# Changing Immunochemistry Platforms : Thyroid Function Test Comparison and Reference Intervals Based on Clinical Needs

Pelanti, Jonna

2022-10-29

Pelanti , J , Lamberg , T , Salopuro , T , Pussinen , C , Suvisaari , J , Joutsi-Korhonen , L , Schalin-Jäntti , C , Itkonen , O & Anttonen , M 2022 , ' Changing Immunochemistry Platforms : Thyroid Function Test Comparison and Reference Intervals Based on Clinical Needs ' , The journal of applied laboratory medicine , vol. 7 , no. 6 , pp. 1438-1444 . https://doi.org/10.1093/jalm/jfac067

http://hdl.handle.net/10138/350580
https://doi.org/10.1093/jalm/jfac067

# Changing Immunochemistry Platforms: Thyroid Function Test Comparison and Reference Intervals Based on Clinical Needs

**Jonna Pelanti [a,b,*] Tea Lamberg,[a] Titta Salopuro,[a] Christel Pussinen,[a] Janne Suvisaari,[a] Lotta Joutsi-Korhonen,[a] Camilla Schalin-Jäntti,[c] Outi Itkonen,[a] and Mikko Anttonen[a]**

**Background:** Diagnosis of thyroid dysfunction relies on thyroid stimulating hormone (TSH), free thyroxine (FT4), and free tri-iodothyronine (FT3) tests against valid reference intervals (RIs). We changed the immunoassay platform from Abbott Architect to Siemens Atellica and aimed to establish Atellica RIs based on laboratory information system (LIS) patient data.

**Methods:** Atellica thyroid hormone immunoassays were verified against those of Architect. Real-life patient results were retrieved from LIS. A single result per patient dataset was used to establish the RIs by the indirect method.

**Results:** Atellica and Architect assays correlated well but Atellica showed a positive bias between 13% and 53%, the largest for FT4. Variations of the Atellica assays were ≤4%. The 95% Atellica RIs were 0.4–3.8 mU/L for TSH, 0.9–1.6 ng/dL for FT4, and 227–416 pg/dL for FT3. Considering the accumulating clinical experience with Atellica, the RIs for clinical use were adjusted as 0.5–4.0 mU/L, 0.9–1.8 ng/dL, and 169–409 pg/dL, respectively.

**Conclusions:** We verified thyroid hormone RIs for Atellica by the indirect method for the first time. Our model proved reliable for selecting results of presumably healthy individuals from LIS data. Critical review of the RIs with local endocrinologists is essential.

## INTRODUCTION

Meaningful reference intervals (RIs) and interpretation of thyroid hormone laboratory test results, i.e., thyroid stimulating hormone (TSH), free thyroxine (FT4), and to some extent free triiodothyronine (FT3), are needed to correctly diagnose thyroid dysfunction. Traditionally, RIs have been established by collecting a minimum of 120 samples from qualified reference individuals and calculating the 2.5th and 97.5th percentiles (direct method) (1), but nowadays laboratory information systems (LIS) (2, 3) enable the use of large patient data sets. The datasets need to be filtered to represent healthy individuals, e.g., by including only one result per patient (4), by excluding all but the first or last result of the patient (3), or biochemically (5). For thyroid hormones, patients with a positive result for thyroid

[a]Department of Clinical Chemistry, University of Helsinki and Helsinki University Hospital, Helsinki, Finland; [b]Labquality Ltd, Helsinki, Finland; [c]Department of Endocrinology, University of Helsinki and Helsinki University Hospital, Helsinki, Finland.
*Address correspondence to this author at: Kumpulantie 15, 00520 Helsinki, Finland. Fax +358 9 8566 8280; e-mail jonna.pelanti@labquality.fi.
Received March 22, 2022; accepted June 30, 2022.
https://doi.org/10.1093/jalm/jfac067

## IMPACT STATEMENT

This article will benefit patients with thyroid disorders by helping laboratories provide thyroid function test results with accurate reference intervals. We have changed the immunoassay platform to Siemens Atellica and produced thyroid test reference intervals utilizing thousands of patient results. We show a way to process patient results in the LIS systems to reliably select results of presumably healthy individuals and prove Atellica thyroid hormone tests are suitable for clinical diagnostics. The reference intervals established in this study agree with previous studies using Siemens Advia Centaur or ACS:180™ system.

peroxidase antibodies should be excluded (6). In addition, outliers need to be removed and the Tukey nonparametric method is one of the most commonly used methods (7, 8).

Despite standardization of diagnostic assays, differences still exist, especially between immunoassays. Thus, when we changed immunochemistry platforms from Abbott® Architect to Siemens Atellica® Solution in October 2019, we needed to determine new RIs.

In this study, we present verification of the Atellica thyroid hormone tests and a robust model of patient data set derivation from LIS to achieve groups of presumably healthy individuals. The established 2.5th to 97.5th percentile Atellica RIs are in good agreement with the literature.

## MATERIALS AND METHODS

This study was performed at the Department of Clinical Chemistry (Diagnostic Center, Helsinki University Hospital), in the core laboratory and 3 satellite laboratories. Our thyroid tests are accredited according to ISO15189:2021 (9). This retrospective study is a quality investigation conducted according to the Declaration of Helsinki and approved by the Medical Research Committee of the Helsinki University Hospital (§15/2019). Based on local legislation, this Committee did not require evaluation by The Ethical Committee. The study did not receive grants from funding agencies in the public, commercial, or not-for-profit sectors.

### Patient Samples

Blood samples were collected into Vacuette lithium-heparin tubes (Greiner Bio-One) before 2 PM. Plasma was separated by centrifugation (2500*g*, 10 min at room temperature).

For method verification, we used plasma samples (TSH [$n = 78$], FT4 [$n = 56$], and FT3 [$n = 52$]) measured by the Abbott Architect assays and covering a wide concentration range. Second, sample sets from the endocrinology ($n = 90$) and pediatric ($n = 32$) outpatient clinics anonymously identified from LIS during a 1-week period were assayed either fresh or after storage at −20 °C.

### Immunoassays and Platforms

The tests were analyzed with Siemens Atellica® Solution IM1600 (Siemens Healthineers) and Architect i2000SR (Abbott Diagnostics) using Siemens Atellica reagent kits TSH 3-Ultra (TSH3-UL), FT4, and FT3, and Abbott Architect reagent kits TSH, FT4, and FT3, respectively (10–12).

### Assay Calibrations and Verification

Two level quality assurance samples (Liquichek Immunoassay Plus Controls, Bio-Rad Laboratories) were used in the daily routine and to estimate the within-run repeatability and within-laboratory total precision according to the Clinical and Laboratory Standards Institute EP10 protocol (13) by making

**Table 1. Thyroid hormone percentiles in various datasets.**

| TSH (mU/L) | *n* | 2.5% | MED | 97.5% |
|---|---|---|---|---|
| All results 15–59 | 166 167 | 0.1 | 1.7 | 8.1 |
| Excluded: hospital wards | 125 931 | 0.2 | 1.7 | 7.3 |
| Excluded: TPO antibody positive results | 118 067 | 0.2 | 1.7 | 6.4 |
| Single results | 94 854 | 0.4 | 1.7 | 5.0 |
| Single results, outliers removed by Tukey | 91 100 | 0.4 | 1.6 | 3.8 |
| Siemens Atellica kit insert (10) | n.d.[a] | 0.55 | n.d. | 4.78 |
| **FT4 (ng/dL)** | ***n*** | **2.5%** | **MED** | **97.5%** |
| All results 15–59 | 113 648 | 0.9 | 1.2 | 1.8 |
| Excluded: hospital wards | 84 197 | 0.9 | 1.2 | 1.7 |
| Excluded: TPO antibody positive results | 77 213 | 0.9 | 1.2 | 1.7 |
| Single result | 60 927 | 0.9 | 1.2 | 1.6 |
| Single result, outliers removed by Tukey | 59 751 | 0.9 | 1.2 | 1.6 |
| Siemens Atellica kit insert (11) | n.d. | 0.9 | n.d. | 1.8 |
| **FT3 (pg/dL)** | ***n*** | **2.5%** | **MED** | **97.5%** |
| All results 15–59 | 10 391 | 207.8 | 318.2 | 779.2 |
| Excluded: hospital wards | 3647 | 220.8 | 311.7 | 649.4 |
| Excluded: TPO antibody positive results | 3076 | 220.8 | 318.2 | 649.4 |
| Single result | 2246 | 220.8 | 311.7 | 526.0 |
| Single result, outliers removed by Tukey | 2110 | 220.8 | 311.7 | 415.6 |
| Siemens Atellica kit insert (12) | n.d. | 227.3 | n.d. | 422.1 |

[a]n.d., not determined.

duplicate measurements twice a day for 10 days. For quality management, traditional Westgard rules (14) were applied. The Atellica platform was calibrated whenever needed. The bias and accuracy were monitored daily, and we participated in several external quality assessment (EQA) schemes (Labquality Ltd, UKNEQAS, Bio-Rad Laboratories).

## Laboratory Software and Datasets

For the RI work-up, thyroid hormone results from year 2020 were retrieved using data application LabDW (Logex Suomi) connected to the LIS program MY+ (Mylab Ltd). LabDW data combines all available patients, samples, analysis, users, and logistic information.

The derivation of LIS datasets from raw data to the filtered data is presented in Table 1. Results from individuals <15 years and ≥60 years and tests ordered by hospital wards and hospital outpatient clinics were excluded, leaving results mainly from primary care and occupational health patients, as suggested by Farrel et al. (15). All patients with a positive result for TPO antibodies (since 2015) were removed and outliers handled according to Tukey (7, 8). With these criteria, 45% of TSH and FT4, and 75% of FT3 results were excluded. If a patient had several results in the dataset, we used only the latest result (denoted as single result dataset). Results outside the measuring range were excluded.

## Statistical Analysis

For assay verification, Validation Manager™ version 63.0–63.4 (Finbiosoft Ltd) was used. The LIS derived data was analyzed with Microsoft Excel

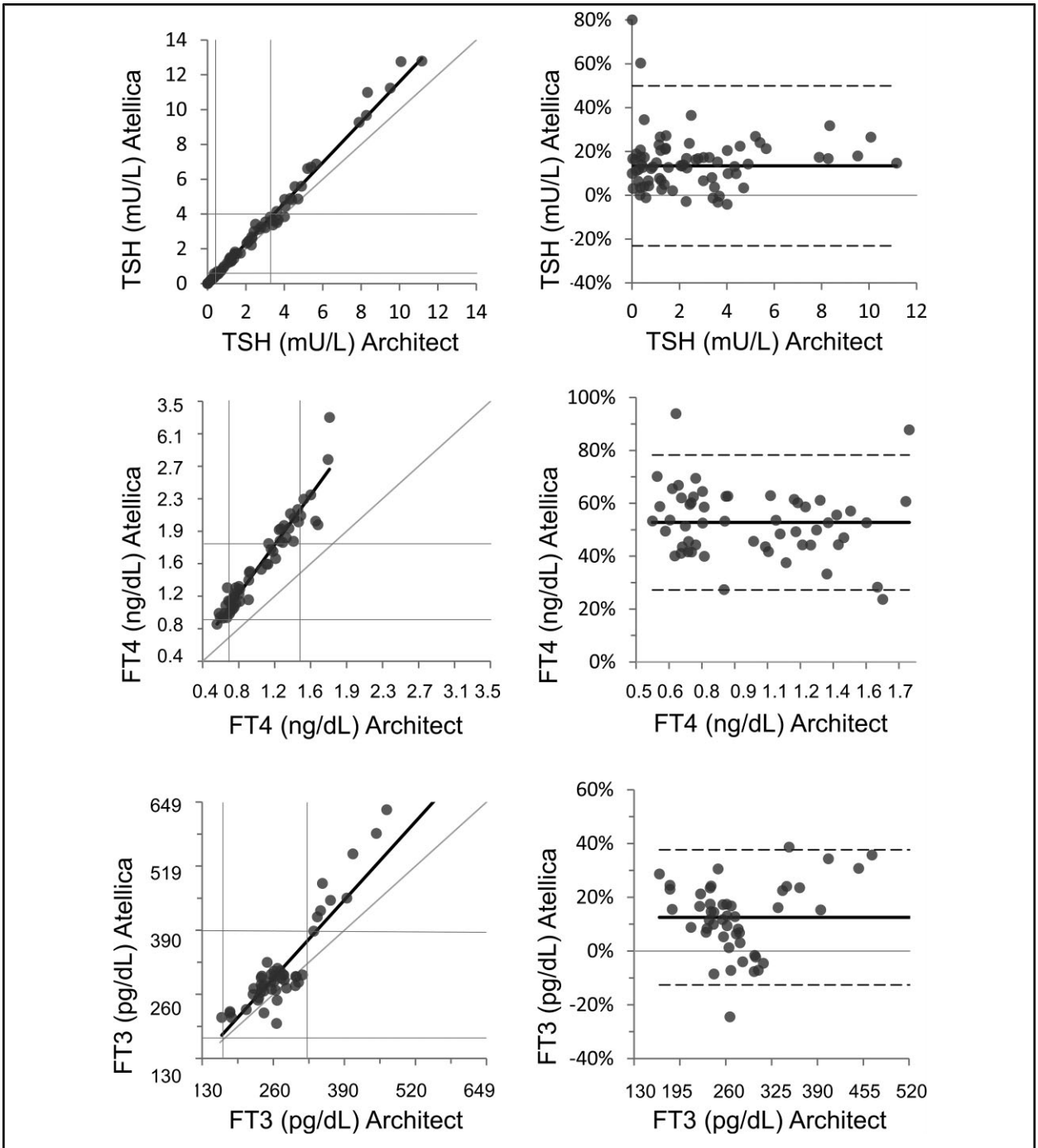


**Fig. 1. Verification of Atellica against Architect immunoassays for thyroid hormones using plasma samples. Passing–Bablok and Blandt–Altman analysis of Atellica and Architect TSH ($n = 74$), FT4 ($n = 55$), and FT3 ($n = 52$) results. Passing–Bablok analysis with regression line (—) (Passing–Bablok fit $y = -0.01025 + 1.161x$ for TSH, $y = 0.06587 + 1.5324x$ for FT4, $y = -0.3903 + 1.223x$ for FT3), LRL and URL of respective assay (horizontal and vertical lines –), and 95% limits of agreement (- - -).**

2010. Normality of distribution was analyzed with MedCalc® Statistical Software version 19.6 (MedCalc Software Ltd), the skewness was estimated according to Ichihara et al. (8) and was also visualized in histograms (data not shown).

## RESULTS

### Assay Verification and Preliminary RIs

For Atellica tests, within-run repeatability and within-laboratory total precision were good (CVs <4%, goal <5%). The correlation with respective Architect assays was excellent ($r = 0.959–0.995$) (Fig. 1) but with a significant positive bias (median 14%/53%/13% for TSH/FT4/FT3) (Fig. 1). The biases in both the endocrinology ($n = 90$) and pediatric ($n = 32$) cohorts were similar. The biases were approved as such since they were mostly expected according to existing EQA data (+10%/+ 24%/ +14% for TSH/FT4/FT3).

The Atellica methods were approved for routine clinical use with go-live in October 2019. Based on verification data, the observed biases between the platforms, previous Architect RIs (0.5–3.6 mU/L for TSH, 0.7–1.5 ng/dL for FT4, and 169– 325 pg/dL for FT3), and Atellica kit inserts (10–12), we set the following preliminary RIs for 15–59-year-old individuals: 0.5–4.0 mU/L for TSH, 0.9–1.8 ng/dL for FT4, and 169– 390 pg/dL for FT3. The Architect TSH RI was based on a previous publication (8), and we extrapolated the upper limits according to the positive biases noted in agreement with local endocrinologists.

### Comparison of Reference Intervals in the Datasets

To verify the RIs, we investigated several datasets of 15–59-year-old individuals (Table 1). The dataset distributions could not be normalized by log-transformations or Box–Cox transformations (data not shown). The positive skewness of TSH, FT4, and FT3 in the single result dataset is visualized in Supplemental Fig. 1. The exclusion of hospital ordered tests and TPO positive cases resulted in a decrease, particularly in the 97.5th percentile of TSH.

The established 95% RIs using the "single result" dataset were 0.39 (0.38–0.4)–3.8 (3.79–3.82) mU/L for TSH (median [95% CI]), 0.9 (0.9–0.9)–1.6 (1.6–1.6) ng/dL for FT4, and 220.8 (214.3–227.3)–415.6 (409.1–422.1) pg/dL for FT3. Concentrations of TSH, FT4, and FT3 were different in women and men ($P < 0.0001$). Comparison of 5-year categories of men and women separately revealed that they were different in all age groups except for 40–59 years for TSH and 50–54 years for FT3 (Supplemental Figs. 2–3, Supplemental Table 1).

## DISCUSSION

In this study, we verified Atellica assays and established RIs for plasma TSH, FT4, and FT3 based on laboratory patient data. This was essential because there were major biases between thyroid hormone concentrations with Abbott Architect and Atellica platforms. The importance of biochemical and other filtering of datasets is reflected by clinically significant decreases in the obtained upper percentiles, especially TSH. Our final dataset with a single result per patient most likely represents individuals as healthy as possible. However, this study is limited by the fact that we had only access to anonymous laboratory data. Therefore, we were not able to remove, e.g., patients diagnosed with thyroid disorder or those using thyroxine medication from our data.

At the time of the platform change, we revised the TSH RI based on previous RI and the identified positive bias (+14%). Our data indicate the upper reference limit (URL) to be 3.8 mU/L, whereas the URL 4.78 mIU/L suggested in the Atellica kit insert is evidently too high for our population. The American Association of Clinical Endocrinologists and the American Thyroid Association recommend that laboratories should use an URL of 4.12 mU/L for

TSH (16). However, this recommendation ignores the differences between different assays. In the literature, the estimated TSH RIs vary, which may be caused by differences in the studied populations.

For the Advia Centaur FT4 assay, previous studies indicate lower reference limit (LRL) values between 0.8 ng/dL and 0.9 ng/dL and URL values between 1.5 ng/dL and 1.6 ng/dL (Supplemental Table 2) (11, 17–21). However, in local clinical practice, there were concerns especially on both ends of our preliminary RIs. Therefore, we made minor changes and implemented the following RIs for TSH 0.5–4.0 mU/L, FT4 0.9–1.8 ng/dL (11–23 pmol/l), and FT3 168.8–409.1 pg/dL (2.6–6.3 pmol/l). The implemented FT4 RI corresponds to the 95% RI obtained from the unfiltered FT4 dataset and Atellica kit insert data (Table 1). These have, according to endocrinologists, worked well. The calculated (Table 1) RIs leave out 19%, 7%, and 17% of TSH, FT4, and FT3 results as compared to 19%, 3%, and 14% of the implemented RIs, respectively. That is, the implemented RIs take in more patients inside the RIs.

In conclusion, our patient-result-derived RIs are in line with expected values reported by the manufacturer and previous reports on thyroid hormone RIs on the Siemens Advia Centaur system (Supplemental Table 2). Challenges associated with immunochemistry platform change underscores the importance of ongoing close collaboration between the clinicians and the laboratory scientists.

## SUPPLEMENTAL MATERIAL

Supplemental material is available at *The Journal of Applied Laboratory Medicine* online.

**Nonstandard Abbreviations**: TSH, thyroid stimulating hormone; FT4, free thyroxine; FT3, free triiodothyronine; RIs, reference intervals; LIS, laboratory information system; EQA, external quality assessment; URL, upper reference limit; LRL, lower reference limit.

**Author Contributions:** *All authors confirmed they have contributed to the intellectual content of this paper and have met the following 4 requirements: (a) significant contributions to the conception and design, acquisition of data, or analysis and interpretation of data; (b) drafting or revising the article for intellectual content; (c) final approval of the published article; and (d) agreement to be accountable for all aspects of the article thus ensuring that questions related to the accuracy or integrity of any part of the article are appropriately investigated and resolved*.

**Authors' Disclosures or Potential Conflicts of Interest:** *Upon manuscript submission, all authors completed the author disclosure form. Disclosures and/or potential conflicts of interest:*

**Employment or Leadership:** M. Anttonen, Finnish Society of Clinical Chemistry, board member; O. Itkonen, Finnish Society of Clinical Chemistry, vice president; J. Pelanti, Finnish Society of Clinical Chemistry, secretary. **Consultant or Advisory Role:** C. Schalin-Jäntti, Recordati Rare Dis. **Stock Ownership:** None declared. **Honoraria:** M. Anttonen, Koulab; O. Itkonen, Ipsen, Koulab. **Research Funding:** None declared. **Expert Testimony:** None declared. **Patents:** None declared.

**Role of Sponsor:** No sponsor was declared.

## REFERENCES


1. CLSI. Defining, establishing, and verifying reference intervals in the clinical laboratory; approved guideline—third edition. CLSI document EP28-A3c. Wayne (PA): Clinical and Laboratory Standards Institute; 2010.
2. Lykkeboe S, Nielsen CG, Christensen PA. Indirect method for validating transference of reference intervals. Clin Chem Lab Med 2018;56:463–70.
3. Kouri T, Kairisto V, Virtanen A, Uusipaikka E, Rajamäki A, Finneman H, et al. Reference intervals developed from data for hospitalized patients: computerized method based on combination of laboratory and diagnostic data. Clin Chem 1994;40:2209–15.
4. Grossi E, Colombo R, Cavuto S, Franzini C. The REALAB project: a new method for the formulation of reference intervals based on current data. Clin Chem 2005;51:1232–40.
5. 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.
6. Schalin-Jäntti C, Tanner P, Välimäki MJ, Hämäläinen E. Serum TSH reference interval in healthy Finnish adults using the Abbott Architect 2000i Analyzer. Scand J Clin Lab Invest 2011;71:344–9.
7. Horn PS, Pesce AJ. Reference intervals: an update. Clin Chim Acta 2003;334:5–23.
8. Ichihara K, Boyd JC. An appraisal of statistical procedures used in derivation of reference intervals. Clin Chem Lab Med 2010;48:1537–51.
9. International Organization for Standardization. ISO 15189:2012. Medical laboratories: requirements for quality and competence. Geneva (Switzerland): International Organization for Standardization; 2012.
10. Siemens Healthineers. Atellica IM Thyroid Stimulating Hormone 3-Ultra (TSH-UL). https://doclib.siemens-healthineers.com/rest/v1/view?document-id=627681 (Accessed March 22, 2021).
11. Siemens Healthineers. Atellica IM Free Thyroxine (FT4). https://doclib.siemens-healthineers.com/rest/v1/view?document-id=757054 (Accessed March 22, 2021).
12. Siemens Healthineers. Atellica IM Free Triiodothyrine (FT3). https://doclib.siemens-healthineers.com/rest/v1/view?document-id=760047 (Accessed March 22, 2021).
13. CLSI. Measurement procedure comparison and bias estimation using patient samples; approved guideline—third edition. CLSI document EP09-A3. Wayne (PA): Clinical and Laboratory Standards Institute; 2013.
14. Westgard JO, Barry PL, Quam EF, Shrmeyer SS, Plaut D, Statland BE. Basic QC practices: training in statistical quality control for healthcare laboratories. 2nd Ed. Madison (WI): Westgard Quality Corp, 2002.
15. Farrell CL, Nguyen L. Indirect reference intervals: harnessing the power of stored laboratory data. Clin Biochem Rev 2019;40:99–111.
16. 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.
17. Barth JH, Spencer JD, Goodall SR, Luvai A. Reference intervals for thyroid hormones on Advia Centaur derived from three reference populations and a review of the literature. Ann Clin Biochem 2016;53:385–9.
18. Reix N, Massart C, d'Herbomez M, Gasser F, Heurtault B, Agin A. Thyroid-stimulating hormone and free thyroxine on the ADVIA Centaur immunoassay system: a multicenter assessment of analytical performance. Clin Biochem 2013;46:1305–8.
19. Wang P, Gao YJ, Cheng J, Kong GL, Wang Y, Wu XY, et al. Serum thyroid hormone reference intervals in the apparently healthy individuals of Zhengzhou area of China. Genet Mol Res 2014;13:7275–81.
20. Zou Y, Wang D, Cheng X, Ma C, Lin S, Hu Y, et al. Reference intervals for thyroid-associated hormones and the prevalence of thyroid diseases in the Chinese population. Ann Lab Med 2021;41:77–85.
21. De Grande LAC, Van Uytfanghe K, Reynders D, Das B, Faix JD, MacKenzie F, et al. Standardization of free thyroxine measurements allows the adoption of a more uniform reference interval. Clin Chem 2017;63:1642–52.