Factors influencing the reference interval of thyroid-stimulating hormone in healthy adults: A systematic review and meta-analysis

Dongyang Xing | Delong Liu | Ri Li | Qi Zhou | Jiancheng Xu

Abstract

Background: Many studies have reported that the thyroid-stimulating hormone (TSH) reference interval is susceptible to external factors, such as age, sex, race, region and iodine intake. However, no meta-analysis has comprehensively explored the effect of these factors on the TSH reference interval.

Methods: Articles published from January 1960 to January 2020 were searched in PubMed, Embase, Cochrane, Scopus, Medline English databases and CNKI, WanFang and CQVIP Chinese databases. In total, 19 studies were ultimately included. All data were analysed using Review Manager 5.3, STATA 16.0 software, GraphPad Prism 8.0 and Microsoft Excel 2010 to draw the TSH concentration curve.

Results: The TSH reference interval was significantly influenced by sex and age. The mean of TSH concentration in females was 0.27 mIU/L higher than that in males. Reference interval of TSH is divided into 20–59 years old and >60 years old age groups in males, and 20–39 years old and >40 years old age groups in females. Regardless of sex, TSH concentrations all increase with age. In iodine-deficient areas, TSH reference intervals were generally lower than those in iodine-sufficient or iodine-excessive areas. The TSH reference interval in Asia and North American countries was generally higher than that in most European countries. In the subgroup analyses of sample size, region and assay methods and manufacturers, the between-group differences were significant.

Conclusion: The TSH reference interval was significantly influenced by sex, age, iodine intake, sample size, region, and assay methods and manufacturers, but other factors should not be ignored. Therefore, it is necessary for each laboratory to validate an appropriate TSH reference interval based on local conditions.

Keywords

age, meta-analysis, reference interval, sex, thyroid-stimulating hormone
INTRODUCTION

Thyroid-stimulating hormone (TSH) is a glycoprotein hormone secreted by specific basophils in the pituitary gland, which is a specific indicator for evaluating individual thyroid function. Clinically common thyroid diseases include hyperthyroidism, hypothyroidism, thyroiditis and so on. With improvements in the understanding of thyroid diseases and health examinations, the incidence of thyroid diseases is increasing. Because many patients are asymptomatic or in a subclinical state, and some countries advocate a TSH priority strategy for the diagnosis of thyroid dysfunction, it is important and necessary to establish a reference interval. The reference interval is currently the most widely used tool for medical decision-making. It is usually defined as the reference value between 2.5% and 97.5%, which is the most important factor in health assessment, disease diagnosis, treatment and prognosis assessment.

Some reports define the lower limit of TSH as 0.2–0.4 mU/L and the upper limit as 4.0–5.0 mU/L, but a much lower upper limit of TSH has previously been suggested based on large population studies. The American Association of Clinical Endocrinologist recommended that each laboratory define their own reference intervals so that patients with thyroid-related diseases in different countries can receive screening and diagnosis in a timely manner.

Because TSH is an important indicator for evaluating thyroid function, so the preconditions that may influence the establishment of its reference interval should be considered comprehensively. Some studies based on multi-population regions have evaluated the TSH reference interval from healthy individuals according to the National Academy of Clinical Biochemistry (NACB) standards and have noted that the TSH reference interval was influenced by age, sex, race, iodine intake and thyroid autoantibodies. Through comprehensive analysis, this paper mainly discusses the effect of age, sex, iodine intake, sample size, region, assay methods and manufacturers and other factors such as race on the establishment of the TSH reference interval and explains why it is necessary for different countries to establish their own TSH reference intervals.

MATERIALS AND METHODS

Search Strategy and Database

Through PubMed, Medline, Embase, Scopus, and Cochrane English databases and WanFang, CNKI and CQVIP Chinese databases, the keywords “reference interval” or “reference value” combined with “TSH” or “thyroid stimulating hormone” were used to search all relevant Chinese and English literature from January 1960 to January 2020.

Exclusion and inclusion criteria

Based on the selection criteria of the selected articles and the recommendations of NABC, the final inclusion criteria were as follows: (a) more than 120 subjects; (b) thyroid peroxidase antibody (TPOAb) or thyroglobulin antibody (TGAb) negative; (c) no family or personal history of thyroid disease; (d) no goiter and no medical history influencing thyroid function (except use of oestrogens); (e) no obvious abnormalities in thyroid ultrasound; and (f) all articles grouped by age or sex.

The exclusion criteria were as follows: (a) related information of TSH reference interval could not be extracted; (b) Agency for Healthcare Research and Quality (AHRQ) score was >7; (c) studies related to pregnancy, children and the elderly; and (d) subjects with organic diseases of cardiovascular, lungs, kidneys and digestive system.

Quality assessment

The AHRQ was selected to assess the quality of the included studies. There are 11 scoring criteria. The selected articles were defined as three classes: high-quality articles (8–11 scores), median-quality articles (4–7 scores) and low-quality articles (1–3 scores). We included all studies and then performed a sensitivity analysis and then combine the NOS score to exclude low-quality articles. If there were arguments about the scores of the articles, the third author participated in the quality assessment.

Data extraction

Two reviewers (Dongyang Xing and Delong Liu) extracted the following data from all eligible studies independently: first author, publication year, country, sample size, age group and mean, the percentiles (2.5th and 97.5th) of TSH reference interval and assay methods and manufacturers. All researchers had no objection to the final results.

Statistical analysis

The random effects model was used to analyse the mean, standard deviation (SD) and sample size of the original data. The TSH concentration was log-transformed due to its abnormal distribution and the reference interval usually is represented by the mean, 2.5th percentile and 97.5th percentile. The results are presented as the weighted mean difference (WMD) with a 95% confidence interval (CI). For the heterogeneity test, I-squared value greater
than 50% was graded as high (the \( p \)-value < .05 was considered a significant difference). TSH differences between different age and sex groups were analysed using STATA 16.0 and Review Manager 5.3 software. At the same time, subgroup analyses of sample size, region, assay methods and manufacturers were also performed. Microsoft Excel 2010 and GraphPad Prism 8.0 were used to draw trend curves of the TSH concentration by sex and age. Microsoft Word 2010 was used to draw flow charts and basic information tables for selected articles.

3 | RESULTS

3.1 | Literature search and study characteristics

A total of 3285 articles were retrieved according to keywords, among which 481 were duplicate articles and 2553 articles were excluded by browsing the titles and abstracts. After reading the remaining 251 articles, 216 articles were excluded. Finally, the remaining 90 studies were included in this meta-analysis (14 published in English and 5 in Chinese). The flow chart is as follows (Figure 1). All articles were published between 1960 and 2020 and were mainly from Korea, Japan, China, India, France, Australia, Turkey, Iran and the Mediterranean. The NOS score of all studies was over 8 (basic information of selected articles is shown in Table 1). There were 9 articles on age comparison and 17 articles on sex comparison (7 articles discussed sex and age at the same time). The funnel plot (Figure 2) showed that each study was distributed at the top of the funnel plot, and the left and right sides were roughly symmetrical. So there was no significant publication bias.

3.2 | Age and sex distribution characteristics of TSH

There were nine studies grouped by age, and distributions of TSH were described by 10-year age group. In females, there were significant differences in TSH reference interval among the 30–39 and 40–49 years old age group (\( p < .05 \)). TSH reference interval on 40–49 years old was 0.24 mIU/L higher than that on 30–39 years old. The 95% CI of WMD was \([-0.44, 0.03]\) (Figure 3A). In males, TSH reference interval on 60–69 years old was 0.29 mIU/L higher than that on 50–59 years old (\( p < .05 \)). The 95% CI of WMD was \([-0.41, −0.16]\) (Figure 3B). The distribution curves of TSH mean values showed different trends in different age groups (Figure 4A, B). In males, the highest mean value of TSH appeared in 60–69 years old. In females, the lowest mean value of TSH appeared in 30–39 years old. The overall trend of TSH concentration increased with age in both males and females. The forest plot results showed a significant difference in the TSH concentration between males and females (\( p < .05 \)). The mean of TSH concentration in females was 0.27 mIU/L higher than that in males, in which heterogeneity was 98%, and the 95% CI of WMD was \([-0.34, −0.19]\) (Figure 5). The overall trend curve drawing by the mean, the lower limits and upper limits of TSH reference

![Flow diagram of included studies](image-url)
intervals in males and females (Figure 6) showed that the TSH concentration in females was generally higher than that in males.

### 3.3 Sample size, region and assay methods and manufacturers distribution characteristics of TSH

Subgroup analyses were performed respectively on the factors that may affect the TSH reference interval, such as sample size, region and assay methods and manufacturers. First of all, we defined the number of sample size in male or female as 1000 for the standard, which was determined according to the basic information of the selected articles. The number of large sample size was higher than 1000 in both male and female, while the number of small sample size ($n = 9$) was lower than 1000 in male or female. In the subgroup analysis of sample size, there was a significant between-group difference ($p < .05$). The total WMD was $0.26 \text{ mIU/L} [95\% \text{ CI } (-0.33,-0.18)]$ (Figure 7). In the forest plot, we found no

### Table 1 Basic information of included studies

| First author | Year | Country       | Sample size | Age  | Detection instruments | Detection methods                   | Quality assessment |
|--------------|------|---------------|-------------|------|-----------------------|-------------------------------------|--------------------|
| Park         | 2018 | Korea         | 5987        | >18  | Roche Diagnostics, Mannheim | Electrochemiluminescence immunooassay | 10                 |
| Yoshihara    | 2011 | Japan         | 1388        | >20  | Roche Diagnostics GmbH  | Electrochemiluminescence immunooassays | 9                  |
| Sasso        | 2019 | Mediterranean | 22,602      | 15–105 | Roche Cobas e801 analyser | Electrochemiluminescence immunooassay | 8                  |
| Gao          | 2014 | China         | 4820        | 15–76 | Roche Cobas e601 Automatic immune analyzer | Electrochemiluminescence immunooassay | 8                  |
| Wang         | 2015 | Urumqi        | 897         | 18–84 | Roche Cobas e601 Automatic immune analyzer | Electrochemiluminescence immunooassay | 8                  |
| Raman        | 2013 | Indian        | 1916        | >18  | Roche Cobas Elecsys 1010 analyser | Electrochemiluminescence immunooassay | 11                 |
| Amouzeger    | 2013 | Tehran, Iran  | 5704        | >20  | Roche Cobas e411 analyser | Electrochemiluminescence immunooassay | 8                  |
| Tame.C       | 2010 | Turkey        | 55,318      | >20  | Roche Elecsys 2010 analysers | Electrochemiluminescence immunooassay | 8                  |
| Frauk        | 2014 | Turkey        | 408         | >18  | Roche Diagnostics GmbH, D-68298 | Electrochemiluminescence immunooassay | 8                  |
| Nerrela      | 2013 | Australia     | 152,261     | >18  | Siemens ADVIA Centaur analyser | Chemiluminescence immunooassay | 8                  |
| Qiu          | 2018 | China         | 106,335     | >18  | Siemens ADVIA Centaur | Chemiluminescence immunooassay | 8                  |
| P. Wang      | 2014 | China         | 211         | 23–77 | Siemens ADVIA Centaur XP analyser | Chemiluminescence immunooassay | 8                  |
| Jang         | 2008 | Korea         | 1591        | 18–65 | Siemens IMMULITE 2000 | Chemiluminescence immunooassay | 9                  |
| Raverot      | 2020 | French        | 295,775     | >20  | Architect i2000 immunoassay analyser | Chemiluminescence | 9                  |
| Li           | 2011 | China         | 5639        | 12–85 | Diagnostic Products Corporation (USA) | Chemiluminescence immunooassay | 8                  |
| Guan         | 2008 | China         | 3761        | >13  | Not mentioned | Not mentioned | 10                 |
| Song         | 2012 | China         | 390         | 18–65 | Beckman Coulter Unicel DXI 800 | Chemiluminescence | 8                  |
| Neda         | 2014 | Serbia        | 22,860      | >18  | Abbott Architect CI8200 analyser | Chemiluminescence immunooassay | 8                  |
| Lin          | 2007 | China         | 5070        | 21–80 | Liaison | Chemiluminescence immunooassay | 9                  |
difference in TSH reference intervals in small sample size \((n = 9; p > .05)\), while there were significant differences in large sample size \((n = 8; p < .05)\). The heterogeneity of the small sample size subgroup was 94%, which was lower than that of large sample size. According to the selected articles, the region is mainly divided into China \((n = 8)\) and other countries \((n = 9)\). In different regions, TSH reference interval was significantly different \((p < .05)\), and the total WMD was \(-0.26 \text{ mIU/L} [95\% \text{ CI} (-0.33, -0.18)]\) (Figure 8). The heterogeneity of other regions \(I^2 = 98\%\) is higher than that of China \(I^2 = 78\%\). According to the basic information of the selected articles, the assay manufacturers are mainly divided into Roche \((n = 10)\) and Siemens \((n = 2)\). There was a significant between-group difference in TSH reference interval \((p < .05)\), and the total WMD was \(-0.30 \text{ mIU/L} [95\% \text{ CI} (-0.37, -0.24)]\) (Figure 9). The studies assayed by Siemens
have no difference \((p > 0.05)\), while there was statistical difference in the studies assayed by Roche \((p > 0.05)\).

**4 | DISCUSSION**

Thyroid disease is a common endocrine disease. TSH plays an important role in the diagnosis, treatment, prevention and control of thyroid disease. Asvold et al.\(^3^0\) mentioned that high TSH within the reference interval may predict increased risk of future hypothyroidism. Conversely, in the early stage of a process that culminates in hyperthyroidism, TSH could be slightly reduced, and therefore, low TSH within the reference interval could predict future hyperthyroidism.\(^3^1\) Therefore, to improve the sensitivity and specificity of the test, the TSH reference interval corresponding to the local population should be established firstly.\(^3^2\) At present, many studies have mentioned that the establishment of the TSH reference interval is affected by some factors. Age and sex are important independent predictors that cannot be ignored.\(^3^3\)–\(^3^5\) Iodine intake and race can lead to significant changes in the TSH reference interval.\(^1^2\),\(^3^6\),\(^3^7\) Sample size, region and assay methods and manufacturers should be considered as important objective factors, too. No researches have been found in published studies on the systematic analysis of factors affecting the TSH reference interval in healthy people. This paper is the first meta-analysis to comprehensively summarize these factors. Significant differences in sex and age were found in TSH through a sex analysis of 17 articles and an age analysis of 9 articles. Iodine intake, sample size, region, assay methods and manufacturers and race also had important effects on the establishment of the TSH reference interval. Therefore, this article will focus on the effects of the above factors.

**4.1 | Age**

National Health and Nutrition Survey III (NHANES III) suggested that the serum TSH concentration increased with age in adults without thyroid disease.\(^3^8\) Furthermore, Jonklaas and Razvi mentioned that the age-related increase in serum TSH was similar in both genders.\(^3^9\) Surks et al.\(^4^0\) reported that the 97.5 percentile of TSH increased from 3.56 mIU/L in the 20–29 years old group to 7.49 mIU/L in the over 80 years old group. In this study, the TSH concentration also increased with age. Some researchers have suggested that the progressive increase in the TSH concentration with ageing could be due to an enhancement in the prevalence of acquired autoimmune thyroid disease and an increase in anti-thyroid antibodies.\(^3^8\),\(^4^0\) Several changes in thyroid may also contribute to the increase in serum TSH with age, such as decreased sensitivity of TSH to the negative feedback of the thyroid hormone,\(^4^1\) decreased biological activity of age-related TSH\(^4^2\) and abnormal free thyroxine (FT4), and the TSH feedback loop may lead to an increased TSH concentration.\(^4^3\) Other studies have

![FIGURE 4 A and B, respectively, showed the changes of the overall mean values of TSH with age in different age groups in males and females](image-url)
confirmed that the increase in TSH with age is due to the normal compensatory phenomenon in the elderly. However, the TSH concentration decreased with age in a few iodine-deficient regions such as Italy and Germany. Van de Ven et al reported that the TSH concentration in iodine-deficient areas was inversely proportional to age. In conclusion, the variation trend between age and TSH concentration is not fixed, and the TSH reference interval in different countries will be distinct due to the influence of living habits and living environment.

The results of this study and most other studies suggest that the TSH concentration increases with age.

### 4.2 | Sex

In this meta-analysis, by comparing the mean values of TSH between males and females, the variation curves of the TSH concentration in diverse regions were obtained. The results showed a higher TSH concentration in females than in males in most regions. Oestrogen is an important factor affecting the TSH concentration. Low oestrogen may cause hypothyroidism and then lead to an increased TSH concentration. Postmenopausal women were the typical group, and the TSH concentration increased significantly. This may be one of the reasons why TSH is generally higher in females than in males. Jonklaas and Razvi also supposed that women tended to have higher TSH concentrations than men because this increase was accounted for by positive thyroid peroxidase antibody status. Although this study cannot represent the relationship of the TSH concentration between males and females in all regions, it is a general trend that the TSH concentration is higher in females than in males.

### 4.3 | Iodine intake and region

Iodine intake was an important factor that could partly explain the variation in the TSH reference interval in different studies.

---

**FIGURE 5** Showed that TSH reference interval in female was higher than in male. The WMD was \(-0.27 \text{ mIU/L} [-0.34, -0.19]\).  

**FIGURE 6** Showed that the relationship between male and female TSH reference intervals in each study.
At the same time, the difference in TSH reference interval in different regions is also closely related to iodine intake. Park et al. found that the TSH reference interval in Korea was significantly higher than that in Western countries. In their study, the mean value and upper limit of the TSH reference interval were 2.16 and 7.03 mIU/L, respectively, while the NHANES III of the United States reported values of 1.40 and 4.12 mIU/L, respectively. These differences may be explained by the state of iodine intake between Western and Korean countries. Other studies reported that the upper and lower limits of the TSH reference interval in regions such as North America and East Asia, where iodine is in abundant supply, are often higher than those in iodine-deficient regions such as Europe. Not only different levels of iodine intake in other countries but also the levels of iodine intake vary from mild to excessive in different regions of China. Guan et al. suggested that it was necessary to consider iodine intake when establishing the TSH reference interval. Their study was conducted in Panshan, Zhangwu and Huanghua, regions with mildly deficient, more than adequate and excessive iodine intake, respectively, and the mean levels of TSH in Panshan, Zhangwu and Huanghua were 1.15, 1.28 and 1.93 mIU/L, respectively. Therefore, iodine deficiency or excess will affect the establishment of the TSH reference interval. It has been reported that salt in regions in which iodine intake is deficient should be supplemented with iodine according to the degree of iodine deficiency. Otherwise, the incidence of hyperthyroidism increases among persons with iodine deficiency who suddenly increase their iodine intake.

### Sample size and assay methods and manufacturers

According to the features of the researches included in this meta-analysis, the sample size and assay methods and manufacturers were analysed by subgroup analysis. In this study, the differences between articles with large sample size were more obvious. Chen et al. proposed that their results were inconsistent with those of Wang et al. The sample size was 211 in the study of Wang and 7693 of Chen. Therefore, the difference may be related to the sample size. It can be seen that the sample size is an important factor influencing the establishment of the TSH reference interval.

#### FIGURE 7

Showed the subgroup analysis of sample size

| Study or Subgroup | male Mean | SD | Total | female Mean | SD | Total | Weight | Mean Difference (IV, Random, 95% CI) | Mean Difference (IV, Random, 95% CI) |
|-------------------|----------|----|-------|-------------|----|-------|--------|-----------------------------------|-----------------------------------|
| **1.1.1 large sample size** |          |    |       |             |    |       |        |                                   |                                   |
| Gao 2014          | 2.24     | 1.02 | 2540  | 2.36        | 1.08 | 2280  |        | -0.12 [-0.18, -0.06]               |                                   |
| Li 2011           | 2.96     | 1.28 | 1036  | 3.01        | 1.28 | 1452  |        | -0.05 [-0.15, 0.05]                |                                   |
| Lin 2007          | 3.12     | 1.34 | 2549  | 3.3         | 1.43 | 2521  |        | -0.18 [-0.26, -0.10]               |                                   |
| Neda 2014         | 2.46     | 0.79 | 11440 | 2.39        | 0.92 | 11420 |        | 0.07 [0.05, 0.09]                  |                                   |
| Park 2018         | 3.77     | 1.63 | 3493  | 4.28        | 1.99 | 3568  |        | -0.51 [-0.60, -0.43]               |                                   |
| Qin 2018          | 2.67     | 1.07 | 51044 | 2.88        | 1.26 | 53851 |        | 0.21 [-0.22, -0.20]                |                                   |
| Sasso 2019        | 1.65     | 0.69 | 7805  | 1.92        | 0.83 | 12099 |        | -0.27 [-0.29, -0.25]               |                                   |
| Tamer.C 2010      | 2.1      | 0.85 | 11779 | 2.24        | 0.92 | 43539 |        | -0.14 [-0.16, -0.12]               |                                   |
| **Total (95% CI)**| 91686    |    | 130730| 57.4%       |     | -0.17 [-0.27, -0.08]               |                                   |

Heterogeneity: Tau² = 0.02; Chi² = 638.78, df = 7 (P < 0.00001); I² = 99%

Test for overall effect: Z = 3.71 (P = 0.0002)

| **1.1.2 small sample size** |          |    |       |             |    |       |        |                                   |                                   |
| Amouzegar 2013      | 1.47     | 0.9  | 953   | 2.01       | 1.4 | 1246  |        | -0.54 [-0.64, -0.44]               |                                   |
| Frauz 2014          | 1.5      | 0.8  | 288   | 1.8        | 1.1 | 140   |        | -0.30 [-0.51, -0.09]               |                                   |
| Guan 2008           | 1.32     | 1.26 | 630   | 1.47       | 1.85 | 1670  |        | -0.15 [-0.28, -0.02]               |                                   |
| Jiang 2008          | 1.28     | 0.84 | 911   | 1.49       | 2.08 | 680   |        | -0.21 [-0.38, -0.04]               |                                   |
| P.Wang 2014         | 3.04     | 1.94 | 94    | 3.63       | 2.69 | 117   |        | -0.59 [-1.22, 0.04]                |                                   |
| Ramamo 2013         | 4.41     | 1.89 | 916   | 5.28       | 2.34 | 1000  |        | -0.87 [-1.06, -0.68]               |                                   |
| Song 2012           | 1.92     | 1.12 | 221   | 2.37       | 1.86 | 169   |        | -0.45 [-0.77, -0.13]               |                                   |
| WangXin 2015        | 4.05     | 1.96 | 198   | 4.76       | 1.32 | 253   |        | -0.71 [-1.03, -0.39]               |                                   |
| Yoshiiara 2011      | 1.48     | 0.55 | 198   | 1.48       | 0.52 | 1190  |        | 0.00 [0.08, 0.08]                  |                                   |
| **Total (95% CI)**  | 4389     |    | 4645  | 42.6%      |     | -0.40 [-0.62, -0.19]               |                                   |

Heterogeneity: Tau² = 0.09; Chi² = 123.20, df = 8 (P < 0.00001); I² = 94%

Test for overall effect: Z = 3.71 (P = 0.0002)

### 4.4 Sample size and assay methods and manufacturers
### FIGURE 8
Showed the subgroup analysis of region

| Study or Subgroup | mean | SD  | Total | mean | SD  | Total | Weight | Mean Difference | Mean Difference |
|-------------------|------|-----|-------|------|-----|-------|--------|----------------|----------------|
|                   | male |     |       | female |     |       |        | IV, Random, 95% CI | IV, Random, 95% CI |
| 1.2.1 China       |      |     |       |       |     |       |        |                |                |
| Gao 2014          | 2.24 | 1.02| 2540  | 2.36 | 1.08| 2280  | 7.2%   | -0.12 [-0.18, -0.06] |                |
| Guan 2008         | 1.32 | 1.26| 630   | 1.47 | 1.85| 1670  | 6.1%   | -0.15 [-0.26, -0.02] |                |
| Li 2011           | 2.96 | 1.28| 1036  | 3.01 | 1.28| 1452  | 6.6%   | -0.05 [-0.15, 0.05]  |                |
| Lin 2007          | 3.12 | 1.34| 2549  | 3.3  | 1.43| 2521  | 7.0%   | -0.18 [-0.26, -0.10] |                |
| P.Wang 2014       | 3.04 | 1.94| 94    | 3.63 | 2.69| 117   | 1.2%   | -0.69 [-1.22, 0.04]  |                |
| Qiu 2018          | 2.67 | 1.07| 51044 | 2.88 | 1.26| 53851 | 7.5%   | -0.21 [-0.22, -0.20] |                |
| Song 2012         | 1.52 | 1.12| 221   | 2.37 | 1.86| 169   | 3.2%   | -0.45 [-0.77, -0.13] |                |
| Wang/Kin 2015     | 4.09 | 1.96| 198   | 4.76 | 1.32| 253   | 3.2%   | -0.71 [-1.03, -0.39] |                |
| Subtotal (95% CI) | 58312|     |       | 62313|     |       | 41.9%  | -0.19 [-0.26, -0.12] |                |

Heterogeneity: Tau² = 0.01; Chi² = 31.60, df = 7 (P < 0.00011); I² = 78%
Test for overall effect: Z = 5.23 (P < 0.00001)

### FIGURE 9
Showed the subgroup analysis of assay methods and manufacturers

| Study or Subgroup | mean | SD  | Total | mean | SD  | Total | Weight | Mean Difference | Mean Difference |
|-------------------|------|-----|-------|------|-----|-------|--------|----------------|----------------|
|                   | male |     |       | female |     |       |        | IV, Random, 95% CI | IV, Random, 95% CI |
| 1.3.1 Roche       |      |     |       |       |     |       |        |                |                |
| Amouzegar 2013    | 1.47 | 0.9 | 953   | 2.01 | 1.4 | 1246  | 6.7%   | -0.54 [-0.64, -0.44] |                |
| Frak 2014         | 1.5  | 0.8 | 268   | 1.8  | 1.1 | 140   | 4.8%   | -0.30 [-0.51, -0.09] |                |
| Jang 2008         | 1.28 | 0.84| 911   | 1.49 | 2.08| 680   | 5.5%   | -0.21 [-0.38, -0.04] |                |
| Neda 2014         | 2.46 | 0.79| 11440 | 2.39 | 0.92| 11420 | 7.5%   | 0.07 [0.05, 0.09]  |                |
| Park 2016         | 3.77 | 1.63| 3493  | 4.28 | 1.99| 3568  | 6.8%   | -0.51 [-0.58, -0.43] |                |
| Raman 2013        | 4.41 | 1.89| 916   | 5.28 | 2.34| 1000  | 5.1%   | -0.87 [-1.06, -0.68] |                |
| Sasso 2019        | 1.65 | 0.69| 7905  | 1.92 | 0.83| 12099 | 7.5%   | -0.27 [-0.28, -0.25] |                |
| Tamer.C 2010      | 2.1  | 0.85| 11779 | 2.24 | 0.92| 43539 | 7.5%   | -0.14 [-0.16, -0.12] |                |
| Yoshinara 2011    | 1.48 | 0.55| 198   | 1.48 | 0.52| 1190  | 6.9%   | 0.00 [-0.08, 0.08]  |                |
| Subtotal (95% CI) | 37763|     |       | 74882|     |       | 58.1%  | -0.29 [-0.42, -0.17] |                |

Heterogeneity: Tau² = 0.03; Chi² = 695.59, df = 8 (P < 0.000011); I² = 99%
Test for overall effect: Z = 4.65 (P < 0.00001)

Total (95% CI) 96075 137195 100.0% -0.26 [-0.33, -0.18]
Heterogeneity: Tau² = 0.02; Chi² = 784.86, df = 16 (P < 0.000011); I² = 98%
Test for overall effect: Z = 6.80 (P < 0.00001)
Test for subarous differences: Chi² = 2.11, df = 1 (P = 0.15); I² = 52.6%
interval was mostly established by CLIA or ECLI. The main principle of CLIA is to use chemiluminescence reagents to mark antigens or antibodies. After the labelled antigens and antibodies go through a series of immune reactions and physicochemical steps with the determinand, the content of the determinand is finally expressed in the form of luminescence intensity. ECLI is a process of chemiluminescence caused by electrochemical reactions, and its principle is similar to that of CLIA. At present, there are various assay manufacturers, including the Siemens ADVIA Centaur Analyzer, Abbot Architect Analyzer, Roche Cobas Analyzer, and Beckman Coulter. In this study, Roche and Siemens were the main assay manufacturers. Due to the diverse range of assay manufacturers, even when the same assay method is used, different reagents for marking antigens or antibodies during the reaction may cause differences in the TSH reference interval. Kalaria et al. also proposed that different assay manufacturers had an impact on the diagnosis of thyroid diseases. Therefore, the establishment of the TSH reference interval can be influenced by different assay methods and manufacturers.

4.5 | Other factors

In addition to the above factors, race as a possible factor should not be ignored. Based on the limited information provided by the original paper on this side, we provided only a brief explanation. It is worth noting that the results of Chen et al. showed that the lower and upper limits of the TSH reference interval in China were higher than that of Western countries. Data from several recent studies showed that the median and upper limit of the TSH reference interval in both African Americans and non-Hispanic Americans were lower than those in white Americans. Jonklaas and Razvi also suggested that African Americans have lower serum TSH concentrations than white people. One possible explanation is racial differences. NHANES III also proposed that the median, upper and lower limits of the reference interval in African America were lower than those in Caucasian. Boucai et al. confirmed that the decrease median, upper and lower limits of the TSH reference interval in the African America were due to the shift of the distributed population to a lower TSH reference value. Studies on genetic and environmental impacts have suggested that the negative feedback pathway of TSH is also affected by genetics. Another report on gene polymorphisms in the thyroid hormone pathway also suggested that the difference in the TSH reference interval may be related to differences between African America and Caucasian. Thus, the importance of race in the establishment of the TSH reference interval has been demonstrated.

5 | LIMITATIONS

According to the 19 selected articles, we made specific explanations on the possible influencing factors. However, due to the limited data and information in original articles, we mainly summarized the following deficiencies: (i) The number of studies was limited, which did not cover the TSH reference interval in healthy people in all countries. (ii) We did not perform a subgroup analysis of iodine intake and race. Due to the basic information in the original article was limit, the number of articles for subgroup analysis was insufficient.

6 | CONCLUSION

This study comprehensively analysed the factors affecting the establishment of the TSH reference interval, including age, sex, iodine intake, sample size, region, assay methods and manufacturers and race. We suggest that TSH is significantly changed by age and sex, and our conclusions are as follows: the TSH reference interval in females was generally higher than that in males and TSH concentration increased with age in both males and females. The establishment of the TSH reference interval was also significantly influenced by sample size, region, assay methods and manufacturers. In the sample size, the effect of large sample size on TSH reference interval is more significant. In addition, the assay and reagents used in each laboratory were diverse manufacturers, and the test results will vary to different degrees. In different regions, the effect of other countries on TSH reference interval is more significant. Therefore, the reference interval provided by the manufacturer should be used selectively, and an appropriate reference interval should be established separately for special groups. As a result, to avoid incorrectly identifying TSH concentrations, we should take the above factors into full consideration when the reference interval was established. At the same time, this meta-analysis can provide some guidance for the clinical diagnosis and treatment of thyroid-related diseases.

ACKNOWLEDGEMENTS

The authors would like to acknowledge all authors of the original studies.

CONFLICT OF INTEREST

Authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

Dongyang Xing and Delong Liu were involved in searching articles, collecting data and drafting of the manuscript; Qi Zhou and Ri Li were involved in interpretation of data and performing statistical analysis; Jiancheng Xu involved in study conception and design and approval the final version of the manuscript.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

ORCID

Jiancheng Xu https://orcid.org/0000-0001-8796-271X
REFERENCES

1. Chan AO, Ip YP, Shek CC. The reference interval of thyroid-stimulating hormone in Hong Kong Chinese. J Clin Pathol. 2011;64(5):433-436.
2. Baskin HJ, Cobin RH, Duick DS, et al. American Association of Clinical Endocrinologists medical guidelines for clinical practice for the evaluation and treatment of hyperthyroidism and hypothyroidism. Endocr Pract. 2002;8(6):457-469.
3. Odhiambo C, Oyaro B, Odipo R, et al. Evaluation of locally established reference intervals for hematometry and biochemistry parameters in Western Kenya. PLoS One. 2015;10(4):e0123140.
4. Stephens AP. The Endocrine Society: current issues in thyroid disease management. Endocr News. 2004;29:23-26.
5. Stathatos N, Wartofsky L. Managing subclinical hypothyroidism in women. Women Health Primary Care. 2005;2:239-246.
6. Bervini S, Trelle S, Kopp PA, Stettler C, Trepp R. Incidence of iodine-induced Hyperthyroidism after administration of iodinated contrast during radiographic procedures: a systematic review and meta-analysis of the literature. Thyroid. 2020. https://doi.org/10.1089/thy.2020.0459
7. Guan H, Shan Z, Teng X, et al. Influence of iodine on the reference interval of TSH and the optimal interval of TSH: results of a follow-up study in areas with different iodine intakes. Clin Endocrinol (Oxf). 2008;69:136-141.
8. Marwahaa RK, Tandonc N, Ganiee MA, et al. Reference range of thyroid function (FT3, FT4 and TSH) among Indian adults. Clin Biochem. 2013;46:341-345.
9. Park SY, Kim HI, Oh HK, et al. Age- and gender-specific reference intervals of TSH and free T4 in an iodine-replete area: data from Korean National Health and Nutrition Examination Survey IV (2013-2015). PLoS One. 2018;13(2):e0190738.
10. Chaker L, Korevaar TI, Medicci M, et al. Thyroid function characteristics and determinants: the rotterdam study. Thyroid. 2016;26(9):1195-1204.
11. Sun X, Shan Z, Teng W. Effects of increased iodine intake on thyroid disorders. Endocrinol Metab (Seoul). 2014;29:240-247.
12. Kratzsch J, Fiedler GM, Leichtle A, et al. New reference intervals for thyrotropin and thyroid hormones based on National Academy of Clinical Biochemistry criteria and regular ultrasonography of the thyroid. Clin Chem. 2005;51:1480-1486.
13. Shoemaker-Hunt S, Hall K, Hoffman L. Advancing patient safety: reviews from the agency for healthcare research and quality’s making healthcare safer III report. J Patient Saf. 2020;163(Suppl 1):S1-S2.
14. Wang P, Gao YJ, Cheng J, et al. Serum thyroid hormone reference intervals in the apparently healthy individuals of Zhengzhou area of China. Genet Mol Res. 2014;13(3):7275-7281.
15. Inal TC, Serteser M, Coskun A, Ozpinar A, Unsal I. Indirect reference intervals estimated from hospitalized population for thyrotropin and free thyroxine. Croat Med J. 2010;51(2):124-130.
16. Kutluturk F, Yildirim B, Ozturk B, et al. The reference intervals of thyroid stimulating hormone in healthy individuals with normal levels of serum free thyroxine and without sonographic pathologies. Endocr Res. 2014;39(2):57-61.
17. Jang YY, Kim CY, Hwang TY, Kim KD, Lee C-H. Reference interval of serum thyroid hormones in healthy Korean adults. J Prev Med Public Health. 2008;41(2):128-134.
18. Lo Sasso B, Vidali M, Scazzone C, Agnello L, Ciaccio M. Reference interval by the indirect approach of serum thyrotropin (TSH) in a Mediterranean adult population and the association with age and gender. Clin Chem Lab Med. 2019;57(10):1587-1594.
19. Li C, Guan H, Teng X, et al. An epidemiological study of the serum thyrotropin reference range and factors that influence serum thyrotropin levels in iodine sufficient areas of China. Endocr J. 2011;58(11):995-1002.
20. Yoshihara A, Noh YJ, Ohye H, et al. Reference limits for serum thyrotropin in a Japanese population. Endocr J. 2011;58(7):585-588.
21. Amouzegar A, Delshad H, Mehran L, et al. Reference limit of thyrotropin (TSH) and free thyroxine (FT4) in thyroperoxidase positive and negative subjects: a population based study. J Endocrinol Invest. 2013;36(11):950-954.
22. Hai-feng GAO, Jun-li GE, Wei Z. Investigation on biology reference intervals for five parameters of thyroid gland function in Baoji Area. J Mod Lab Med. 2014;6:121-123.(in Chinese).
23. Xinling W, Chen Xiang X, Ziqi GY. Research on TSH hormone reference interval of normal population in Urumqi Area. Label Immunossays Clin Med. 2015;22(9):839-841,878. (in Chinese).
24. Ling Q, Wang Danhen X, Tao CX, Qi S. Influence of gender, age and season on thyroid hormone reference interval. Natl Med J China. 2018;20:1582-1587.(in Chinese).
25. Mlinikovic N, Ignjatovic S, Zarkovic M, et al. Indirect estimation of age-related reference limits of thyroid parameters: a cross-sectional study of outpatients’ results. Scand J Clin Lab Invest. 2014;74(5):378-384.
26. Ming L, Pingyan J, Lifang X, Lizhen X. Investigation on the reference intervals for 5 items for thyroid function in adults in Yunnan Yuxi. Lab Med. 2017;32(2):114-117.(in Chinese).
27. Xin S, Xiang-yi L, Jiang-ping W, Xiao-hua W, Yong-jun G. An initial investigation of the serum thyroid hormone reference range in the apparently healthy individuals of Beijing and Shanghai area of China. Chinese J Lab Med. 2012;2:156-159 (in Chinese).
28. Raverot V, Bonjour J, Abellon J, et al. Age- and sex-specific TSH upper-limit reference intervals in the general French population: there is a need to adjust our actual practices. J Clin Med. 2020;9(3):792.
29. Hadlow NC, Rothacker KM, Wardrop R, Brown SJ, Lim EM, Walsh JP. The relationship between TSH and free T4 in a large population is complex, non-linear and differs by age and gender. J Clin Endocrin Metab. 2013;98(7):2936-2943.
30. Ásvold BO, Vatten LJ, Midthjell K, Bjørø T. Serum TSH within the reference range as a predictor of future hypothyroidism and hyperthyroidism: 11-year follow-up of the hunt study in Norway. J Clin Endocrinol Metab. 2012;97(1):93-99.
31. Teng W, Shan Z, Teng X, et al. Effect of iodine intake on thyroid diseases in China. N Engl J Med. 2006;354(26):2783-2793.
32. Baloch Z, Carayon P, Conte-Devolx B, et al. Laboratory medicine practice guidelines. Laboratory support for the diagnosis and monitoring of thyroid disease. Thyroid. 2003;13:1-126.
33. Boucai L, Hollowell JG, Surks MI. An approach for development of age- and sex-specific thyrotropin reference limits. Thyroid. 2011;21:5-11.
34. Boucai L, Surks MI. Reference limits of serum TSH and free T4 are signifi cantly influenced by race and age in an urban outpatient medical practice. Clin Endocrinol (Oxf). 2009;70:788-793.
35. Surks MI, Boucai L. Age- and race-based serum thyrotropin reference limits. J Clin Endocrinol Metab. 2010;95:496-502.
36. Spencer CA, Hollowell JG, Kazarosyan M, Braverman LE. National Health and Nutrition Examination Survey III thyroid-stimulating hormone (TSH)-thyroperoxidase antibody relationships demonstrate that TSH upper reference limits may be skewed by occult thyroid dysfunction. J Clin Endocrinol Metab. 2007;92:4239-4240.
37. Walker JA, Illions EH, Huddleston JF, Smallridge RC. Racial comparisons of thyroid function and autoimmunity during pregnancy and the postpartum period. Obstet Gynecol. 2005;106:1365-1371.
38. Hollowell JG, Staehling NW, Flanders WD, et al. Serum TSH, T(4), and thyroid antibodies in the United States population (1988 to 1994): National Health and Nutrition Examination Survey (NHANES III). J Clin Endocrinol Metab. 2002;87:489-499.
39. Jonklaas J, Razvi S. Reference intervals in the diagnosis of thyroid dysfunction: treating patients not numbers. Lancet Diabetes Endocrinol. 2019;7(6):473-483.
40. Surks MI, Hollowell JG. Age-specific distribution of serum thyrotropin and antithyroid antibodies in the US population: implications for the prevalence of subclinical hypothyroidism. *J Clin Endocrinol Metab*. 2007;92(12):4575-4582.

41. Boelaert K. Thyroid dysfunction in the elderly. *Nat Rev Endocrinol*. 2013;9:194-204.

42. Szkudlinski MW, Fremont V, Ronin C, Weintraub BD. Thyroid-stimulating hormone and thyroid-stimulating hormone receptor structure-function relationships. *Physiol Rev*. 2002;82(2):473-502.

43. Zhai X, Zhang L, Chen L, et al. An age-specific serum thyrotropin reference range for the diagnosis of thyroid diseases in older adults: a cross-sectional survey in China. *Thyroid*. 2018;28(12):1571-1579.

44. Tanda ML, Piantanida E, Lai A, et al. Thyroid autoimmunity and environment. *Horm Metab Res*. 2009;41:436-442.

45. Tozzi R, D'Aurizio F, Metus P, et al. Reference intervals for serum thyrotropin in an area of Northern Italy: the Pordenone thyroid study (TRIPP). *J Endocrinol Invest*. 2018;41(8):985-994.

46. Clerico A, Trenti T, Aloe R, et al. A multicenter study for the evaluation of the reference interval for TSH in Italy (ELAS TSH Italian Study). *Clin Chem Lab Med*. 2018;57(2):259-267.

47. Zhai X, Zhang L, Chen L, et al. An age-specific serum thyrotropin reference range for the diagnosis of thyroid diseases in older adults: a cross-sectional survey in China. *Thyroid*. 2018;28(12):1571-1579.

48. van de Ven AC, Netea-Maier RT, Smit JW. Thyrotropin versus thyroid-stimulating hormone: a cross-sectional study in healthy euthyroid individuals. *Scand J Clin Lab Invest*. 2011;71:344-349.

49. Yang F, Teng W, Shan Z, et al. Epidemiological survey on the relationship between different iodine intakes and prevalence of hyperthyroidism. *Eur J Endocrinol*. 2002;146:613-618.

50. Cai J, Fang Y, Jing D, et al. Reference intervals of thyroid hormones in a previously iodine-deficient but presently more than adequate area of Western China: a population-based survey. *Endocr J*. 2016;63:381-388.

51. Zimmermann MB, Andersson M. Update on iodine status worldwide. *Curr Opin Endocrinol Diabetes Obes*. 2012;19(5):382-387.

52. Abdel-Dayem MM, Elgendy MS. Effects of chronic estradiol treatment on the thyroid gland structure and function of ovariectomized rats. *BMC Res Notes*. 2019. https://doi.org/10.1186/s13104-019-4452-0

53. Vadiveloo T, Donnan PT, Murphy MJ, Leese GP. Age- and gender-specific TSH reference intervals in people with no obvious thyroid disease in Tayside, Scotland: the Thyroid Epidemiology, Audit, and Research Study (TEARS). *J Clin Endocrinol Metab*. 2013;98(3):1147-1153.

54. Boucai L, Surks MI. Reference limits of serum TSH and free T4 are significantly influenced by race and age in an urban outpatient medical practice. *Clin Endocrinol (Oxf)*. 2009;70:788-793.

55. Knudsen N, Bulow I, Jorgensen T, et al. Comparative study of thyroid function and types of thyroid dysfunction in two areas in Denmark with slightly different iodine status. *Eur J Endocrinol*. 2000;143(4):485-491.

56. Xing D, Li R, Liu D, Zhou Q, Xu J. Factors influencing the reference interval of thyroid-stimulating hormone in healthy adults: A systematic review and meta-analysis. *Clin Endocrinol (Oxf)*. 2021;00:1–12. https://doi.org/10.1111/cen.14454