



Research Article

Copyright © All rights are reserved by Ghassan Shannan

Establishing Population-Specific Reference Intervals for Free Thyroxine: A Syrian Study Demonstrating Critical Discrepancies with Manufacturer Ranges

Ghassan Shannan^{1*}, Adnan Alkhatib², Mahmoud Elias³, Zeina Malek⁴ and Nasser Thallaj⁵

¹Faculty of Pharmacy Dean, Arab International University, Damascus, Syria

²Consultant clinical pathology CEO. Alkhatib Medical Laboratory

³IT Services Department, Higher institute of Applied Sciences and Technology

⁴Faculty of Pharmacy, Arab International University, Damascus, Syria

⁵Faculty of Pharmacy, Arab International University, Damascus, Syria

*Corresponding author: Ghassan Shannan, Faculty of Pharmacy Dean, Arab International University, Damascus, Syria.

To Cite This article: Ghassan Shannan*, Adnan Alkhatib, Mahmoud Elias, Zeina Malek and Nasser Thallaj, Establishing Population-Specific Reference Intervals for Free Thyroxine: A Syrian Study Demonstrating Critical Discrepancies with Manufacturer Ranges. *Am J Biomed Sci & Res.* 2026 32(4) AJBSR.MS.ID.004178, DOI: [10.34297/AJBSR.2026.32.004178](https://doi.org/10.34297/AJBSR.2026.32.004178)

Received: 📅 July 15, 2026 **Published:** 📅 October 16, 2026

Abstract

Background: The diagnostic accuracy of thyroid function testing depends critically on the reference interval used to interpret results. Manufacturer-supplied intervals, derived predominantly from Western European and North American cohorts, may misrepresent thyroid hormone physiology in genetically and geographically distinct populations, creating a systematic risk of misclassifying euthyroid individuals as hypothyroid or hyperthyroid.

Objective: To establish the first population-specific, age- and sex-stratified free thyroxine (FT4) reference intervals for the Syrian population and to quantify their divergence from the manufacturer-supplied generic range for the Elecsys FT4 III assay.

Methods: Using the indirect method per CLSI EP28-A3c, 10,000 consecutive FT4 measurements from a high-volume accredited laboratory in Damascus (Jan.2020 to Dec.2024) were stratified into six demographic subgroups by biological sex and age category (children 2-16 yr, adolescents 16-18 yr, adults > 18 yr). Reference intervals were derived by iterative mean $\pm$ 2SD truncation. Inferential analyses comprised Kruskal-Wallis omnibus testing, Dunn post-hoc comparisons with Bonferroni correction, Spearman rank correlations stratified by sex and developmental window, and multiple linear regression with a sex-by-age interaction term.

Results: All six Syrian subgroups fell below the manufacturer lower threshold of 0.93 ng/dL, with adult deviations reaching -0.22 ng/dL, while upper limits exceeded the manufacturer ceiling of 1.70 ng/dL by 0.12-0.19 ng/dL across all strata. The pooled adult RI was 0.74-1.85 ng/dL; sex-specific adult intervals were 0.71-1.82 ng/dL in females and 0.72-1.89 ng/dL in males. Inter-stratum heterogeneity was confirmed (Kruskal-Wallis $H = 41.7$, $p < 0.001$). Age showed a significant but weak negative association with FT4 in adults ($\rho = -0.09$ to -0.14 ; $p < 0.001$) but was absent in the 2-16-year paediatric window ($\rho \approx -0.03$ - 0.04 ; $p > 0.20$), supporting a single sex-stratified pediatric interval. Sex independently predicted FT4 in adults ($\beta = +0.043$ ng/dL; $p < 0.001$) with no significant sex-by-age interaction ($p = 0.347$).

Conclusion: The manufacturer-supplied interval is structurally misaligned with the Syrian population at both boundaries, generating a predictable and dual risk of diagnostic misclassification. The locally derived, sex- and age-specific intervals established in this study should be adopted as operative standards in Syrian clinical laboratories, pending prospective multi-center validation.

Keywords: Free thyroxine (FT4), Reference intervals, Population-specific, Syrian population, Thyroid function tests, Immunoassay, Diagnostic accuracy, Laboratory medicine, Geographic variation.

Introduction

Measurement of serum Free Thyroxine (FT4) and Thyroid-Stimulating Hormone (TSH) is important for the diagnosis of thyroid abnormalities. Reference intervals for these biomarkers are very critical for their interpretation. The use of population specific reference intervals is gaining momentum for precise diagnosis. This manuscript focuses on the impact of using age, sex and geographic location specific reference intervals of FT4 for an accurate diagnosis of thyroid abnormalities. FT4 values vary in different periods of life: infancy, childhood and old age [1]. highlighted the necessity of providing pediatric reference intervals for thyroid hormones to support a more reliable diagnosis and management in children. According to [2], age- specific reference intervals for FT4 improve the diagnosis of thyroid dysfunction especially in the elderly. As thyroid function can decline with advancing age [3], proposed that more specific reference intervals should be available for the elderly population to avoid misdiagnosis. Sex influences thyroid hormone levels, which can affect thyroid function assessment. Generally, women are reported to have higher FT4 levels than men [4]. Based on the above, age- and sex-specific reference intervals (RIs) might help improve the clinical diagnosis and management of thyroid diseases [5]. The authors argued that taking into account the influence of age and sex can prevent overlooking thyroid disease, which may occur more frequently in the elderly population, especially in elderly women who may have distinct presentation of the disease and consequently can be more frequently misdiagnosed than men.

The establishment of universal reference intervals is made more complex by geography. Thyroid function is known to vary with geography, time, and genetic, dietary and environmental factors [6,7]. Studies have also shown that there can be considerable differences in the reference intervals for TSH and FT4 derived from studies that have used different assay methodologies [8]. The study by [9] on the thyroid function reference intervals for adults in Lebanon highlighted the importance of local dietary habits and environmental factors when deriving a more relevant and specific reference interval for the Lebanese population.

The clinical implications of the use of population specific reference ranges for the measurement of FT4 in the diagnosis of hypothyroidism and hyperthyroidism have been huge and have improved the detection of subclinical thyroid disease. Recently [10] have suggested that it may be unnecessary to diagnose overt hypothyroidism and hyperthyroidism if population specific reference ranges are not used. It therefore follows that the use of locally derived reference ranges for the measurement of Thyroid Function Tests (FTS) is essential in order that any legitimate changes in thyroid function due to age, sex and other factors in different populations are recognised. In addition, the societal value should not be ignored, because a lack of awareness about the importance of population specific reference intervals in primary care clinicians, can hinder the efforts to achieve the national health care objectives

such as early diagnosis and management of thyroid disease, which can influence the health of the society. [11] argued that serum FT4 levels affect the implementation of preventive measures in order to reduce the risk of metabolic syndrome in populations. To achieve the national health care objectives and in order to provide better health care services to the society, healthcare providers should be aware of the significance of population specific reference intervals. Aging, Sex and Geography (AGE) specific reference intervals are critical to modern endocrinology practice. Xie et al. proposed the notion that accurate and precise hormone testing would contribute to more accurate diagnosis and thus more appropriate and precise therapeutic approaches. Thus, precision medicine at the population level for different age, sex and geography would allow more personalized management of hormone abnormality based on exact hormone concentrations to prevent misdiagnosis.

The impact of population-specific reference intervals for FT4 in the diagnosis of thyroid diseases is of great significance. By considering factors such as age, sex and geographic location, the accuracy of thyroid disease diagnosis and treatment will be increased. With the ever-increasing number of research studies, it can be seen that a uniform reference range for FT4 is not suitable for thyroid function evaluation due to the diversity of thyroid function in different population. Considering the increasing number of studies focusing on population-specific reference intervals, it can be seen that adapting age, sex and geographic location to the clinical work will not only improve the accuracy of thyroid disease diagnosis and treatment, but also be in line with the development trend of personalized medicine. Therefore, the continuous research and education of population-specific reference intervals for thyroid disease diagnosis and treatment are still needed to improve the clinical efficacy and prognosis of patients with thyroid diseases [12,13].

Materials and Methods

Study Design and Ethical Framework

This investigation was conducted as a retrospective observational study employing the indirect method for Reference Interval (RI) derivation, in accordance with the framework established by the Clinical and Laboratory Standards Institute (CLSI) guideline EP28-A3c and the recommendations of the International Federation of Clinical Chemistry (IFCC) Working Group on Reference Intervals and Decision Limits. The indirect approach was selected because the source dataset comprised routine clinical specimens rather than a pre-screened population of confirmed healthy volunteers, rendering a direct (a priori selection) design impractical at this scale. All procedures were conducted in full compliance with the Declaration of Helsinki (2013 revision) and were approved by the Institutional Review Board of Alkhatib Medical Laboratory, Damascus, Syria (reference: AML-IRB-2024-07). Because the data were extracted retrospectively from anonymized laboratory records without any prospective patient contact or intervention,

formal written consent was waived by the approving body. Patient confidentiality was preserved throughout: records were de-identified prior to extraction, retaining only age, biological sex, and the FT4 measurement.

Data Source and Specimen Eligibility

FT4 results were extracted from the electronic laboratory information system (LIS) of Alkhatib Medical Laboratory, a high-throughput accredited clinical biochemistry facility in Damascus, Syria, serving a broad urban and peri-urban catchment area. The dataset encompassed 10,000 consecutive FT4 measurements generated between January 2020 and December 2024 (inclusive), representing the unselected output of routine thyroid function testing ordered by attending clinicians across general practice, endocrinology, and antenatal care settings. Records were eligible for inclusion if the requesting clinician's indication was documented as routine health screening, general wellness assessment, or antenatal thyroid surveillance-categories operationally interpreted as consistent with an apparently healthy or low-pretest-probability population, in keeping with the indirect method's implicit assumption of a predominantly euthyroid sample.

Records were excluded if any of the following flags were present in the LIS: a concurrent diagnosis of overt thyroid disease (hypothyroidism, hyperthyroidism, thyroiditis, or thyroid malignancy); a prescription for levothyroxine, antithyroid agents,

amiodarone, lithium, or other medications with established thyroid-interference potential; a documented acute non-thyroidal illness at the time of sampling; or a patient age below 2 years, given the distinct physiological norms of the neonatal and early infancy periods. Results flagged by the analyser for instrument or analytical error were likewise discarded before database lock. All specimens in the eligible pool were serum samples collected by standard venepuncture, processed within 2 hours of collection, centrifuged at $2,000 \times g$ for 10 minutes at 20°C, and stored at 2-8°C until analysis, which was completed within 8 hours in all cases. Pre-analytical variables including haemolysis index, icterus index, and lipaemia index were recorded automatically by the analyser; samples exceeding platform-specified interference thresholds ($H \geq 3+$, $I \geq 3+$, or $L \geq 3+$) were removed prior to analysis.

Analytical Platform and Quality Assurance

All FT4 measurements were performed on a single analyser using the Elecsys FT4 III assay [14], an automated electrochemiluminescence immunoassay (ECLIA) operating on a competitive binding principle. The use of a single platform throughout the five-year collection window was an intentional design feature, eliminating inter-method bias as a potential source of systematic error in the reference interval estimates. Principal analytical characteristics of the assay are summarized in (Table 1), and corroborate the fitness-for-purpose of this platform for population-level RI work.

Table 1: Analytical Characteristics of the Elecsys FT4 III Assay Used Throughout the Study Period. CV, coefficient of variation; ECLIA, electrochemiluminescence immunoassay; QC, quality control; RI, reference interval; WHO, World Health Organization. CV data represent manufacturer-claimed performance verified independently by the participating laboratory against third-party commercial controls across the 2020–2024 study window.

Parameter	Specification / Value
Assay platform	Elecsys FT4 III (Roche Diagnostics, Mannheim, Germany)
Detection principle	Electrochemiluminescence immunoassay (ECLIA); competitive binding format
Measurand	Free thyroxine (FT4) in human serum or plasma
Reporting units	ng/dL (SI conversion : $\times 12.871 = \text{pmol/L}$)
Measuring range	0.023 – 7.77 ng/dL (0.30 – 100 pmol/L)
Within-run CV	$\leq 2.1\%$ across the clinical decision range
Between-run CV	$\leq 3.4\%$ (manufacturer-reported; verified by participating laboratory)
Manufacturer RI (adults)	0.93 – 1.70 ng/dL (12.0 – 21.9 pmol/L)
Calibration traceability	Traceable to a candidate reference measurement procedure; standardised against the WHO 1st International Standard for T4 (67/049)
Internal QC scheme	Westgard multi-rule protocol applied daily; results released only when all QC targets met

Internal Quality Control (IQC) was maintained daily using two levels of commercial aqueous controls (BioRad Lyphochek Immunoassay Plus, levels 1 and 2), interpreted according to Westgard multi-rule criteria (1^{3s} , 2^{2s} , R^{4s} , 4^{1s} , 10^x rules). Results were released into the LIS only after all IQC criteria were satisfied. The laboratory participates in the United Kingdom National External Quality Assessment Service (UK NEQAS) for Thyroid

Function scheme; z-scores for FT4 over the study period remained within ± 2 SD of the peer-group mean for all but two assessments, in which instrument recalibration restored performance. Those two run sequences were identified retrospectively in the LIS and the corresponding patient results were excluded from the study dataset as a precautionary measure.

Demographic Stratification

The eligible dataset was partitioned into six mutually exclusive demographic strata on the basis of biological sex and age category, as follows: female children (2-16 yr), female adolescents (16-18 yr), female adults (>18 yr), male children (2-16 yr), male adolescents (16-18 yr), and male adults (>18 yr). The boundaries of these categories reflect established physiological transitions in Hypothalamic-Pituitary-Thyroid (HPT) axis function: the shift from childhood to adolescence is associated with rising gonadal steroid concentrations that modulate Thyroxine-Binding Globulin (TBG) synthesis and central TSH secretory dynamics; the transition into adulthood marks a stabilization of the reproductive hormonal milieu. Stratification by biological sex was considered essential given consistent epidemiological and physiological evidence for sex-specific differences in thyroid hormone metabolism, TBG concentration, and HPT sensitivity. Records for which age was missing or implausible (>100 yr), or for which biological sex was absent or recorded as indeterminate, were excluded from the stratified analyses but retained in an aggregate pool for preliminary distributional inspection. This additional exclusion resulted in the removal of 265 records (2.65% of the pre-stratification dataset), yielding a final analysable sample of 9,735 specimens distributed across the six strata, with sizes ranging from $n = 82$ (male adolescents) to $n = 5,676$ (female adults).

Reference Interval Derivation

Within each stratum, the arithmetic mean ($\bar{x}$) and Standard Deviation (SD) of FT4 values was computed from the raw data. To mitigate the influence of extreme values arising from undetected subclinical thyroid disease—an inherent limitation of the indirect method—a standard iterative truncation procedure was applied: all observations falling outside the range of $\bar{x} \pm 2SD$ were excluded. The mean and SD were then recalculated from the trimmed distribution, and this refined $\bar{x} \pm 2SD$ was adopted as the stratum-

specific reference interval, spanning from the 2.5th to the 97.5th percentile of the central reference population under the assumption of approximate normality post-truncation.

The implicit normality assumption was examined within each stratum using formal tests of distributional fit (detailed in Section 2.6). In strata where residual skewness persisted after truncation, the RI boundaries were cross-validated against empirical 2.5th and 97.5th percentiles obtained by non-parametric ranking, to assess whether the parametric and non-parametric estimates diverged by more than 5%—the threshold beyond which a Box-Cox power transformation or fully non-parametric derivation would have been adopted. In practice, this threshold was not exceeded in any stratum, confirming the adequacy of the parametric approach. Aggregate (sex-pooled, all-age) reference intervals were additionally derived for females ($n = 6,699$) and males ($n = 3,036$) using the same methodology applied to each pooled, trimmed distribution.

Statistical Analysis

Beyond RI derivation, a structured programmed of inferential and correlational analyses was pre-specified to characterize the relationships between FT4 and the primary demographic covariates of age and sex, and to quantify the clinical relevance of deviations from the manufacturer's generic interval. All statistical procedures were implemented in R [15] and verified against output from Python 3.11 (scipy 1.11 and stats models 0.14). The full analytical pipeline is summarized in (Table 2). Distributional normality was assessed within each stratum using the Shapiro-Wilk test for strata with $n \leq 5,000$, and the D'Agostino-Pearson omnibus test for the larger female adult stratum ($n = 5,676$), where the Shapiro-Wilk statistic loses discriminatory sensitivity. Homogeneity of variance across the six strata was evaluated with Levene's test (centre = median), selected for its robustness to departure from normality relative to Bartlett's test.

Table 2. Summary of Statistical Procedures, Software, and Analytical Objectives OLS, ordinary least squares; RI, reference interval; LRL, lower reference limit; URL, upper reference limit; CI, confidence interval; CV, coefficient of variation. Effect size r for rank-based comparisons interpreted as: < 0.10, negligible; 0.10–0.29, small; 0.30–0.49, moderate. Cohen's f^2 for regression: 0.02 = small, 0.15 = medium. Significance level $\alpha = 0.05$ for all tests; Bonferroni correction applied wherever multiple pairwise comparisons were made simultaneously.

Analytical Objective	Statistical Procedure	Software / Reference	Outputs Generated
Outlier truncation for RI derivation	Iterative mean $\pm$ 2SD exclusion	Applied within each stratum prior to final RI calculation	Trimmed distributions; final mean $\pm$ SD per stratum
Normality assessment	Shapiro-Wilk test ($n \leq 5,000$); D'Agostino-Pearson omnibus (larger strata)	R base package ; scipy.stats	W / K^2 statistics; p-values; Q-Q plots
Variance homogeneity	Levene's test (centre = median)	R car package (Fox & Weisberg, 2019)	F-statistic; df; p-value
Omnibus group comparison	Kruskal-Wallis H-test across 6 strata	R kruskal.test()	H-statistic; df; p-value
Post-hoc pairwise tests	Dunn test with Bonferroni correction	R dunn.test package (Dinno, 2017)	z-statistic; adjusted p-value; effect size $r = z/\sqrt{n}$

Continuous correlations	Spearman ρ (age vs. FT4, stratified by sex and developmental window)	R cor.test(); 2,000-iteration bootstrap CI	ρ ; 95% CI; p-value; ρ^2 (variance explained)
Effect decomposition	Multiple linear regression (OLS): FT4 ~ sex + age + sex×age; adults only (n = 9,735)	R lm (); age centred at mean (42.3 yr)	β coefficients; SE; t; p; adjusted R ² ; Cohen's f ²
Boundary discrepancy	Δ LRL and Δ URL vs. manufacturer RI; Cohen's d-analogue = displacement ÷ pooled SD	Calculated from stratum-specific SD estimates	Absolute boundary deviations; standardised effect sizes
Significance threshold	$\alpha = 0.05$ (two-tailed) for all inferential tests; Bonferroni-corrected where multiple comparisons were made	Pre-specified before data analysis	All p-values reported to three decimal places (< 0.001 as floor)

Inter-stratum differences in FT4 distributions were assessed using the Kruskal-Wallis H-test as the primary omnibus comparison, given confirmed heteroscedasticity and the conservative desire to avoid parametric assumptions across all six groups simultaneously. Significant omnibus results were followed by Dunn's test with Bonferroni correction for all pairwise contrasts. Effect sizes for rank-based comparisons were expressed as $r = z / \sqrt{n}$, where n denotes the combined sample size of each pair, and interpreted against established thresholds (negligible: $r < 0.10$; small: 0.10-0.29; moderate: 0.30-0.49). The continuous relationship between age and FT4 within each sex was quantified using Spearman's rank correlation coefficient (ρ), with 95% confidence intervals derived from 2,000-iteration bias-corrected and accelerated (BCa) bootstrap resampling. Correlations were computed for (i) the full age range within each sex, (ii) the paediatric sub-window (2-16 yr) within each sex, and (iii) the adult sub-window (>18 yr) within each sex, to identify whether the age-FT4 relationship showed developmental heterogeneity across the lifespan.

To disentangle the independent and interactive contributions of sex and age to FT4 concentration, an ordinary least squares multiple linear regression model was fitted to the full adult dataset (n = 9,735; age ≥ 18 yr): FT4 = $\beta_0 + \beta_1(\text{sex}) + \beta_2(\text{age}) + \beta_3(\text{sex} \times \text{age}) + \epsilon$. Sex was coded binary (male = 1, female = 0, reference category: female adults); age was mean-centred prior to entry (grand mean = 42.3 yr) to reduce multicollinearity and permit direct interpretation of main effects at the mean of the other predictor. Model diagnostics included visual inspection of residual-versus-fitted and Q-Q plots for residual normality and homoscedasticity, and computation of Variance Inflation Factors (VIF) to detect multicollinearity (threshold VIF > 5). Overall model fit was reported as adjusted R² and the omnibus F-statistic. The practical effect size at the model level was quantified as Cohen's f² = R² / (1-R²), interpreted against conventional benchmarks (0.02 = small, 0.15 = medium).

A boundary-discrepancy analysis was performed for each stratum by computing the absolute deviation between the population-derived RI limits (LRL and URL) and the corresponding limits of the manufacturer-supplied generic interval (0.93-1.70 ng/dL). A standardized effect size analogous to Cohen's d was derived for each boundary displacement by dividing the absolute deviation by the pooled standard deviation of the respective stratum, providing a scale-independent index of clinical relevance (thresholds: <0.20 = negligible; 0.20-0.49 = small; 0.50-0.79 = moderate).

Comparison with Manufacturer-Supplied Reference Interval

The generic adult RI published in the Elecsys FT4 III package insert (0.93-1.70 ng/dL; 12.0-21.9 pmol/L) served as the primary external comparator throughout this study. This interval was derived by the manufacturer from a cohort of healthy adult volunteers predominantly recruited from Western European and North American centres, using a direct prospective design with rigorous exclusion criteria. The comparison was made at the level of absolute boundary values and standardised boundary displacements (Section 2.6), and the clinical consequences of applying the generic interval to the Syrian dataset were characterised in terms of the proportion of the Syrian population that would be reclassified as outside the reference space under each interval.

Results

Study Cohort and Distribution of Specimens

Of the 10,000 FT4 measurements retrieved from the laboratory information system across Syrian clinical facilities between January 2020 and December 2024, all records met pre-specified analytical quality criteria and were carried forward for stratified reference interval derivation. The total dataset comprised 6,699 female (67.0%) and 3,036 male (30.4%) specimens, with the remainder representing records of undetermined or unregistered sex that were excluded from sex-stratified analyses but retained for visual inspection of the aggregate distribution. Age ranged from 2 to well beyond 18 years across both sexes. Stratification into three biologically grounded categories-children (2-16 yr), adolescents (16-18yr), and adults (>18yr)-was adopted to reflect recognised developmental transitions in Hypothalamic-Pituitary-Thyroid (HPT) axis regulation.

Within the six demographic strata, specimen counts varied considerably, from n = 82 in the male adolescent group to n=5,676 in female adults. Although the smaller subgroups carry wider confidence intervals around their reference limits, the aggregate sample sizes in every stratum exceeded the minimum of 40 observations recommended by the CLSI EP28-A3c guideline for non-parametric reference interval estimation, and exceeded 120 in all but the male adolescent cohort. Notably, the adolescent strata were the most challenging to populate, consistent with the well-documented under-representation of this age group in routine

laboratory databases owing to lower healthcare utilization rates during this period of life.

Age and Sex-Specific Reference Intervals

Using the indirect method with iterative mean ± 2 SD truncation, age- and sex-specific reference intervals (RIs) for FT4 were derived for each stratum. The resulting intervals, together with descriptive statistics, are presented in (Table 3). Across female cohorts, a discernible pattern emerged whereby the upper reference limit (URL) contracted modestly from childhood (URL = 1.83 ng/dL)

through adolescence (URL = 1.74 ng/dL), before widening again in adulthood (URL = 1.82 ng/dL). This adolescent contraction may reflect the modulatory influence of rising oestrogen concentrations on Thyroxine-Binding Globulin (TBG) capacity and on central HPT setpoint, an interpretation supported by prior physiological studies. In contrast, the Lower Reference Limit (LRL) among females reached its nadir in adulthood (LRL = 0.71 ng/dL), approximately 9-10% below the paediatric and adolescent lower bounds, suggesting a gradual broadening of the lower tail of the distribution with reproductive maturation and beyond.

Table 3: Age- and Sex-Specific Reference Intervals for FT4 in the Syrian Population (2020–2024), with Descriptive Statistics IQR, interquartile range; RI, reference interval (2.5th–97.5th percentile). Mean $\pm$ SD and median values were computed from the truncated distributions following iterative exclusion of observations outside mean ± 2 SD. Reference intervals represent the central 95% of the trimmed distribution in each stratum.

Demographic Group	Age Range (yr)	n	Mean $\pm$ SD (ng/dL)	Median (ng/dL)	IQR (ng/dL)	Reference Interval (ng/dL)
Female Children	2–16	912	1.31 $\pm$ 0.26	1.29	0.35	0.80 – 1.83
Female Adolescents	16–18	123	1.27 $\pm$ 0.24	1.25	0.31	0.81 – 1.74
Female Adults	> 18	5,676	1.27 $\pm$ 0.28	1.25	0.38	0.71 – 1.82
Male Children	2–16	886	1.36 $\pm$ 0.24	1.34	0.3	0.89 – 1.82
Male Adolescents	16–18	82	1.34 $\pm$ 0.26	1.33	0.34	0.83 – 1.85
Male Adults	> 18	2,061	1.31 $\pm$ 0.29	1.3	0.39	0.72 – 1.89
All Females (pooled)	All	6,699	1.27 $\pm$ 0.28	1.25	0.37	0.72 – 1.82
All Males (pooled)	All	3,036	1.31 $\pm$ 0.28	1.29	0.38	0.77 – 1.88

Male cohorts displayed a directionally different trajectory at the LRL: children exhibited the highest lower bound among all subgroups at 0.89 ng/dL, which declined to 0.83 ng/dL in adolescents and reached 0.72 ng/dL in adults—a net downward shift of 0.17 ng/dL spanning three life stages. The URL, by contrast, remained relatively stable across male subgroups (1.82–1.89 ng/dL), with adult males exhibiting the highest overall URL in the entire dataset. When both sexes were pooled within each age category irrespective of sex-stratification, the aggregate female RI was 0.72–1.82 ng/dL (n = 6,699) and the aggregate male RI was 0.77–1.88 ng/dL (n = 3,036), yielding interval widths of 1.10 and 1.11 ng/dL, respectively—virtually indistinguishable in absolute breadth but meaningfully displaced along the concentration axis.

Inferential Statistical Comparisons Across Demographic

Table 4: Post-hoc Pairwise Comparisons of FT4 Distributions Across Demographic Strata (Dunn Test, Bonferroni-Corrected) p_{aaa} , p-value adjusted for multiple comparisons using the Bonferroni correction. Effect size $r = z / \sqrt{n}$, where n is the total sample size contributing to each pairwise comparison. Thresholds: $r < 0.10$ = negligible; 0.10–0.29 = small; 0.30–0.49 = moderate. Δ Median indicates the difference (female group minus male, or older minus younger).

ssPairwise Comparison	Δ Median (ng/dL)	z-statistic	p_{aaa} (Bonferroni)	Effect Size r	Interpretation
Female Adults vs. Male Adults	0.01	2.38	0.017	0.04	Statistically significant; negligible effect

Strata

To determine whether observed inter-stratum differences in FT4 distributions reflected genuine population-level heterogeneity rather than sampling variability, a Kruskal-Wallis H-test was performed across all six demographic subgroups. The omnibus test returned $H(5) = 41.7$, $p < 0.001$, conclusively rejecting the null hypothesis of distributional equivalence. Levene's test for homogeneity of variances confirmed significant heteroscedasticity across strata ($F(5, 9,729) = 5.14$, $p < 0.001$), reinforcing the appropriateness of stratum-specific rather than pooled reference intervals. Post-hoc pairwise comparisons were conducted using the Dunn test with Bonferroni correction; results are detailed in (Table 4).

Female Children vs. Male Children	-0.05	3.14	0.002	0.07	Significant; small effect: males higher at baseline
Female Adults vs. Female Adolescents	0	0.89	0.374	0.03	Non-significant; no meaningful developmental shift
Female Adults vs. Female Children	0.04	3.71	<0.001	0.09	Significant; small effect across development
Male Adults vs. Male Children	-0.04	2.99	0.003	0.08	Significant; marginal attenuation of FT4 with age
Male Adults vs. Male Adolescents	-0.03	1.64	0.101	0.05	Non-significant; continuity across puberty

The most pronounced statistically significant difference was observed between female children and male children ($r = 0.07$), where male prepubertal subjects maintained a substantially higher LRL (0.89 vs. 0.80 ng/dL), consistent with the influence of male sex steroids-even at low pubertal concentrations-on thyroidal secretory tone. In contrast, no significant difference was detected between female adults and female adolescents ($p = 0.374$), nor between male adults and male adolescents ($p = 0.101$), suggesting that the transition through puberty does not fundamentally alter the central tendency of FT4 within each sex, even if boundary limits shift. Across all significant comparisons, effect sizes (r) remained in the small range (0.04-0.09), underscoring that while differences are statistically reproducible at this sample size, they are modest in absolute clinical magnitude and argue against broad partition on

the basis of age alone.

Correlation Between FT4 and Age: Spearman Analysis by Stratum

To quantify the continuous relationship between FT4 concentration and age within each sex, Spearman rank-order correlation coefficients (ρ) were computed for both the full age range and separately for paediatric and adult sub-windows. Results, including 95% bootstrap confidence intervals and proportion of variance explained (ρ^2), are presented in (Table 5). The non-parametric approach was selected in recognition of the right-skewed age distribution inherent to routine laboratory datasets and the non-linear nature of HPT maturation across the lifespan.

Table 5: Spearman Rank Correlation Between FT4 Concentration and Age, Stratified by Sex and Developmental Window 95% confidence intervals derived from 2,000-iteration bootstrap resampling. ρ^2 denotes the proportion of variance in FT4 ranks accounted for by age ranks within each stratum. All p-values are two-tailed. Non-significant results ($p \geq 0.05$) are italicized.

Analytical Stratum	Spearman ρ	95% CI	p-value	ρ^2 (var. expl.)	Directional Interpretation
Females: Age vs. FT4 (all ages)	-0.14	[-0.17, -0.11]	< 0.001	2.00%	Weak negative association; modest age-related FT4 decline
Males: Age vs. FT4 (all ages)	-0.09	[-0.13, -0.05]	< 0.001	0.80%	Weak negative; attenuated relative to females
Females: Age vs. FT4 (children only)	-0.04	[-0.10, +0.03]	0.213	0.20%	Non-significant; stable FT4 throughout childhood
Males: Age vs. FT4 (children only)	-0.03	[-0.10, +0.04]	0.381	0.10%	Non-significant; corroborates female pediatric pattern
Females: Age vs. FT4 (adults only)	-0.11	[-0.14, -0.08]	< 0.001	1.20%	Sustained gradual decline into adulthood
Males: Age vs. FT4 (adults only)	-0.08	[-0.12, -0.04]	< 0.001	0.60%	Parallel adult decline, slightly less pronounced

The data reveal a consistent, albeit weak, negative association between advancing age and FT4 across the full lifespan in both females ($\rho = -0.14$; 95% CI: -0.17, -0.11) and males ($\rho = -0.09$; 95% CI: -0.13, -0.05), with females exhibiting a marginally steeper gradient. The proportion of variance explained by age was correspondingly modest-2.0% in females and 0.8% in males-signifying that, while the age trajectory is statistically real and directionally reproducible, the overwhelming majority of inter-individual variability in FT4 is

attributable to factors other than chronological age alone, including genetic polymorphisms in thyroid hormone synthesis, iodine intake, body mass index, and TSH receptor sensitivity.

Within the paediatric window (2-16 yr), correlations were non-significant in both sexes (females: $\rho = -0.04$, $p = 0.21$; males: $\rho = -0.03$, $p = 0.38$). This finding carries a directly applicable clinical message: a single, sex-stratified reference interval is likely sufficient across the 2-16year paediatric range, without the need

for additional fine-grained age partitioning that would further fragment already modest sample sizes. The data do not support a developmental inflection point within childhood that would necessitate separate intervals for, say, children aged 2-8 versus 9-16. In adults, both sexes displayed modest but statistically robust negative correlations (females: $\rho = -0.11$; males: $\rho = -0.08$), consistent with well-established age-related attenuation of thyroid secretory reserve.

Multiple Linear Regression and Effect Decomposition: Sex, Age, and Their Interaction

Table 6: Multiple Linear Regression of FT4 on Sex and Age in Syrian Adults ($n = 9,735$): Coefficient Estimates and Model Diagnostics
Reference category: adult females (sex coded 0). Age centred at mean (42.3 yr) prior to entry to facilitate interpretation of main effects and reduce multicollinearity. *Adjusted R^2 and omnibus F-statistic pertain to the overall model. Cohen's $f^2 = 0.021 / (1 - 0.021) \approx 0.022$, indicating a small but non-trivial effect size at the model level.

Model Parameter	Coefficient (β)	Std. Error	t-statistic	p-value	Interpretation
Intercept (reference: female adults)	1.268	0.004	316.2	< 0.001	Estimated mean FT4 in reference group
Sex (male vs. female)	0.043	0.009	4.78	< 0.001	Males average 0.043 ng/dL higher; modest but consistent
Age (continuous, years)	-0.002	0.001	-2.34	0.019	Each additional year associated with 0.002 ng/dL decrease
Sex $\times$ Age interaction	0.001	0.001	0.94	0.347	No significant sex-by-age moderation; parallel trajectories
Model fit: Adjusted $R^2 = 0.021$; F (3, 9731) = 69.8; $p < 0.001$ — age and sex jointly explain 2.1% of FT4 variance	—	—	—	—	Small but statistically robust proportion of explained variance

Sex emerged as a significant independent predictor of FT4, with males averaging 0.043 ng/dL higher than females after adjusting for age ($\beta = +0.043$, SE = 0.009, $t = 4.78$, $p < 0.001$). Though modest in absolute magnitude, this sex differential maps onto meaningful boundary displacement at the population level—particularly at the URL—as corroborated by the reference interval data. Age contributed a small but independent negative effect ($\beta = -0.002$ ng/dL per year; $p = 0.019$), consistent with the Spearman findings. Critically, the sex $\times$ age interaction term was non-significant ($\beta = +0.001$, $p = 0.347$), indicating that the age-related decline in FT4 proceeds at essentially parallel rates in both sexes, without evidence of sex-specific acceleration or deceleration of the thyroid ageing process. The model's adjusted R^2 of 0.021 confirms that sex and age jointly account for a statistically robust but numerically small proportion

of sex and age to FT4 variability, a multiple linear regression model was fitted to the full adult dataset ($n = 9,735$; subjects ≥ 18 yr), treating sex (binary: male = 1, female = 0) and age (continuous, years) as predictors and including a sex-by-age interaction term to test for possible moderation. Ordinary least squares estimation was used; model diagnostics confirmed approximately homoscedastic residuals upon visual inspection of residual-versus-fitted plots, with a modest positive skew at the upper tail consistent with the known right-skew of FT4 distributions at population level. Results are summarised in (Table 6).

of total FT4 variance, reaffirming that the largest contributors to inter-individual dispersion remain unmeasured in this dataset.

Deviation from the Manufacturer-Supplied Generic Reference Interval

A formal boundary discrepancy analysis was performed by comparing each population-specific RI against the manufacturer-supplied generic interval of 0.93-1.70 ng/dL. For each subgroup, the absolute deviation at the lower reference limit (Δ LRL) and Upper Reference Limit (Δ URL) was calculated, and an approximate standardised effect size (analogous to Cohen's d) was derived by expressing the boundary displacement as a fraction of the pooled standard deviation across the Syrian and manufacturer-implied distributions. Findings are detailed in (Table 7).

Table 7: Boundary Discrepancy Analysis: Syrian Population-Specific FT4 Reference Intervals versus Manufacturer-Supplied Generic Range Δ LRL = Syrian LRL – Manufacturer LRL; Δ URL = Syrian URL – Manufacturer URL. Negative Δ LRL values (red) indicate that the Syrian lower limit falls below the manufacturer threshold, increasing the risk of misclassifying truly hypothyroid patients as borderline-normal. Positive Δ URL values (green) indicate that the Syrian upper limit exceeds the manufacturer threshold, raising the risk of under-detecting hyperthyroidism. *Cohen's d-analogue computed as boundary displacement $\div$ pooled SD of the respective stratum. Thresholds: < 0.20 = negligible; 0.20–0.49 = small; 0.50–0.79 = moderate.

Group	Syrian RI (ng/dL)	Manufacturer RI (ng/dL)	Δ LRL	Δ URL	Cohen d*	Misclassification Risk
Female Adults	0.71–1.82	0.93–1.70	–0.22	0.12	0.34	Under-diagnosis of hypothyroidism; over-diagnosis of hyperthyroidism
Male Adults	0.72–1.89	0.93–1.70	–0.21	0.19	0.37	Particularly elevated URL; hyperthyroid false positives in healthy males
All Syrian Adults (pooled)	0.74–1.85	0.93–1.70	–0.19	0.15	0.35	Systematic dual-boundary displacement across the entire adult population
Female Children	0.80–1.83	0.93–1.70	–0.13	0.13	0.24	Moderate boundary shift; paediatric diagnostic thresholds affected
Male Children	0.89–1.82	0.93–1.70	–0.04	0.12	0.16	Lower limit nearly concordant; upper limit discrepancy persists

The boundary discrepancy analysis exposes two structurally distinct forms of misalignment between the manufacturer's generic RI and the Syrian population data. At the lower boundary, all Syrian subgroups fell below the manufacturer's LRL of 0.93 ng/dL, with adult subgroups showing the largest divergence (Δ LRL ranging from –0.19 to –0.22 ng/dL). From a clinical standpoint, applying the generic LRL to Syrian patients would result in a substantial number of biochemically euthyroid individuals being flagged as having reduced FT4, potentially triggering unnecessary TSH and thyroid antibody reflex testing, radioiodine scintigraphy workup, or empirical levothyroxine initiation in patients who do not require it. The small-effect-size Cohen's d analogues (0.16–0.37) indicate that these boundary displacements, while consistently unidirectional, are not so large as to arise from analytical platform bias alone and are more plausibly explained by genuine population-level differences in thyroid hormone setpoint.

At the upper boundary, Syrian adults-particularly males-exhibited FT4 URLs that systematically exceeded the manufacturer's URL of 1.70 ng/dL by 0.12–0.19 ng/dL. The magnitude of the upper boundary displacement was greatest in male adults (Δ URL = +0.19 ng/dL; Cohen's d-analogue = 0.37), and the pooled adult cohort showed a Δ URL of +0.15 ng/dL relative to the generic range. This upward displacement of the URL carries the clinically inverse risk: Syrian patients whose FT4 falls between 1.70 and 1.85–1.89 ng/dL would be incorrectly classified as hyperthyroid if the manufacturer's RI is applied, despite falling within the locally appropriate reference space. In a clinical context where FT4 is often used to adjudicate borderline TSH suppression or to titrate antithyroid therapy, such misclassification could translate into premature dose escalation, unnecessary beta-blockade, or inappropriate radioiodine ablation referrals.

The dual-directional nature of these discrepancies-with

simultaneous downward displacement of the LRL and upward displacement of the URL-indicates not merely a scale offset but a systematic broadening and rightward shift of the entire FT4 distribution in the Syrian population relative to the manufacturer's reference cohort. This pattern is consistent with previously reported population-specific variation in thyroid hormone physiology attributable to differences in iodine status, selenium and zinc intake, genetic variation in *TSHR* and *DIO2* loci, body composition, and possibly the chronic low-grade inflammatory milieu associated with prolonged population-level stress exposure-a factor of particular relevance given the socioeconomic context of the Syrian population during the 2020–2024 study window.

Synthesis of Findings

Taken in aggregate, the analyses presented here converge on four analytically grounded conclusions. First, FT4 reference intervals in the Syrian population are demonstrably sex- and age-dependent, with statistically significant inter-stratum heterogeneity (Kruskal-Wallis $H = 41.7$, $p < 0.001$) justifying the use of partitioned rather than universal intervals. Second, the magnitude of sex and age effects-while statistically robust-is small, with regression modelling attributing only 2.1% of total FT4 variance to these two factors combined (adjusted $R^2 = 0.021$); the biological implication is that substantial inter-individual variability persists within each stratum and cannot be resolved by demographic partitioning alone. Third, age has negligible influence on FT4 within the 2–16-year pediatric window in both sexes ($p \approx -0.03$ to -0.04 , $p > 0.20$), supporting the use of a single pediatric interval per sex without further age sub-stratification. Fourth, and most practically urgent, the manufacturer-supplied generic RI of 0.93–1.70 ng/dL is structurally misaligned with the Syrian population at both boundaries, exhibiting a downward LRL displacement of 0.13–0.22 ng/dL and an upward URL displacement of 0.12–0.19 ng/dL across

all adult and pediatric subgroups, with effect sizes in the small-to-moderate range that carry direct diagnostic consequences. These findings collectively provide the quantitative and statistical foundation for recommending the adoption of the population-derived, sex- and age-specific reference intervals presented in Table 1 as the operative standards for FT4 interpretation in Syrian clinical laboratories, pending prospective validation in independent cohorts drawn from this population.

Discussion

Thyroid gland plays significant role in physiology by regulating the body functions through thyroid hormones namely Thyroxine (T4) and Thyroid-Stimulating Hormone (TSH). Accurate determination of thyroid gland functions is essential for the management and diagnosis of thyroid diseases especially in populations with a different genetic, environmental or even different lifestyle. The aim of this review is to discuss the biological, clinical and diagnostic implications of the discrepancies in the Thyroid Function Tests (TFTs) reference intervals in Syrian population based on the latest studies.

Reference intervals that do not account for individual biological variation or that are based on homogenous populations contain potential for inaccuracy, particularly in multifaceted individuals. The establishment of reference intervals should also consider ethnic and demographic factors that may have influence [16]. Thyroid dysfunction in Syria is said to present differently, owing to the genetic makeup of the Syrian population and the environmental factors present in their region. Hence, the thyroid function reference values established in other regions of the world may not be the same for the Syrian population, which may consequently lead to thyroid dysfunction remaining undiagnosed or being misdiagnosed [17]. In pregnant women, it has been established that inappropriate levels of TSH and FT4 can have detrimental effects on the health of the mother as well as the fetus [8,18]. reviewed studies on pregnancy-specific thyroid hormone reference intervals and found a clear cross-ethnic variation which could impact the need for ethnicity-specific references. This supports the need for locally conducted studies taking into account racial differences in thyroid physiology. Similarly, several studies suggest that local and ethnicity specific references for the TFTs need to be derived for the Syrian population as it has not been extensively studied [19].

The clinical implications of altered thyroid function Reference Intervals (RIVs) are numerous. For example, treatment of patients with thyroid dysfunction may be inappropriately directed due to clinical judgment that is based on misaligned RIVs. For example, over-treatment and under-treatment of hypothyroid patients have been reported due to non-specific RIVs that result in inappropriate interpretation of TSH values [17]. Developing ethnicity-specific RIVs for thyroid function will help in better assessment of thyroid status and diagnosis of various thyroid conditions, such as subclinical hypothyroidism that is of clinical importance and significantly affects patient's quality of life [20,21] have shown that

the RIVs for TSH vary in different ethnic groups living in Iran and recommend local population including Syrian population to have local and comprehensive RIVs, which match the characteristics of the population under study.

Alongside the biological and clinical importance of the subject, there are several diagnostic challenges due to the use of common reference intervals. The potential for misdiagnosis still exists [22] due to lack of taking into account population specific criteria, and a high TSH level in a patient from Syria cannot be evaluated by referring to the common criteria. Hence, there is a great need for establishing ethnicity-specific population-based reference intervals for safe evaluation of the patient and to get the best management and therapeutic decisions that in turn lead to better therapeutic outcome and improved quality of life [23]. More recently, it was proven that ethnicity-specific reference intervals can lead to a more accurate diagnosis and therefore to more effective treatment for a wider variety of populations [24].

It was noted that gender and age factors make up some of the factors that contribute to the challenges faced in trying to determine misaligned reference ranges. "Thyroid hormone variability is affected by numerous factors including gender and age and is therefore a non-binary factor that contributes to the challenges in determining misaligned reference ranges; and specific reference ranges may differ in relation to different populations," the review by *Al-Hetar et al.* stated. The team adds that there are many other studies that show changes in thyroid hormones levels at different ages. The use of a one size fits all approach to thyroid function test results may miss the mark as far as accurately and effectively treating a large number of patients. A large number of patients may be incorrectly diagnosed and given inappropriate treatment as a result of a failure to adjust lab results to appropriate reference ranges.

Another factor affecting the standardization of reference intervals is the progress of laboratory techniques and assay methods. In their study [8], demonstrated method-dependent differences in TSH assays that demand reassessment of reference standards, especially in populations where healthcare and laboratory infrastructure may vary significantly. Without tailoring reference intervals to the specific healthcare context, not only will laboratory diagnosis remain associated with the risk of misleading and inappropriate management of patients with suspected hyper or hypothyroidism, but it will also ensure continuous misdiagnosis and inadequate management of affected individuals. Therefore, the Syrian health sector urgently needs reliable studies and a uniform basis for laboratory diagnosis of thyroid dysfunction.

The impacts of widened reference intervals on Thyroid Function Tests (TFTs) in the Syrian population have been determined at a biological, clinical and diagnostic level. Localized reference intervals are necessary to ensure accurate diagnosis and management of thyroid diseases. Recent studies have revealed the complexity of the regulation of the thyroid hormones and highlight the necessity

of culturally and ethnically sensitive investigations. This research affords an urgent need for further investigation into a number of reference intervals to ensure accurate investigations of the TFTs for a robust and applicable Syrian population in order to reduce the observed gap between healthcare and patient therapeutic pathways in this region.

Limitations

Several limitations inherent to the study design require acknowledgement. The indirect method, while pragmatically necessary given the scale of the dataset, cannot fully exclude the contribution of undetected subclinical thyroid dysfunction to the tails of each stratum's distribution, notwithstanding the iterative outlier exclusion applied. The single-centre design, although the participating laboratory serves a broad urban and peri-urban catchment, may not capture the full geographical and socioeconomic heterogeneity of the Syrian population - particularly rural communities with distinct iodine exposure profiles. Stratum-specific sample sizes were modest for the adolescent subgroups (female adolescents: $n = 123$; male adolescents: $n = 82$), widening the confidence intervals around those intervals and limiting their precision. Finally, the derived RIs are platform-specific to the Elecsys FT4 III assay and are not directly transferable to other immunoassay platforms without cross-verification, given the known inter-method variability that persists in FT4 measurement despite ongoing standardization efforts [25-27].

Conclusion

This investigation yields a finding that is both statistically robust and clinically consequential: the manufacturer-supplied reference interval for the Elecsys FT4 III assay (0.93-1.70 ng/dL) does not represent the Syrian population. Every one of the six demographic strata examined-spanning children, adolescents, and adults of both sexes - returned a lower reference limit below the manufacturer's threshold, with adult deviations reaching -0.22 ng/dL, while the upper limit was exceeded across all subgroups, most markedly in adult males (+0.19 ng/dL). Translated into practice, this means that a meaningful proportion of biochemically euthyroid Syrian patients would be misclassified as hypothyroid at the lower boundary, or remain undetected as hyperthyroid at the upper boundary, whenever the generic interval is applied uncritically. Neither error is benign: the former risks unnecessary levothyroxine exposure and reflex investigation cascades; the latter risks delayed antithyroid treatment in patients who genuinely need it.

The statistical architecture of the study provides independent corroboration at each analytical level. The Kruskal-Wallis omnibus test ($H = 41.7$, $p < 0.001$) and Levene's test for variance homogeneity ($F = 5.14$, $p < 0.001$) confirm that the six strata are not exchangeable and cannot legitimately share a single reference interval. Spearman correlations establish that age-related FT4 decline, though statistically real in adults ($\rho = -0.09$ to -0.14), is absent within the 2-16-year pediatric window ($\rho \approx -0.03$ to -0.04 ;

$p > 0.20$), removing any statistical justification for further age sub-partitioning in childhood. Multiple linear regression additionally confirms sex as an independent determinant of FT4 ($\beta = +0.043$ ng/dL; $p < 0.001$), with age-related decline proceeding at parallel rates in both sexes (interaction term $p = 0.347$) - a finding that simplifies the practical implementation of stratified thresholds without sacrificing analytical rigour.

Placed within the wider literature, these results are not unexpected. Comparable divergence between population-specific and manufacturer-supplied thyroid hormone RIs has been documented in Lebanon [9], and the biological rationale - encompassing differences in dietary iodine status, genetic polymorphisms in *TSHR* and *DIO2*, and the chronic physiological burden of prolonged conflict - provides a plausible mechanistic explanation specific to the Syrian context. What distinguishes the present study is its scale ($n = 10,000$), its five-year temporal span, and the formal quantification of boundary displacements through standardized effect sizes, which place the magnitude of the discrepancy beyond the threshold attributable to assay imprecision alone. We recommend the immediate adoption of the sex- and age-specific reference intervals derived in this study as the operative standards for FT4 reporting in Syrian clinical laboratories using the Elecsys FT4 III platform. Implementation requires no additional laboratory infrastructure - only the substitution of locally validated limits for generic manufacturer values within the laboratory information system. Looking ahead, prospective multi-center validation across geographically and socioeconomically diverse Syrian cohorts, together with cross-platform verification on additional immunoassay systems, will be necessary to extend the reach and generalizability of these findings and to establish locally anchored thyroid hormone reference standards as a durable feature of Syrian laboratory medicine [28-31].

Acknowledgements

None.

Conflict of Interest

None.

References

1. Surup H, Vogel M, Koerner A, Hiemisch A, Oelkers L, et al. (2021) Pediatric reference intervals for thyrotropin, free triiodothyronine, and free thyroxine and the relevance of body mass index and puberty in measurement interpretation. *Thyroid* 31(8): 1192-1202.
2. Jansen HI, Dirks NF, Hillebrand JJ, Ten Boekel E, Brinkman JW, et al. (2024) Age-specific reference intervals for thyroid-stimulating hormones and free thyroxine to optimize diagnosis of thyroid disease. *Thyroid* 34(11): 1346-1355.
3. Allen GFG, Blampied S, Patel KA, McDonald TJ, Hattersley AT, et al. (2026) Age-related reference intervals for thyroid function tests are beneficial in adults over 60 years. *Ann Clin Biochem* 63(5): 394-402.
4. Li Q, Tang Y, Yu X, Qin G, Tian L, et al. (2025) Thyroid function reference intervals by age, sex, and race: a cross-sectional study. *Ann Intern Med* 178(7): 921-929.

5. Taylor PN, Lansdown A, Witczak J, Khan R, Rees A, et al. (2023) Age-related variation in thyroid function--a narrative review highlighting important implications for research and clinical practice. *Thyroid Res* 16(1): 7.
6. Wang D, Yu S, Ma C, Li H, Qiu L, et al. (2019) Reference intervals for thyroid-stimulating hormone, free thyroxine, and free triiodothyronine in elderly Chinese persons. *Clinical Chemistry and Laboratory Medicine (CCLM)* 57(7): 1044-1052.
7. Çatak Z, Fener H, Ayyıldız H, Çokgüler ZŞ (2026) Establishing Age-and Sex-Specific Reference Intervals for Thyroid Function Tests in the Older People of Eastern Anatolia: A Population-Based Indirect Approach. *Medicina*, 62(3): 425.
8. Okosieme OE, Agrawal M, Usman D, Evans C (2021) Method-dependent variation in TSH and FT4 reference intervals in pregnancy: a systematic review. *Ann Clin Biochem* 58(5): 537-546.
9. Salameh C, Al Achkar A, El Boustany J, Sleilaty G, Gannagé Yared MH (2022) Reference intervals for thyroid-stimulating hormone, free thyroxine, free triiodothyronine, and total triiodothyronine in the Lebanese adult population. *Annals of Clinical Biochemistry*, 59(4): 264-271.
10. Jonklaas J, Razvi S (2019) Reference intervals in the diagnosis of thyroid dysfunction: treating patients not numbers. *The Lancet Diabetes Endocrinology* 7(6): 473-483.
11. Mehran L, Amouzegar A, Bakhtiyari M, Mansournia MA, Rahimabad PK, et al. (2017) Variations in serum free thyroxine concentration within the reference range predicts the incidence of metabolic syndrome in non-obese adults: a cohort study. *Thyroid* 27(7): 886-893.
12. D Aurizio F, Kratzsch J, Gruson D, Petranović O, Včariček P, Giovanella L (2023) Free thyroxine measurement in clinical practice: how to optimize indications, analytical procedures, and interpretation criteria while waiting for global standardization. *Crit Rev Clin Lab Sci* 60(2): 101-140.
13. Xie T, Su M, Feng J, Pan X, Wang C, et al. (2023) The reference intervals for thyroid hormones: a four-year investigation in Chinese population. *Front Endocrinol (Lausanne)* 13: 1046381.
14. Roche Diagnostics (2020) Elecsys FT4 III: Method Sheet. Mannheim, Germany: Roche Diagnostics.
15. R Core Team (2023) R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.
16. Sivakumar S, Makija I, Bhagat R, Maurya S, Chatterjee NS, et al. (2025) Ethnicity-based variations in biological reference interval-A systematic scoping review. *Clin Chim Acta* 578: 120539.
17. Razvi S, Bhana S, Mrabeti S (2019) Challenges in interpreting thyroid stimulating hormone results in the diagnosis of thyroid dysfunction. *J Thyroid Res* 2019(1): 4106816.
18. Khalil AB, Salih BT, Chinengo O, Bardies MRD, Turner A, et al. (2018) Trimester specific reference ranges for serum TSH and Free T4 among United Arab Emirates pregnant women. *Pract Lab Med* 12: e00098.
19. Mahomed M (2025) Determination of reference intervals and decision limits for thyroid stimulating hormone and Thyroxine on cord blood samples. Master's thesis. University of Cape Town.
20. Veltri F, Belhomme J, Kleynen P, Grabczan L, Rozenberg S, et al. (2017) Maternal thyroid parameters in pregnant women with different ethnic backgrounds: Do ethnicity-specific reference ranges improve the diagnosis of subclinical hypothyroidism? *Clinical Endocrinology* 86(6): 830-836.
21. Meamar R, Feizi A, Aminorroaya A, Amini M, Nasri M, et al. (2021) Thyroid stimulating hormone reference range: Iranian thyroid cohort study. *Acta Bio Med* 92(5): e2021283.
22. Abdalla O (2021) Genome Wide Association Study of Serum Levels of Thyroid Hormones in Qatari Population. Master's thesis. Hamad Bin Khalifa University (Qatar).
23. Oron T, Lazar L, Feldhamer I, Manor M, Koren Morag N, et al. (2020) Pediatric reference values of TSH should be personalized according to BMI and ethnicity. *Eur J Endocrinol* 183(4): 419-426.
24. Tran C, Topliss DJ, Schneider HG, Seeman E, Clayton Chubb D, et al. (2025) Establishing the TSH reference intervals for healthy adults aged over 70 years: the Australian ASPREE cohort study. *Clin Chem Lab Med* 63(3): 611-620.
25. Colonnello E, Graziani A, Rossetti R, Voltan G, Masi D, et al. (2025) The Chronobiology of Hormone Administration: "Doctor, what time should I take my medication?". *Endocr Rev* 46(5): 670-689.
26. Spencer CA (2023) Laboratory thyroid tests: a historical perspective. *Thyroid* 33(4): 407-419.
27. Javaid, U., Kennedy, D., Addison, C., Tsalidis, V. and Razvi, S., 2022. Frequency, determinants and costs of thyroid function testing in a laboratory serving a large population. *European Journal of Endocrinology*, 186(5): 553-560.
28. Dinno A (2017) dunn.test: Dunn's Test of Multiple Comparisons Using Rank Sums. R package version 1.3.5.
29. Fox J, Weisberg S (2019) An R Companion to Applied Regression. 3rd ed.
30. World Health Organization (2019) WHO International Standard 1st International Standard for Thyroxine (T4).
31. Yamada S, Horiguchi K, Akuzawa M, Sakamaki K, Yamada E, et al. (2023) 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* 33(4): 428-439.