Supplementary Materials

Table of contents

Supplementary Text

Supplementary Methods

Harmonization of laboratory test records

To enable combination of laboratory data from different institutions under the Hong Kong Hospital Authority with different laboratory-specific reference intervals (RI), we harmonized the laboratory test records of thyroid-stimulating hormone (TSH) and free thyroxine (FT4) using the methodology recommended by Chuang-Stein, which was applicable to all laboratory tests without making assumptions on their distributions [19, 20]. Raw laboratory test results from different institutions were standardized to have a lower reference limit (LRL) and upper reference limit (URL) of 0 and 1, respectively, taking into consideration the width of the original laboratory-specific RIs. Laboratory records with raw values within the RIs were transformed to be unit-free values between 0 and 1, while those with raw values outside the RIs were transformed to be negative values or >1. Units of all the standardized test results were eventually restored by adapting the most frequently adopted RIs in Clinical Data Analysis and Reporting System: 0.27-4.2 mIU/L for TSH and 12-22 pmol/L for FT4, which were known as the harmonized RIs in this study. At this point, the harmonized TSH levels can become negative or close to 0.

Incidence rate of thyroid dysfunctions

A new case of thyroid dysfunction was defined as an individual who was first classified as having a specific thyroid dysfunction during 2010-2019. For each of the four thyroid dysfunctions, crude incidence rates were calculated by dividing the number of new cases by the mid-year adult population, as reported by the Census and Statistics Department, the Government of the Hong Kong Special Administrative Region [24]. Age- and sex-standardized incidence rates for all individuals aged ≥20, as well as age-standardized rates for both sexes, were computed using the direct standardization method compared to the age and sex distribution from the World Bank Group [25]. All incidence rates were expressed per 100,000 person-years.

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

The RIs for both the TSH and FT4 function tests provided by Abbott were reported to be inaccurate and differ significantly from the indirect RIs [28]. Unlike other manufacturers that use one-step assays and biotin-streptavidin capture systems, Abbott's thyroid function testing platform utilizes a two-step chemiluminescent microparticle immunoassay without biotin [29]. In addition, Abbott provided RIs based on a central 99% interval, in contrast to the 95% interval typically employed by other manufacturers [27, 28]. Due to the intrinsic difference between Abbott and other thyroid function testing platforms, we performed a sensitivity analysis by stratifying the dataset into platform-specific sub-cohorts. The first sub-cohort comprised thyroid function tests potentially measured by Abbott, identified by their characteristic laboratory-specific RI (TSH: 0.35-4.94 mIU/L; FT4: 9.0-19/19.1 pmol/L) [27, 28], while the second sub-cohort included all the remaining thyroid function tests measured by other platforms. For each sub-cohort, we recalculated the RIs based on the 2.5th and 97.5th percentiles for every age- and sex-stratified subgroup in each calendar year. Using these two sets of platform-stratified age- and sex-specific RIs, we repeated our primary analysis to determine the influence of these platform-stratified RIs on thyroid status reclassification and longitudinal incidence trends of thyroid dysfunctions.

Supplementary Results

Secular trends in the URLs and LRLs of TSH and FT4

Secular trends in URLs and LRLs of TSH and FT4 are shown in ***Additional file 2: Fig. S5***. Age- and sex-specific URLs of TSH revealed at least two joinpoints from 2006 to 2019 in all age groups in both sexes. At least one joinpoint was identified for age- and sex-specific LRLs of TSH, except for the age groups of 30-39 in females and 18-29 in males (***Additional file 1: Table S7***). For FT4, at least one joinpoint was detected for the age- and sex-specific URLs during 2006-2019 in all age groups in females. In males, no joinpoints could be identified for age- and sex-specific URLs in the age groups ≤39 and 50-59. Age- and sex-specific LRLs of FT4 in all age groups in females showed significant overall increasing trends from 2006 to 2019 (Average annual percent change [AAPC], $P < 0.05$). Similar rising trends were observed in males (AAPC, $P < 0.05$), except for age-groups of ≤39 and ≥70 in males (***Additional file 1: Table S8***).

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

Approximately 16.5% (or 657,776) of 3,978,396 TSH and 20.6% (or 242,715) of 1,176,088 FT4 measurements were potentially originated from Abbott thyroid function testing platform based on the laboratory-specific RIs. Compared to non-Abbott platforms (TSH URL: 4.40 to 7.1 mIU/L; FT4 LRL: 9.08 to 12.82; FT4 URL: 20.8 to 26.38 pmol/L), age- and sex-specific RIs established for Abbott platform had lower TSH URLs (2.67 to 4.46 mIU/L) and FT4 URLs (18.90 to 24.25 pmol/L), but higher FT4 LRLs (11.66 to 13.41 pmol/L) (***Additional file 1: Tables S10 and S11***). TSH LRLs remained comparable between Abbott (-0.03 to 0.18 mIU/L) and non-Abbott platforms (-0.01 to 0.29 mIU/).

The application of these platform-stratified age- and sex-specific RIs resulted in a higher proportion of cases (2.35%) being categorized as overt hyperthyroidism but a similarly lower proportion (2.37%) as subclinical (***Additional file 2: Fig. S7***) compared to our original age- and sex-specific RIs (***Fig. 3***). While we observed an increased incidence of overt hyperthyroidism and a corresponding decreased incidence of subclinical hyperthyroidism, the significant rising trend in incidence of hyperthyroidism observed from 2010 to 2019 remained statistically robust (***Additional file 2: Fig. S8***). Whereas, the trend in hypothyroidism shifted slightly, with joinpoint observed in 2015 for overt hypothyroidism and in 2016 for subclinical hypothyroidism (***Additional file 2: Fig. S8***).

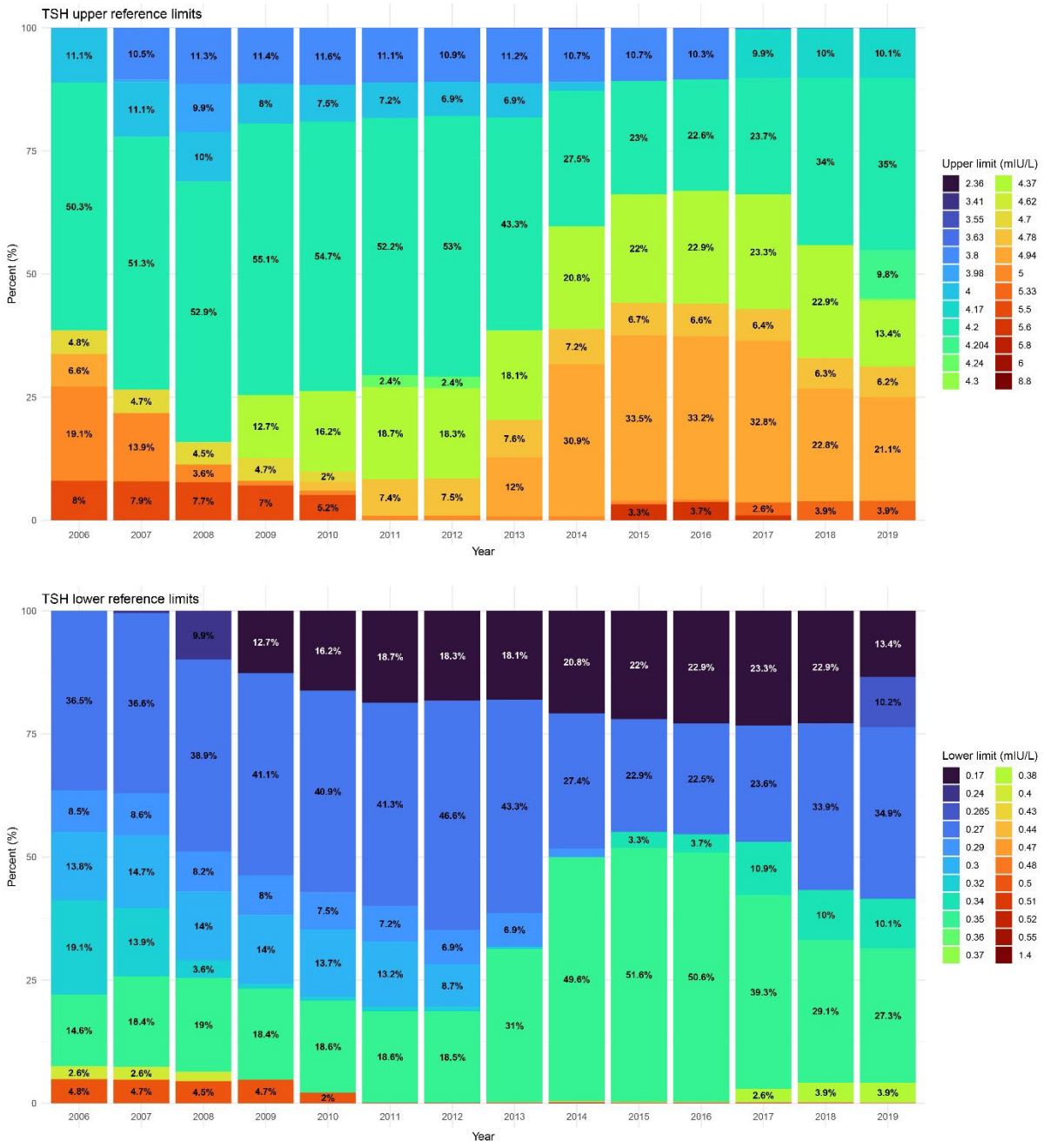
Supplementary Discussion

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

Sensitivity analyses revealed that the Abbott thyroid function testing platform yielded lower TSH and FT4 URLs compared to other platforms. Specifically, an approximately 2 pmol/L lower FT4 URL in Abbott platform shifted the classification of thyroid status between overt and subclinical hyperthyroidism. Yet, the overall incidence of hyperthyroidism remained unchanged. These results were highly consistent with a study conducted in the Netherlands [27], which demonstrated that indirect FT4 URLs derived from Abbott platforms were notably lower and exhibited a decrease of approximately 2 pmol/L over time compared to other manufacturers [27].

Supplementary Figures

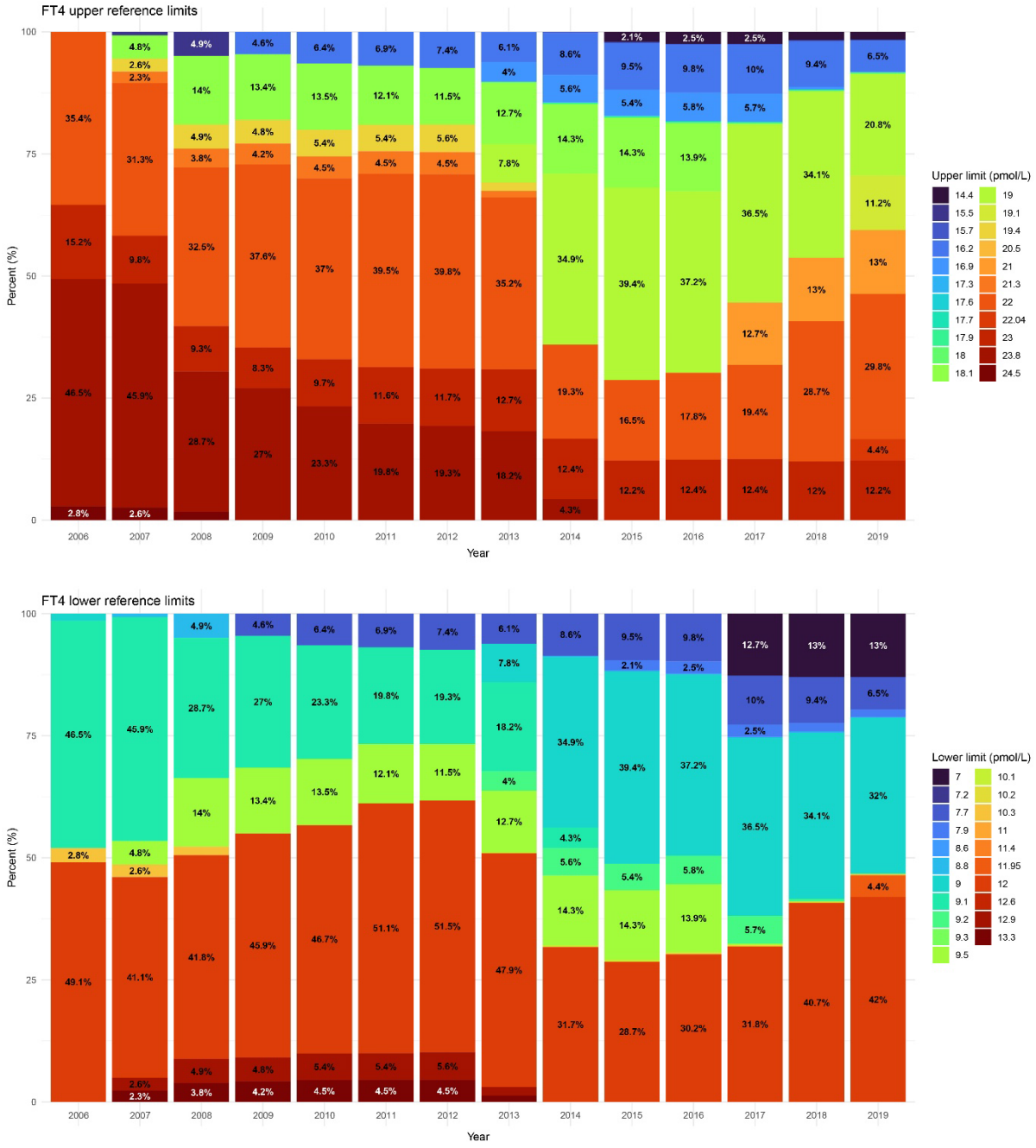
Fig. S1. Annual distribution of laboratory-specific lower and upper reference limits for TSH from 2006 to 2019



Abbreviation: LRL, lower reference limit; TSH, thyroid-stimulating hormone; URL, upper reference limit.

* The percentages indicate the proportion of TSH tests adopting the specific lower reference limit or upper reference limit among all tests conducted in Hospital Authority-managed institutions annually from 2006 to 2019.

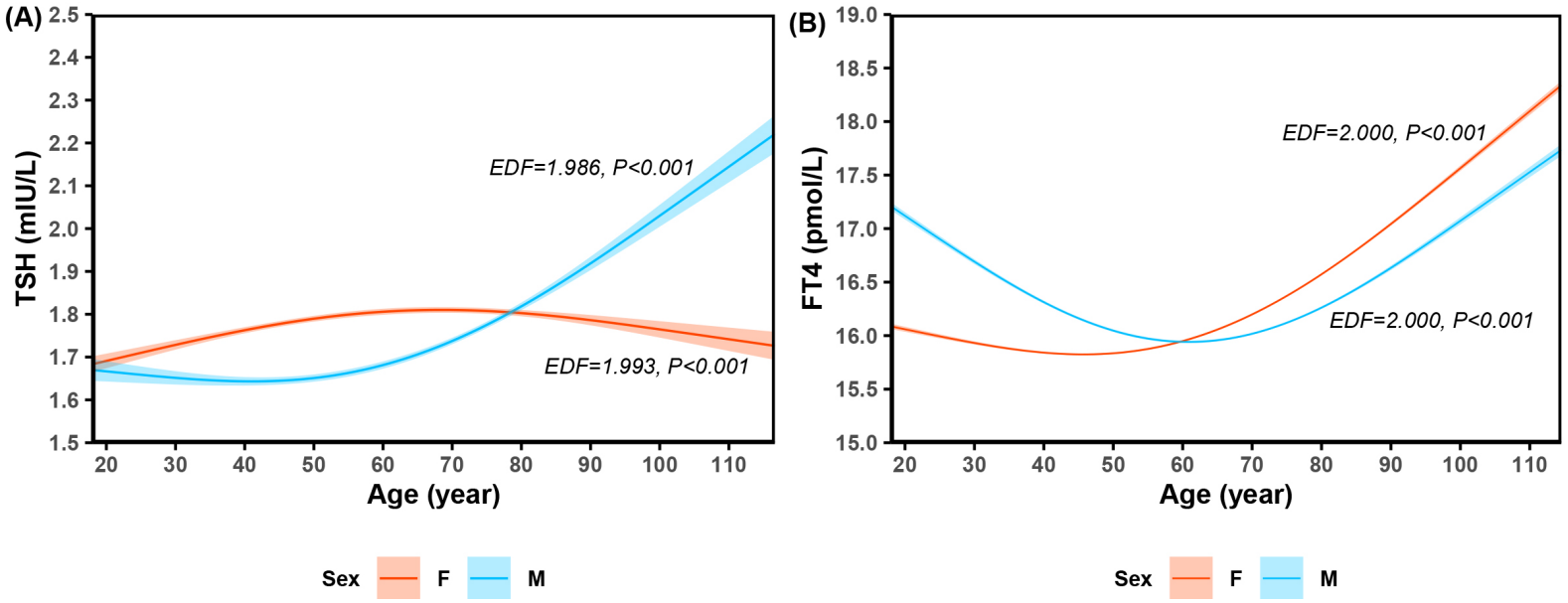
Fig. S2. Annual distribution of laboratory-specific lower and upper reference limits for FT4 from 2006 to 2019



Abbreviation: FT4, free thyroxine; LRL, lower reference limit; URL, upper reference limit.

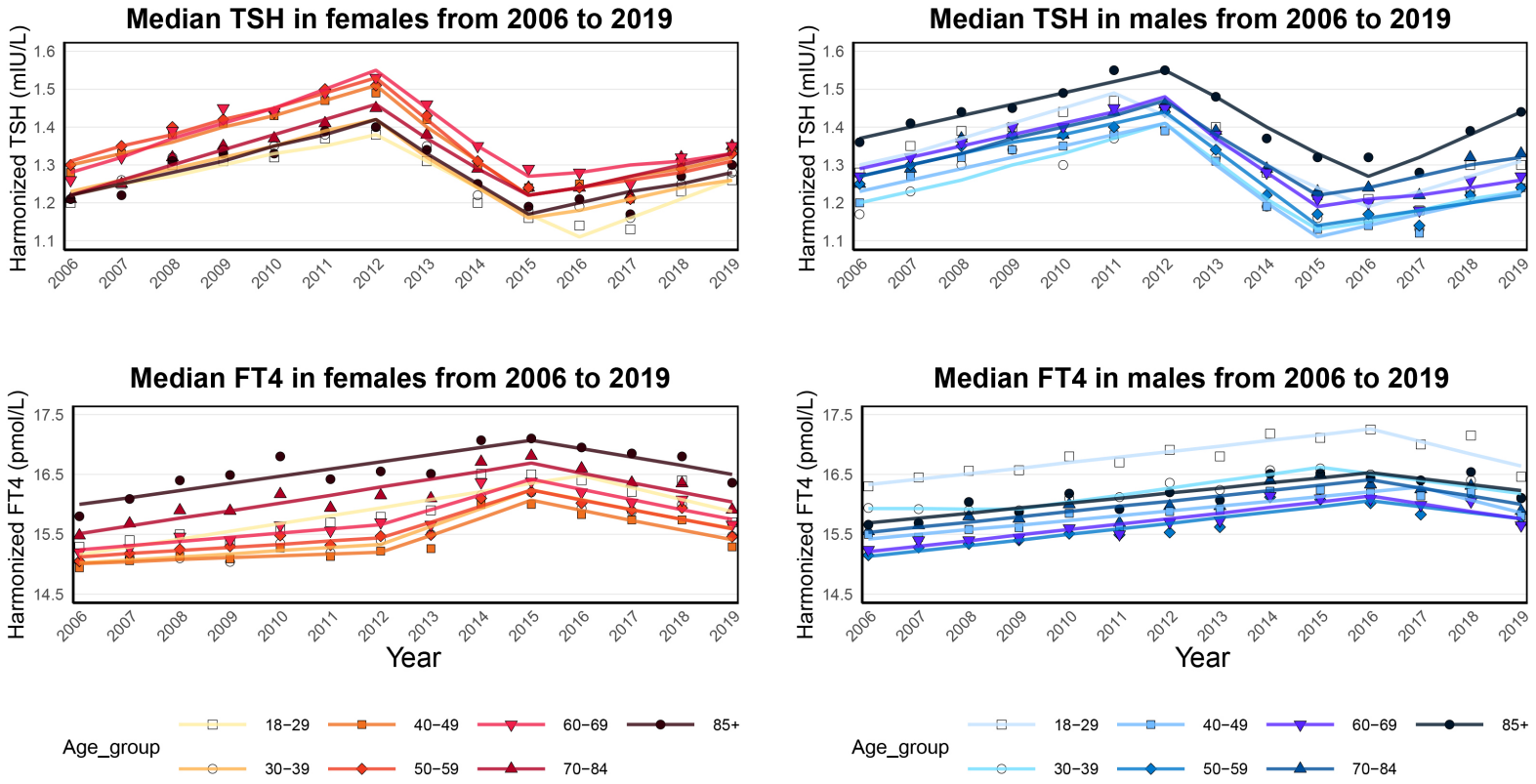
* The percentages indicate the proportion of FT4 tests adopting the specific lower reference limit or upper reference limit among all tests conducted in Hospital Authority-managed institutions annually from 2006 to 2019.

Fig. S3. Generalized additive model plots presenting the non-linear relationship of harmonized TSH and FT4 levels with age, stratified by sex



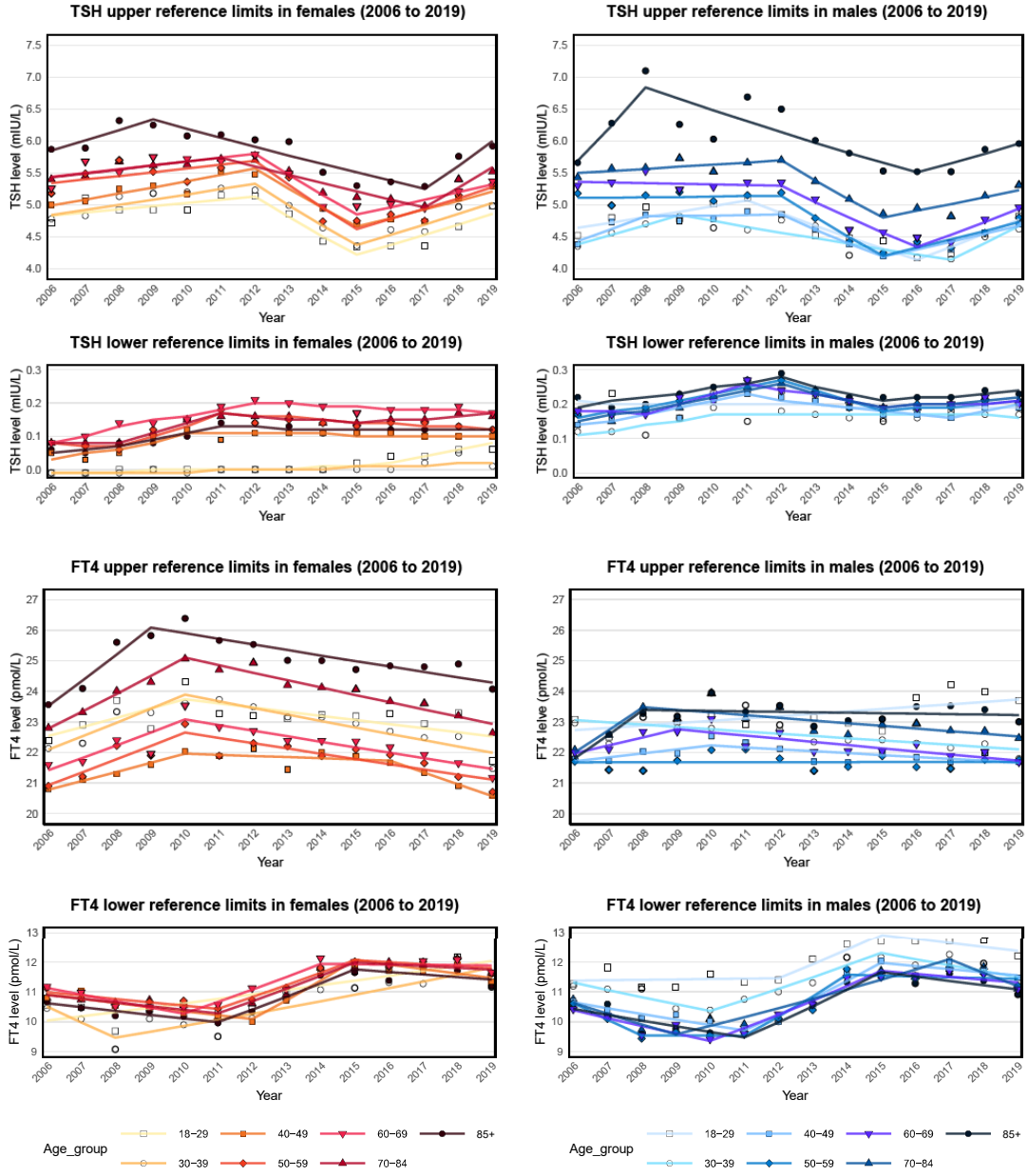
Abbreviation: EDF, effective degrees of freedom; TSH, thyroid-stimulating hormone; FT4, free thyroxine.
* EDF values exceeding 1 in the generalised additive models denote nonlinear associations in the fitted statistical relationships.

Fig. S4. Secular trends in median TSH and FT4 levels among females and males from 2006 to 2019



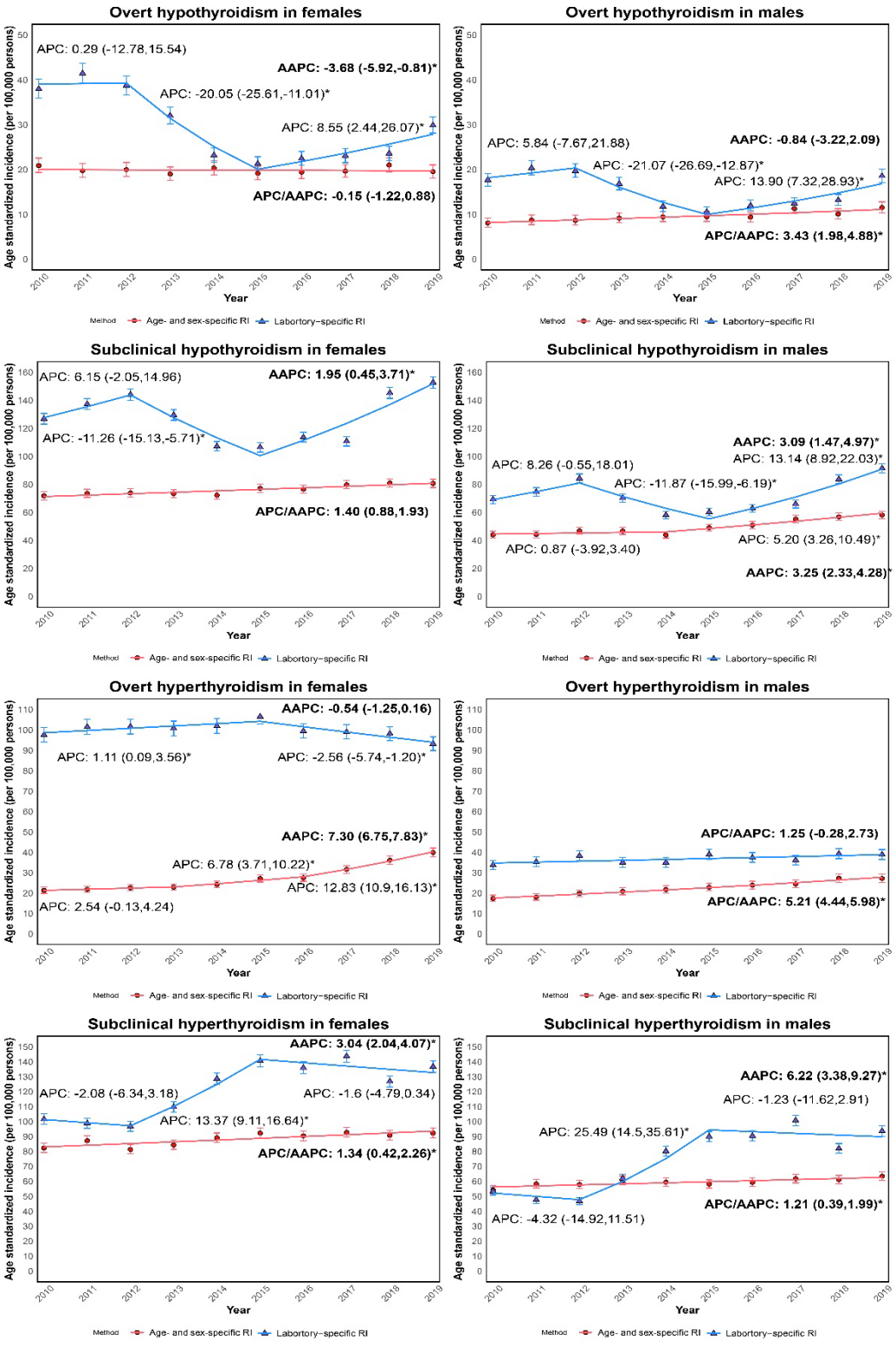
Abbreviation: TSH, thyroid-stimulating hormone; FT4, free thyroxine.

Fig. S5. Secular trends in upper and lower reference limits of TSH and FT4 among males and females from 2006 to 2019



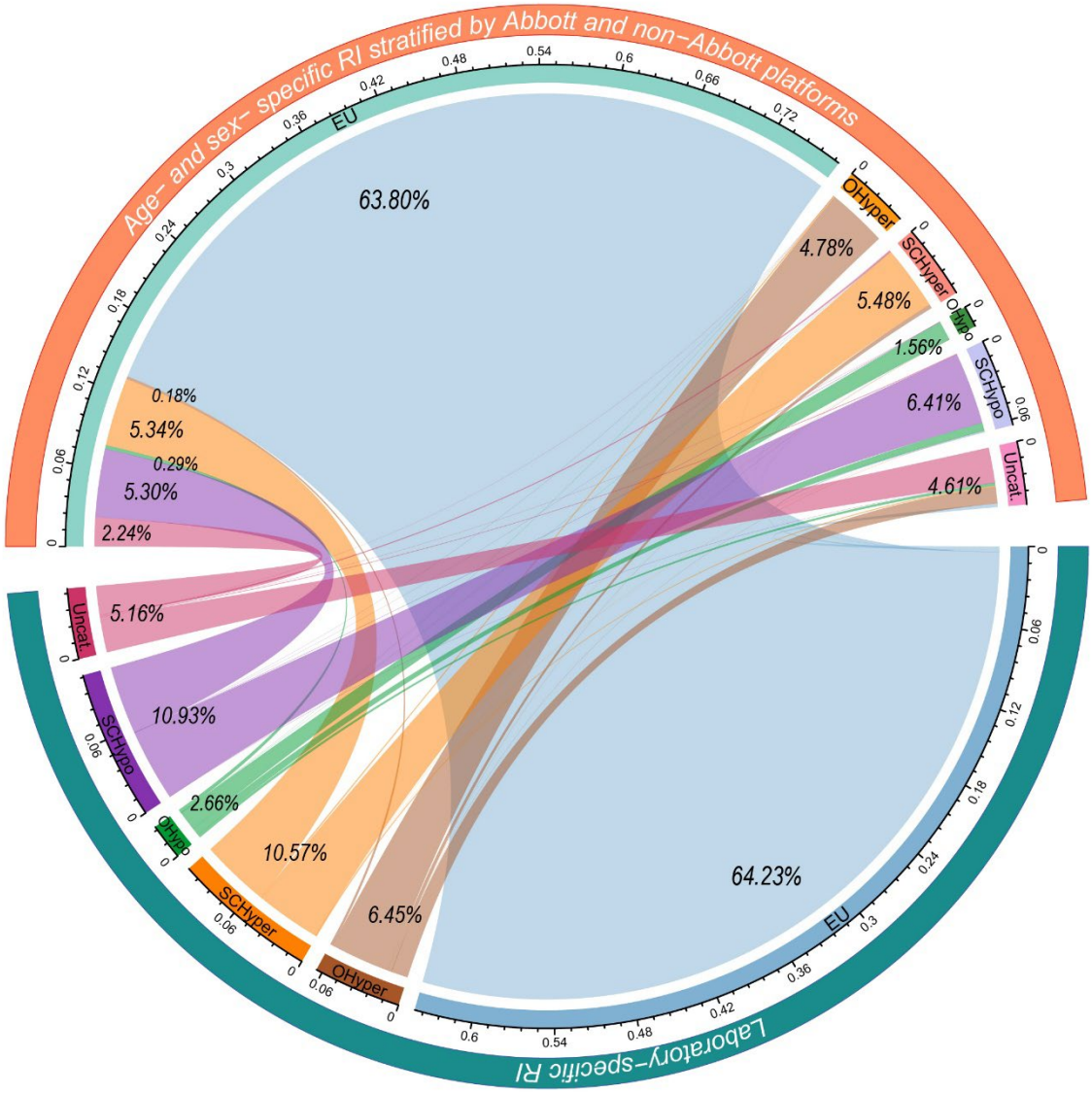
Abbreviation: TSH, thyroid-stimulating hormone; FT4, free thyroxine.

Fig. S6. Secular trends in age-standardized incidence rates of thyroid dysfunction by sex from 2010 to 2019



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

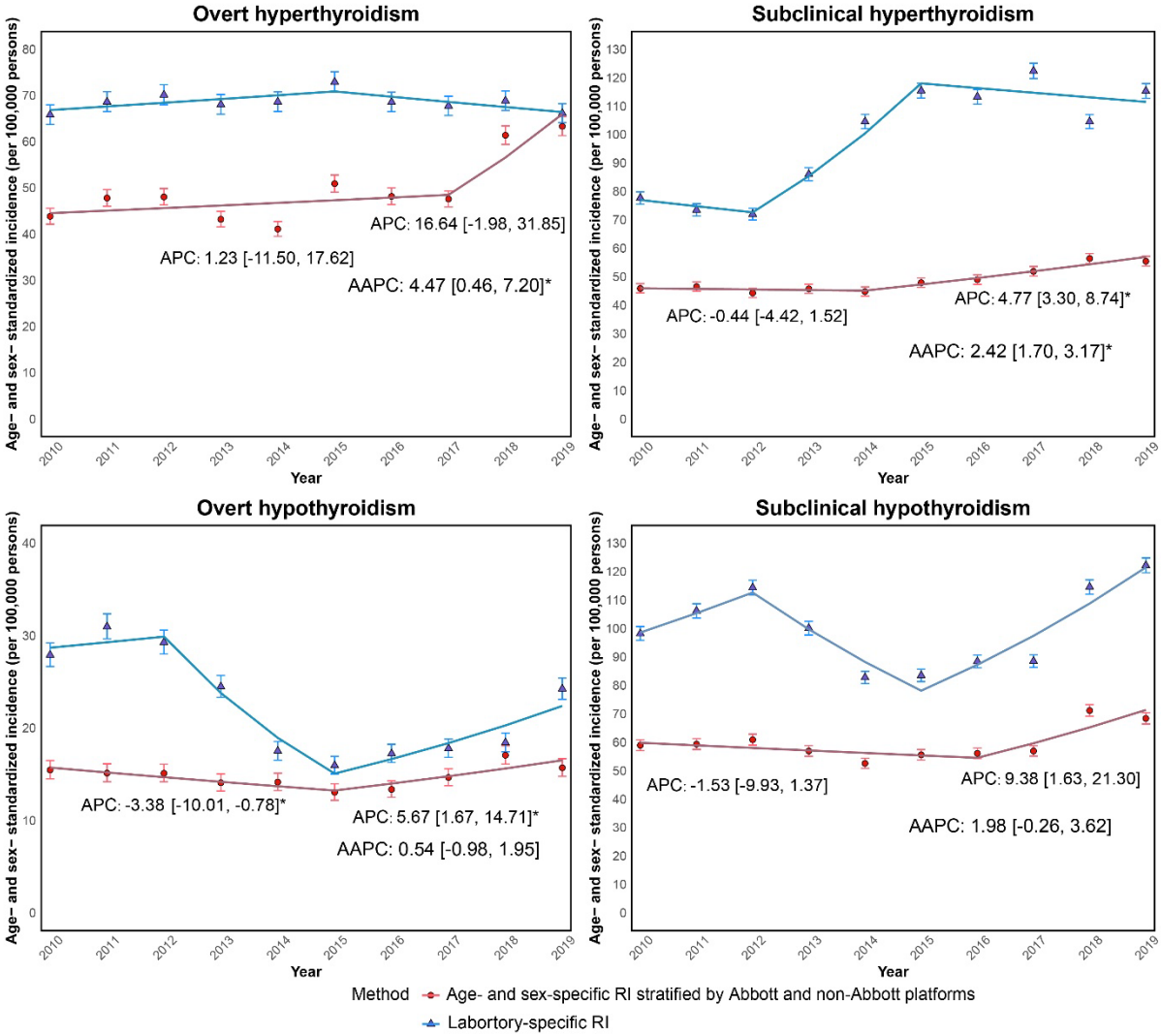
Fig. S7. Sensitivity analysis comparing thyroid status reclassification using laboratory-specific RIs versus age- and sex-specific RIs stratified by Abbott and non-Abbott thyroid function testing platforms



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.

The distribution of thyroid status defined by laboratory-specific RIs showed minimal changes after TSH/FT4 measurements were separately stratified into Abbott and non-Abbott cohorts. For instance, the proportion of euthyroid cases was 64.23% in the stratified dataset compared to 64.02% in the combined dataset.

Fig. S8. Sensitivity analysis of longitudinal trends in standardized incidence rates of thyroid dysfunction (2010-2019): comparison of laboratory-specific RIs versus age- and sex-specific RIs stratified by Abbott and non-Abbott platforms



* means significant APC or AAPC

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