, in establishing the age- and sex-specific RIs, we did not exclude individuals with positive thyroid antibodies as recommended by the National Academy of Clinical Biochemistry [21]. Nevertheless, a study

among Hong Kong Chinese in 2011 found similar RI of TSH before and after excluding these individuals [53], supporting the reliability of our study design. Second, as the study was conducted in a population-based cohort in Hong Kong, the generalizability of our findings was limited. The age- and sex-specific RIs established in this study may not be directly applicable to populations in other geographical regions of different ethnicities and iodine intake. Third, although we applied stringent exclusion criteria for documented thyroid-related disease diagnoses and treatments, utilizing real-world data from EHR database inherently carries the risk of including individuals with unrecognized or unrecorded conditions. While pharmaceutical and surgical records effectively capture most overt cases, we cannot entirely rule out the presence of undiagnosed, asymptomatic individuals within our study cohort. Finally, as our standardized incidence rates were calculated using World Bank population data for individuals aged 20 years and above, our estimated incidence of thyroid dysfunctions may be comparatively lower than those reported in studies including all age groups.

Conclusions

This study established the first age- and sex-specific RIs for TSH and FT4 among adults in Hong Kong. Unlike conventional laboratory-specific RIs which apply a uniform standard to all individuals, age- and sex-specific RIs account for intrinsic physiological variation of TSH and FT4 across the lifespan and between sexes. Our findings demonstrate that the age- and sex-specific RIs can reduce misclassification of thyroid dysfunctions in different

age-sex subgroups, thereby improving diagnostic accuracy and enabling more precise monitoring of secular trends. These advancements may ultimately contribute to optimal management of thyroid dysfunctions and better patient outcomes. Finally, given TSH's higher sensitivity than FT4, we recommend review of the age- and sex-specific RIs of TSH at least every five years. To maintain clinical utility in response to observed secular trends, FT4 RIs should be reviewed concurrently, or at a maximum of every ten years. Further studies in geographically and ethnically diverse populations are warranted to validate the broader applicability and generalizability of these findings.

List of abbreviations

AAPC: average annual percent change

AF: atrial fibrillation

APC: annual percent change

CDARS: Clinical Data Analysis and Reporting System

EHR: electronic health record

FT4: free thyroxine

HA: Hospital Authority

IEMO: Institute for Evidence-Based Medicine in Old Age

IQR: interquartile range

LRLs: lower reference limits

Ohyper: overt hyperthyroidism

Ohypo: overt hypothyroidism

RIs: reference intervals

Schyper: subclinical hyperthyroidism

Schypo: subclinical hypothyroidism

TSH: thyroid-stimulating hormone

TRUST: Thyroid Hormone Replacement for Untreated Old Adults with Subclinical Hypothyroidism Trial

URLs: upper reference limits

Declarations

Ethics approval and consent to participate

Ethics approval was obtained from the institutional review board of the University of Hong Kong and the HA Hong Kong West Cluster (reference: UW22-502), and the Hong Kong Polytechnic University (reference: HSEARS20221113001). As electronic health records in CDARS are anonymized, individual consent was not required.

Consent for publication

Not applicable.

Data Availability

This study is conducted using the anonymized dataset from the CDARS. We are unable to share the data used in this study since the data custodian, the Hong Kong HA, has not provided us the permission. Nevertheless, CDARS data can be accessed upon approval of application via the HA Data Sharing Portal for research purposes (https://www3.ha.org.hk/data/Provision/Index/).

Competing Interests

The authors declare no competing interests.

Funding

This study is supported by the Health and Medical Research Fund, Health

Bureau, the Government of the Hong Kong Special Administrative Region (Reference: 20211121).

Authors' Contributions

GHYL contributed to funding acquisition. JWL, GHYL and CLC contributed to the conceptualization and study design. GHYL, GMQG, KCBT, and CLC contributed to data acquisition. JWL and GHYL contributed to methodology, data analyses, data validation, and wrote the original draft of the manuscript. JWL and GHYL have accessed and verified the underlying data. JWL, CMT, GMQG, KKYC, KCLL, KCBT, SKKL, EYNC, CLC, and GHYL had full access to the data in the study, contributed to the data interpretation, critically reviewed and revised the manuscript, approved the final manuscript and accepted the responsibility to submit it for publication. All authors read and approved the final manuscript.

Acknowledgements

None.

Authors' social media handles

X: @LuJiawen75839; @GloriaLi412951

Supplementary Information

Additional file 1: Supplementary Tables S1-S11. Table S1: Thyroid function test records by institution. Table S2: ICD-9-CM diagnosis and BNF codes used in cohort identification. Table S3: Harmonized TSH and FT4 levels by age and sex. Table S4: Secular trends in median values of TSH and FT4. Table S5: Age- and sex-specific TSH and FT4 RIs. Table S6: Linear regression analysis of TSH and FT4 RI values and widths. Table S7: Secular trends in age- and sex-specific reference limits of TSH. Table S8: Secular trends in age- and sex-specific reference limits of FT4. Table S9: Reclassification rate of thyroid function statuses using laboratory-specific versus age- and sex-specific RIs.

Table S10: Age- and sex-specific reference intervals for TSH and FT4 from Abbott platform. Table S11: Age- and sex-specific reference intervals for TSH and FT4 from non-Abbott platforms.

Additional file 2: Supplementary Methods, Supplementary Results, Supplementary Discussion, and Figures S1-S8. Fig. S1: Distribution of laboratory-specific TSH limits, 2006-2019. Fig. S2: Distribution of laboratory-specific FT4 limits, 2006-2019. Fig. S3: Non-linear associations of TSH and FT4 with age by sex. Fig. S4: Secular trends in median TSH and FT4 by sex. Fig. S5: Secular trends in TSH and FT4 reference limits by sex. Fig. S6: Age-standardized incidence rates of thyroid dysfunction by sex. Fig. S7: Sensitivity analysis of thyroid status reclassification by platform. Fig. S8: Sensitivity analysis of incidence trends by platform.

References

1. Taylor PN, Medici MM, Hubalewska-Dydejczyk A, Boelaert K. Hypothyroidism. Lancet. 2024;404:1347–64.
2. Jonklaas J, Razvi S. Reference intervals in the diagnosis of thyroid dysfunction: treating patients not numbers. Lancet Diabetes Endocrinol. 2019;7:473–83.
3. Hollowell JG, Staehling NW, Flanders WD, Hannon WH, Gunter EW, Spencer CA, et al. Serum TSH, T(4), and thyroid antibodies in the United States population (1988 to 1994): National Health and Nutrition Examination Survey (NHANES III). J Clin Endocrinol Metab. 2002;87:489–99.
4. Zhao L, Teng D, Shi X, Li Y, Ba J, Chen B, et al. The Effect of Increased Iodine Intake on Serum Thyrotropin: A Cross-Sectional, Chinese Nationwide Study. Thyroid. 2020;30:1810–9.
5. Yamada S, Horiguchi K, Akuzawa M, Sakamaki K, Yamada E, Ozawa A, et al. The Impact of Age- and Sex-Specific Reference Ranges for Serum Thyrotropin and Free Thyroxine on the Diagnosis of Subclinical Thyroid Dysfunction: A Multicenter Study from Japan. Thyroid. 2023;33:428–39.
6. Boucai L, Hollowell JG, Surks MI. An approach for development of age-, gender-, and ethnicity-specific thyrotropin reference limits. Thyroid. 2011;21:5–11.
7. Vadiveloo T, Donnan PT, Murphy MJ, Leese GP. 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;98:1147–53.
8. Hoogendoorn EH, Hermus AR, de Vegt F, Ross HA, Verbeek AL, Kiemeney LA, et al. Thyroid function and prevalence of anti-thyroperoxidase antibodies in a population with borderline sufficient iodine intake: influences of age and sex. Clin Chem. 2006;52:104–11.
9. Park SY, Kim HI, Oh HK, Kim TH, Jang HW, Chung JH, et al. Age- and

gender-specific reference intervals of TSH and free T4 in an iodine-replete area: Data from Korean National Health and Nutrition Examination Survey IV (2013-2015). PLoS One. 2018;13:e0190738.
10. Bremner AP, Feddema P, Leedman PJ, Brown SJ, Beilby JP, Lim EM, et al. Age-related changes in thyroid function: a longitudinal study of a community-based cohort. J Clin Endocrinol Metab. 2012;97:1554–62.
11. Li Q, Tang Y, Yu X, Qin G, Tian L, Cheng L, et al. Thyroid Function Reference Intervals by Age, Sex, and Race: A Cross-Sectional Study. Ann Intern Med. 2025;178:921–9.
12. Waring AC, Arnold AM, Newman AB, Bùžková P, Hirsch C, Cappola AR. Longitudinal Changes in Thyroid Function in the Oldest Old and Survival: The Cardiovascular Health Study All-Stars Study. J Clin Endocrinol Metab. 2012;97:3944–50.
13. Jones GRD, Haeckel R, Loh TP, Sikaris K, Streichert T, Katayev A, et al. Indirect methods for reference interval determination - review and recommendations. Clin Chem Lab Med. 2018;57:20–9.
14. Katayev A, Balciza C, Seccombe DW. Establishing reference intervals for clinical laboratory test results: is there a better way? Am J Clin Pathol. 2010;133:180–6.
15. Raverot V, Bonjour M, Abeillon du Payrat J, Perrin P, Roucher-Boulez F, Lasolle H, et al. Age- and Sex-Specific TSH Upper-Limit Reference Intervals in the General French Population: There Is a Need to Adjust Our Actual Practices. J Clin Med. 2020;9:792.
16. Jansen HI, Dirks NF, Hillebrand JJ, Ten Boekel E, Brinkman JW, Buijs MM, et al. Age-Specific Reference Intervals for Thyroid-Stimulating Hormones and Free Thyroxine to Optimize Diagnosis of Thyroid Disease. Thyroid. 2024;34:1346–55.
17. Cheung NT, Fung V, Kong JH. The Hong Kong Hospital Authority's information architecture. Stud Health Technol Inform. 2004;107:1183–6.
18. Ribera A, Sugahara O, Buchannan T, Vazquez N, Lyle AN, Zhang L, et al. Evaluation of the Current State of Thyroid Hormone Testing in Human Serum-Results of the Free Thyroxine and Thyrotropin Interlaboratory Comparison Study. Thyroid. 2025;35:471–84.
19. Chuang-Stein C. Some Issues Concerning the Normalization of Laboratory Data Based on Reference Ranges. Drug Inf J. 2001;35:153–6.
20. Chuang-Stein C. Summarizing Laboratory Data with Different Reference Ranges in Multi-Center Clinical Trials. Drug Inf J. 1992;26:77–84.
21. Baloch Z, Carayon P, Conte-Devolx B, Demers LM, Feldt-Rasmussen U, Henry JF, et al. Laboratory medicine practice guidelines. Laboratory support for the diagnosis and monitoring of thyroid disease. Thyroid. 2003;13:3–126.
22. National Committee for Clinical Laboratory Standards. How to define and determine reference intervals in the clinical laboratory; approved guideline. 2nd ed. NCCLS document C28-A2. Wayne, PA: National Committee for Clinical Laboratory Standards; 2000.
23. Kim HJ, Fay MP, Feuer EJ, Midthune DN. Permutation tests for joinpoint regression with applications to cancer rates. Stat Med. 2000;19:335–51.

24. Census and Statistics Department, Government of the Hong Kong Special Administrative Region. Population estimates. Census and Statistics Department. 2026. Available from: https://www.censtatd.gov.hk/en/page_1273.html. Accessed 5 May 2026.
25. World Bank Group. Global population (number). Gender Data Portal. 2026. Available from: https://genderdata.worldbank.org/en/indicator/sp-pop. Accessed 5 May 2026.
26. Ye Y, Sing CW, Hubbard R, Lam DCL, Li HL, Li GH, et al. Prevalence, incidence, and survival analysis of interstitial lung diseases in Hong Kong: a 16-year population-based cohort study. Lancet Reg Health West Pac. 2024;42:100871.
27. Dirks NF, den Elzen WPJ, Hillebrand JJ, Jansen HI, Boekel ET, Brinkman J, et al. Should we depend on reference intervals from manufacturer package inserts? Comparing TSH and FT4 reference intervals from four manufacturers with results from modern indirect methods and the direct method. Clin Chem Lab Med. 2024;62:1352–61.
28. Westbye AB, Aas FE, Dahl SR, Zykova SN, Kelp O, Dahll LK, et al. Large method differences for free thyroid hormone assays in the hyperthyroid range can affect assessment of hyperthyroid status: Comparison of Abbott Alinity to Roche Cobas, Siemens Centaur and equilibrium dialysis LC-MS/MS. Clin Biochem. 2023;121-2:110676.
29. Favresse J, Burlacu MC, Maiter D, Gruson D. Interferences With Thyroid Function Immunoassays: Clinical Implications and Detection Algorithm. Endocr Rev. 2018;39:830–50.
30. 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;92:4575–82.
31. Zhai X, Zhang L, Chen L, Lian X, Liu C, Shi B, et al. An Age-Specific Serum Thyrotropin Reference Range for the Diagnosis of Thyroid Diseases in Older Adults: A Cross-Sectional Survey in China. Thyroid. 2018;28:1571–9.
32. Fu J, Wang Y, Liu Y, Song Q, Cao J, Wang P. Reference intervals for thyroid hormones for the elderly population and their influence on the diagnosis of subclinical hypothyroidism. J Med Biochem. 2023;42:258–64.
33. Chaker L, Cappola AR, Mooijaart SP, Peeters RP. Clinical aspects of thyroid function during ageing. Lancet Diabetes Endocrinol. 2018;6:733–42.
34. Carlé A, Laurberg P, Pedersen IB, Perrild H, Ovesen L, Rasmussen LB, et al. Age modifies the pituitary TSH response to thyroid failure. Thyroid. 2007;17:139–44.
35. Kwakkel J, Wiersinga WM, Boelen A. Differential involvement of nuclear factor-kappaB and activator protein-1 pathways in the interleukin-1beta-mediated decrease of deiodinase type 1 and thyroid hormone receptor beta1 mRNA. J Endocrinol. 2006;189:37–44.
36. Boelen A, Kwakkel J, Alkemade A, Renckens R, Kaptein E, Kuiper G, et al. Induction of type 3 deiodinase activity in inflammatory cells of mice with chronic local inflammation. Endocrinology. 2005;146:5128–34.
37. Mammen JS. Interpreting Elevated TSH in Older Adults. Curr Opin

Endocr Metab Res. 2019;5:68–73.
38. Hadlow NC, Rothacker KM, Wardrop R, Brown SJ, Lim EM, Walsh JP. The relationship between TSH and free T4 in a large population is complex and nonlinear and differs by age and sex. J Clin Endocrinol Metab. 2013;98:2936–43.
39. Turner MR, Camacho X, Fischer HD, Austin PC, Anderson GM, Rochon PA, et al. Levothyroxine dose and risk of fractures in older adults: nested case-control study. BMJ. 2011;342:d2238.
40. Evron J, Moretti B, Evans R, Burns J, Hummel SL, Esfandiari NH, et al. Association between over- and under-replacement with thyroid hormone and incident heart failure. J Clin Endocrinol Metab. 2025 Dec 17:dgaf677.
41. Jonklaas J, Bianco AC, Bauer AJ, Burman KD, Cappola AR, Celi FS, et al. Guidelines for the treatment of hypothyroidism: prepared by the american thyroid association task force on thyroid hormone replacement. Thyroid. 2014;24:1670–751.
42. Ross DS, Burch HB, Cooper DS, Greenlee MC, Laurberg P, Maia AL, et al. 2016 American Thyroid Association Guidelines for Diagnosis and Management of Hyperthyroidism and Other Causes of Thyrotoxicosis. Thyroid. 2016;26:1343–421.
43. Stott DJ, Rodondi N, Kearney PM, Ford I, Westendorp RGJ, Mooijaart SP, et al. Thyroid Hormone Therapy for Older Adults with Subclinical Hypothyroidism. N Engl J Med. 2017;376:2534–44.
44. Mooijaart SP, Du Puy RS, Stott DJ, Kearney PM, Rodondi N, Westendorp RGJ, et al. Association Between Levothyroxine Treatment and Thyroid-Related Symptoms Among Adults Aged 80 Years and Older With Subclinical Hypothyroidism. JAMA. 2019;322:1977–86.
45. January CT, Wann LS, Calkins H, Chen LY, Cigarroa JE, Cleveland JC Jr, et al. 2019 AHA/ACC/HRS Focused Update of the 2014 AHA/ACC/HRS Guideline for the Management of Patients With Atrial Fibrillation. Circulation. 2019;140:e125–e51.
46. Van Gelder IC, Rienstra M, Bunting KV, Casado-Arroyo R, Caso V, Crijns H, et al. 2024 ESC Guidelines for the management of atrial fibrillation developed in collaboration with the European Association for Cardio-Thoracic Surgery (EACTS). Eur Heart J. 2024;45:3314–414.
47. Pasquali R, Casanueva F, Haluzik M, van Hulsteijn L, Ledoux S, Monteiro MP, et al. European Society of Endocrinology Clinical Practice Guideline: Endocrine work-up in obesity. Eur J Endocrinol. 2020;182:G1–32.
48. Cheng S, He J, Han Y, Han S, Li P, Liao H, et al. Global burden of atrial fibrillation/atrial flutter and its attributable risk factors from 1990 to 2021. Europace. 2024;26:euae195.
49. Zhang C, Zhang X, Zhou R, Ji K, Ding X, Chu M, et al. Secular trend in disease burden of atrial fibrillation/flutter in China from 1992 to 2021 and its projection in 25 years. BMC Public Health. 2025;25:2064.
50. Hu J, Xie J, Liu B, Peng W, Liu B, Yan Y, et al. Major risk factors of obesity in China and recommendations for future prevention and control efforts: a systematic review and meta-analysis. Lancet Reg Health West Pac.

2026;68:101833.
51. Wartofsky L, Dickey RA. The evidence for a narrower thyrotropin reference range is compelling. J Clin Endocrinol Metab. 2005;90:5483–8.
52. Garber JR, Cobin RH, Gharib H, Hennessey JV, Klein I, Mechanick JI, et al. Clinical practice guidelines for hypothyroidism in adults: cosponsored by the American Association of Clinical Endocrinologists and the American Thyroid Association. Endocr Pract. 2012;18:988–1028.
53. Chan AO, Iu YP, Shek CC. The reference interval of thyroid-stimulating hormone in Hong Kong Chinese. J Clin Pathol. 2011;64:433–6.

Figure titles and legends

Fig. 1: Workflow of the study design

Fig. 2. Cohort derivation for establishment of age- and sex-specific TSH and FT4 reference intervals

Abbreviation: TSH, thyroid-stimulating hormone; FT4, free thyroxine.

Fig. 3: Age- and sex-specific reference intervals for TSH and FT4 by calendar year (2006-2019)

Abbreviations: FT4, free thyroxine; TSH, thyroid-stimulating hormone.
*Red circles denote female-specific values; blue circles denote male-specific values. A color gradient (light to dark) indicates temporal progression from 2006 to 2019. Dashed horizontal lines represent harmonized laboratory reference intervals for TSH and FT4.

Fig. 4: Thyroid status reclassification using age- and sex-specific RIs with reference to laboratory-specific RIs

The chord diagram illustrates the reclassification of thyroid statuses defined using the age- and sex-specific RIs compared with the conventional laboratory-specific RIs. The bottom arc represents the initial thyroid statuses defined by laboratory-specific RIs, while the top arc represents the reclassified thyroid statuses defined by age- and sex-specific RIs. The width of each connecting ribbon is proportional to the number of individuals following that specific reclassification path. Ribbons connecting different categories (e.g. from ‘Schyper’ on the bottom arc to ‘EU’ on the top arc) indicate individuals whose thyroid status was reclassified. Conversely, ribbons connecting segments of the same category (e.g. from ‘EU’ to ‘EU’) represent individuals whose thyroid status remained unchanged using both sets of RIs. The outer scale is graduated in units of 0.01, where each unit corresponds to 1% of the total cohort.
Abbreviations: EU, euthyroidism; Ohyper, overt hyperthyroidism; Ohypo, overt hypothyroidism; Schyper, subclinical hyperthyroidism; Schoypo, subclinical hypothyroidism; RI, reference intervals; Uncat. refers to individuals who could not be categorized into any of the above thyroid status.

Fig. 5: Secular trends in standardized incidence rate of thyroid dysfunctions from 2010 to 2019

Abbreviations: AAPC, average annual percent change; APC, annual percent change; RI, reference interval.

Table 1. Secular trends in standardized incidence rates of thyroid dysfunctions

	2010	2011	2012	2013	2014	2015	2016	2017	201
Laboratory-specific RIs defined groups									

Overt hypothyroidism	Incidence [d]	27.89	30.96	29.26	24.47	17.5	15.94	17.22	17.78	18.4
	APC (95% CI) [c]	2.07 (-11.16,18.02)		▲[b]	-20.41 (-26.2,-11.82)*[a]		▲[b]	10.4 (3.96,27.6		
Subclinical hypothyroidism	Incidence [d]	98.18	106.15	114.32	100.06	82.74	83.46	88.38	88.47	114.6
	APC (95% CI) [c]	6.9 (-0.54,14.89)		▲[b]	-11.49 (-15.11,-6.62)*[a]		▲[b]	11.68 (8.09,18.		
Overt hyperthyroidism	Incidence [d]	65.79	68.59	70.09	68.02	68.6	72.88	68.54	67.71	68.8
	APC (95% CI) [c]	1.18 (0.02,5.15)*[a]					▲[b]	-1.61 (-6.01,-0.0		
Subclinical hyperthyroidism	Incidence [d]	77.66	73.42	71.93	85.99	104.52	115.38	113.24	122.32	104.
	APC (95% CI) [c]	-2.87 (-9.02,5.16)		▲[b]	17.56 (11.34,22.68)*[a]		▲[b]	-1.41 (-6.57,1.		
Age- and sex-specific RIs defined groups										
Overt hypothyroidism	Incidence [d]	14.55	14.29	14.36	14.13	14.93	14.34	14.44	15.5	15.5
	APC (95% CI) [c]	0.98 (0.2,1.76)*[a]								
Subclinical hypothyroidism	Incidence [d]	58.02	59.08	60.45	60.15	58.18	63.23	63.81	67.57	68.9
	APC (95% CI) [c]	0.63 (-2.85,2.57)			▲[b]		3.29 (1.64,7.08)*[a]			
Overt hyperthyroidism	Incidence [d]	19.36	19.94	21.21	21.86	22.95	24.96	25.64	28.03	31.6
	APC (95% CI) [c]	5.05 (2.24,6.01)*[a]					▲[b]	9.72 (7.07,		
Subclinical hyperthyroidism	Incidence [d]	68.41	72.83	69.67	72.58	74.34	75.33	74.96	77.33	76.0
	APC (95% CI) [c]	1.28 (0.57,1.99)*[a]								

Abbreviation: AAPC, average annual percent change; APC, annual percent change; CI, confidence interval; RI, reference intervals.

[a]* Indicates a statistically significant APC during this period; [b]▲ marks a significant joinpoint identified in this year;

[c] % was the unit for APC and AAPC; [d] 100,000 was the unit for standardized incidence.