Method dependent variation in TSH and FT4 reference intervals in pregnancy: a systematic review

Journal:	*Annals of Clinical Biochemistry*
Manuscript ID	ACB-21-103.R1
Manuscript Type:	Research Article
Date Submitted by the Author:	n/a
Complete List of Authors:	Okosieme, O; Cwm Taf Morgannwg University Health Board, Diabetes and Endocrinology; Cardiff University, Thyroid Research Group, Division of Infection and Immunity Agrawal, Medha; Cwm Taf Morgannwg University Health Board, Endocrine and Diabetes Department, Prince Charles Hospital Usman, Danyal; Cwm Taf Morgannwg University Health Board, Endocrine and Diabetes Department, Prince Charles Hospital Evans, Carol; University Hospital of Wales, Clinical Biochemistry
Keywords:	Thyroid hormones < Analytes, Thyroid disease < Clinical studies, Pregnancy < Clinical studies

Method dependent variation in TSH and FT4 reference intervals in pregnancy: a systematic review

Onyebuchi E Okosieme[1,2], Medha Agrawal[1,2], Danyal Usman[1,2], and Carol Evans[3]

1. Thyroid Research Group, School of Medicine, Cardiff University
2. Diabetes Department, Prince Charles Hospital, Cwm Taf University Health Board
3. Department of Medical Biochemistry & Immunology, University Hospital of Wales, Cardiff

Onyebuchi Okosieme: OkosiemeOE@Cardiff.ac.uk
Danyal Usman: usman@cardiff.ac.uk
Medha Agrawal: medhaagrawal@doctors.org.uk
Carol Evans: carol.evans9@wales.nhs.uk

Short title: TSH and FT4 reference intervals in pregnancy

Keywords: TSH, FT4, thyroid function, reference intervals, hyperthyroidism, hypothyroidism, pregnancy

Correspondence: Dr Onyebuchi Okosieme, Diabetes Department, Prince Charles Hospital, Cwm Taf Morgannwg Health Board, Gurnos Estate, Merthyr Tydfil CF47 9DT

Funding statement: No funding was received for this study.

Conflict of interest: We declare that no conflict of interest exists with respect to this study.

Ethical approval: Not required.

Contributorship: Concept and design: OEO, CE
Data acquisition: OEO, MA, UD, CE
Data analysis: OEO
Writing and editing: OEO, MA, UD, CE

ABSTRACT

Background: Gestational TSH and FT4 reference intervals may differ according to assay method but the extent of variation is unclear and has not been systematically evaluated. We conducted a systematic review of published studies on TSH and FT4 reference intervals in pregnancy. Our aim was to quantify method-related differences in gestation reference intervals, across four commonly used assay methods, Abbott, Beckman, Roche, and Siemens.

Methods: We searched the literature for relevant studies, published between January 2000 and December 2020, in healthy pregnant women without thyroid antibodies or disease. For each study, we extracted trimester-specific reference intervals (2.5–97.5 percentiles) for TSH and FT4 as well as the manufacturer provided reference interval for the corresponding non-pregnant population.

Results: TSH reference intervals showed a wide range of study-to-study differences with upper limits ranging from 2.33 to 8.30 mU/L. FT4 lower limits ranged from 4.40–13.93 pmol/L, with consistently lower reference intervals observed with the Beckman method. Differences between non-pregnant and first trimester reference intervals were highly variable, and for most studies the TSH upper limit in the first trimester could not be predicted or extrapolated from non-pregnant values.

Conclusions: Our study confirms significant intra and inter-method disparities in gestational thyroid hormone reference intervals. The relationship between pregnant and non-pregnant values is inconsistent and does not support the existing practice in many laboratories of extrapolating gestation references from non-pregnant values. Laboratories should invest in deriving method-specific gestation reference intervals for their population.

INTRODUCTION

Thyroid dysfunction is common in females of reproductive age and occurs in 2-5% of pregnant women[1, 2]. Uncorrected thyroid dysfunction in pregnancy has deleterious effects on fetal and maternal health including an increased risk of pregnancy loss and offspring intellectual impairment[3, 4]. Prompt detection and correction of thyroid dysfunction is therefore essential for optimal fetal and maternal outcomes[5-7]. However, the laboratory diagnosis of thyroid dysfunction in pregnancy is confounded by a series of adaptive physiological changes that translate to clinically meaningful differences between pregnant and non-pregnant thyroid hormone reference intervals. In addition, thyroid hormone concentrations change through the course of pregnancy. Total thyroid hormone concentrations rise in early pregnancy due to increased production of thyroxine-binding globulin (TBG) together with stimulation of the thyroid stimulating hormone (TSH) receptor by human chorionic gonadotrophin[8]. The increased thyroid hormone output is in turn accompanied by a fall in TSH concentration through pituitary thyroid feedback[9]. Free thyroid hormones, on the other hand, are maintained within the normal range, but free thyroxine (FT4) immunoassays are susceptible to method dependent bias in pregnancy due to variations in albumin and TBG concentrations.

The challenges of method-dependent bias in TSH and FT4 reference intervals are well-recognised[10, 11], but the extent of assay related variation in pregnancy is unclear and has not been systematically evaluated. Current international guidelines advocate the use of trimester-specific normative values derived from a healthy pregnant population in the evaluation of thyroid dysfunction in pregnancy[12]. In reality many laboratories lack gestation-specific reference intervals and apply arbitrary non-pregnant cut-offs, creating the potential for misdiagnosis and inappropriate therapy. In the absence of gestation specific reference intervals, the American Thyroid

Association (ATA) guidelines recommend that the first trimester upper and lower TSH reference limits should be set at 0.5 and 0.4 mU/L below the corresponding upper and lower non-pregnant limits, respectively. These empirical cut-offs are selected to reflect the magnitude of the anticipated difference in the non-pregnant and pregnant values based on the expected TSH drop in early gestation[12]. However, the validity of this approach for different assay methods has not been systematically evaluated.

Thus, we conducted a systematic review of published studies on TSH and FT4 reference intervals in pregnancy. Our primary aim was to quantify method-related differences in reference intervals across four frequently used manufacturer assays, namely, Abbott, Beckman, Roche, and Siemens. In addition, we examined the relationship between pregnant and non-pregnant reference intervals, and thus, the validity of extrapolating gestation reference intervals from non-pregnant intervals for the different assay methods.

METHODS

Search strategy

We searched Medline for published articles on thyroid hormone reference intervals in pregnancy between January 2000 to December 2020. We used various combinations of the search terms: "thyroid function", "FT4", "thyroxine", "TSH", “thyrotropin”, "pregnancy", "gestation", "reference range", and "reference interval". We sourced additional publications from references in individual articles. Relevant articles were selected after reading through titles and abstracts or full texts when the title or abstract information was insufficient to exclude the study.

Study selection and data extraction

We selected articles in which thyroid hormones were measured using one of four assay methods, Abbott Architect, Beckman Access or Dxl, Roche Cobas or Elecsys,

and Siemens Advia Centaur. We included only studies that reported reference intervals as 2.5–97.5 centiles with gestational age information at the time of blood sampling. We excluded studies if they were not in English, had less than 120 patients, did not exclude women with positive antibodies or thyroid disease, or were conducted in areas with known excess or deficient iodine nutrition status. The extracted information comprised first author, country of study, population ethnicity, number of subjects, age distribution, trimester of sampling, TSH and FT4 reference intervals, and reference intervals for the corresponding non-pregnant population. Non-pregnant reference intervals were extracted from the manufacturer provided values as reported by the authors. Where manufacturer reference intervals were not stated, study derived non-pregnant reference intervals were used if available. Study selection and data extraction were independently conducted by two reviewers (MA, DU) and differences were resolved by consensus and referral to other reviewers (OO, CE).

Study quality

We assessed the methodological quality of studies using the Newcastle Ottawa Scale (NOS) for the assessment of non-randomised studies. The NOS was adapted for this study to assess study selection (3 points), representativeness of the sample to a healthy pregnant population (3 points), and the assessment and reporting of reference intervals (3 points).

Data analysis

Reference intervals were summarised for each study as 2.5–97.5 percentiles and grouped by assay method and trimester of pregnancy. Where multiple results were available in the same trimester, we selected the data point most representative of that trimester. We were unable to undertake a conventional meta-analysis as most studies did not include standard measures of variance for the lower and upper reference

intervals. Thus, we described the range for the lower and upper reference limits for each assay method in each trimester and compared study-to-study as well as inter-method variation. In addition, we summarised the TSH and FT4 lower and upper reference limits using median and interquartile range, with each study represented as an unweighted data point. Method dependent differences in reference limits were then compared using the Kruskal Wallis test with the Bonferroni correction applied for multiple group comparisons. The Kruskal–Wallis test is a non-parametric method for comparing two or more independent samples while the Bonferroni correction was applied to reduce the risk of a type 1 error from multiple comparisons. To explore the validity of extrapolating gestational reference intervals from non-pregnant values, we summarised the magnitude of the difference between non-pregnant (NP) and first trimester (T1) reference limits (NP–T1) for each study. Inter-method differences in NP–T1 medians were also compared with the Kruskal Wallis test and Bonferroni correction. All analysis was conducted using Stata, version 15.1, StataCorp, Texas, USA.

RESULTS

Study Selection

The study selection flow chart is presented in figure 1. After excluding duplicate retrievals, we identified 779 studies which we screened by reading through their titles or abstracts. The full-text of 134 articles were assessed for eligibility of which 91 studies were excluded for various reasons including unavailability of 2.5–97.5 percentile reference intervals, non-exclusion of thyroid disease or antibody-positive individuals, use of assay methods other than those being assessed, samples <120 subjects, and populations with iodine deficiency or excess (figure 1). The final study sample thus comprised 43 studies [13-55].

Study characteristics

The characteristics of included studies are shown in supplementary table 1. Out of the 43 selected studies, 19 were conducted in Asian countries, predominantly China (n=16) while 15 studies were from European countries. Other studies were from North America (n=3), South America (n=3), the Middle East (n=2), and Australia (n=1). The studies included a total number of 132,794 pregnant women, comprising 68,097 samples analysed by Abbott (14 studies), 15,164 by Beckman (9 studies), 30,903 by Roche (15 studies), and 21,819 by Siemens (11 studies). Nineteen studies excluded women with antibodies to either thyroid peroxidase (TPOAb) or thyroglobulin (TgAb) [13, 14, 17-20, 22, 31, 33, 34, 36, 39, 41, 44, 45, 47, 51, 52, 55], while 24 studies did not measure TgAbs and excluded women with positive TPOAb only [15, 16, 21, 23-25, 27-30, 32, 35, 37, 38, 40, 42, 43, 46, 48-50, 53, 54]. The median age of patients ranged from 24 to 35 years with TSH and FT4 reference intervals determined during the 1st, 2nd, and 3rd trimesters in 42, 28, and 26 studies, respectively. Studies that presented data separately for patients with different ethnicities and with multiple assay methods are presented separately. The quality scores ranged from 6–9, and most studies scored between 7 and 8 points.

TSH reference intervals

TSH reference intervals (2.5–97.5 percentile) for the 1st to 3rd trimesters are shown in figures 2–4, respectively. In the first trimester, the TSH lower limit ranged from 0.01–0.59 mu/L, with most studies reporting a TSH lower limit <0.20 mU/L (figure 2a). The upper limit showed greater study-to-study variation and within-method variation which were observed for all assay methods in the first trimester (figures 2a). The Abbott assays showed the widest variation, with a TSH upper limit range of 2.33–8.30 mU/L, including a study by Dhatt et al, that reported extremely high upper limits in women of Arab and Asian ethnicity [15](figure 2a). The intra-method variation in TSH upper limits

continued into the 2nd and 3rd trimesters while the lower limits remained <0.50 mu/L in the 2nd trimester and <0.60 mU/L in the third trimester (figures 3a, 4a). Comparisons of medians across methods showed no significant method related difference for the lower or upper TSH limit in all trimesters (P>0.05, supplementary table 2). Three studies with inter-method measurements in the same subjects (Fan[16], Springer[42], Liu[18]) also reported no consistent pattern of method-related differences in TSH reference intervals. Distribution of TSH lower and upper limits by trimester and assay methods are shown in figure 5. TSH limits for each assay were progressively higher in each trimester (figures 5a, 5b).

FT4 reference intervals

FT4 reference intervals (2.5–97.5 percentile) are shown in figures 2–4. Reference intervals varied across studies in all trimesters and was present within as well as across assay methods. The Beckman method consistently yielded lower FT4 reference intervals than other assay methods. FT4 lower limits in the 1st trimester ranged from 7.16–12.37, 5.90–10.81, 10.30–13.41, and 9.01–13.93 pmol/L for the Abbott, Beckman, Roche, and Siemens assays, respectively. The upper limits ranged from 15.96–24.60, 13.20–18.66, 18.00–22.50, and 16.73–26.49 pmol/L for the Abbott, Beckman, Roche, and Siemens assays, respectively. The Beckman upper limit reported in some studies was lower than the Roche or Siemens lower limit in other studies. Relatively lower Beckman concentrations were also observed in the study by Liu *et al* which measured FT4 using the Beckman, Abbot and Roche assays in the same patients[18]. The distribution of FT4 lower and upper limits by trimester and assay method is presented in figure 5. FT4 reference intervals got progressively lower with each trimester but method related differences persisted in the 2nd and 3rd trimesters. Comparison of median lower and upper FT4 limits consistently showed lower

Beckman values compared to other methods, in all trimesters (P<0.05, supplementary table 2).

Difference between non-pregnant and first trimester reference intervals

To examine the validity of extrapolating gestational reference intervals from non-pregnant values, we determined the difference between non-pregnant and first trimester reference limits (NP–T1) for TSH and FT4 (figure 6). For the TSH lower limit most NP–T1 values were in the 0–0.5 mU/L range, and thus roughly consistent with the recommendation to derive gestation TSH lower limit by subtracting 0.4 from the non-pregnant lower limit. In contrast there was greater variation for the upper limit with differences ranging from –3.98 to +2.72 mU/L. TSH upper limit NP–T1 was >1.0 mU/L in 18 studies meaning that the recommended subtraction of 0.5 mU/L from the non-pregnant upper limit would have over-estimated the gestation TSH upper limits by at least 0.5 mU/L in these samples.

TSH upper limit NP–T1 was negative in 8 studies indicating that the 0.5 mU/L subtraction would under-estimate gestation TSH upper limits in these samples. Only 15 studies (4 Abbott, 1 Beckman, 6 Roche, 4 Siemens) had a TSH upper limit NP–T1 in the 0–1.0 mu/L range i.e., roughly equivalent with the 0.5 mU/L difference. No single assay method showed a consistent pattern of difference between non-pregnant and gestation upper TSH limit. Using the ratio of the non-pregnant and gestation TSH upper limits (NP/T1) also gave highly variable results (data not shown). NP–T1 for the FT4 lower and upper limits were also variable and ranged from –2.76 to +2.50 pmol/L for the lower limit and –6.0 to +6.0 pmol/L for the upper limit with no specific method related patterns (figure 6).

Ethnicity

We explored the influence of ethnicity on reference intervals by grouping the data according to the two most frequently represented ethnic groups in the studies i.e., Chinese and Caucasians (21 studies each). Supplementary figure 1 shows the distribution of TSH and FT4 reference limits according to trimester, assay method, and ethnicity. Statistical comparison of reference limits by ethnicity was not feasible due to small group numbers. However, Roche assays tended to report higher TSH upper limits for Chinese compared to Caucasian patients (median TSH 4.80 vs 3.40 mU/L, supplementary figure 1b). A study of reference intervals in women of Arab and Asian ethnicity by Dhatt *et al* reported no difference in TSH reference intervals but showed lower FT4 reference intervals in Arab compared to Asian women trimesters 1 and 2[15] (figures 3 and 4).

DISCUSSION

We have undertaken a systematic review of published reports on thyroid hormone reference intervals in pregnancy with the aim of evaluating the variation across assay methods. We observed marked variation for the TSH upper limit with a wide range of study-to-study differences affecting all analytical methods. The Beckman assays yielded comparatively lower FT4 reference intervals that were incongruent with other methods. We also explored the validity of existing strategies in many laboratories of estimating gestational reference intervals from intervals derived from the non-pregnant population. Marked variation was observed in the difference between non-pregnant and first trimester reference intervals, and no single assay method showed a consistent pattern of difference. Our study thus confirms significant method related disparities in gestational thyroid hormone reference intervals and highlights the limitations of applying general population reference intervals in pregnancy.

Method related differences in FT4 and TSH measurements have been well documented in the non-pregnant population[10, 56]. In addition, the UK National External Quality Assessment Scheme (NEQAS) also reported method related variation in thyroid function reference intervals including relatively lower FT4 concentrations for the Beckman assays [57]. ~~but~~ However, only a few studies have systematically addressed these differences in pregnancy. In the study by Springer *et al*, gestational thyroid hormone reference intervals were established with 7 different analytical systems[42]. The authors reported significant inter-method differences for both TSH and FT4 intervals, with the lowest FT4 intervals observed with the Beckman assay[42]. [57]Several authoritative narrative reviews of pregnancy reference intervals have previously confirmed these assay dependent differences in FT4 and TSH intervals and highlighted their potential clinical implications [58, 59]. A meta-analysis of TSH and FT4 gradients from the non-pregnant to pregnant state also showed assay related variation and suggested that the upper TSH cut-off in pregnancy could be approximated by subtracting 22% from the non-pregnant TSH upper limit[60]. However, this analysis was limited to studies conducted exclusively in Chinese populations[60]. In contrast we were unable to show a consistent pattern of difference between the non-pregnant and pregnant TSH upper limit, perhaps due to inclusion of a wider range of studies in our analysis.

Our findings have implications for clinical practice. Uncorrected hypothyroidism carries an increased risk of fetal loss[6] and offspring intellectual impairment[61]. Furthermore, ~~unwarranted~~ maternal over-treatment with Levothyroxine ~~administration~~ in pregnancy may increase the risk of cognitive dysfunction and attention deficit hyperactivity disorders in children[61, 62]. Over-estimating TSH upper limits would miss cases of gestational hypothyroidism while under-estimation would wrongly diagnose

hypothyroidism, putting women without thyroid dysfunction at risk of unnecessary and potentially harmful therapy. The need for assay-dependent reference intervals is even more pressing for FT4 reference intervals due to the striking method discrepancies observed in these series. These considerations remain pertinent given that many laboratories lack gestation specific reference intervals and continue to apply non-pregnant intervals in pregnancy. Our findings show that gestation reference intervals cannot reliably be deduced from the non-pregnant range and that the ATA recommendation to subtract 0.5 mU/L from the non-pregnant upper limit would over or under-estimate the upper TSH limit in the majority of samples.

Ideally each laboratory should derive its own gestational reference intervals based on the assay method and local population. This is not always practicable, particularly for small laboratories with limited resources. One approach would be for health authorities to collaborate at regional level to establish reference intervals for the commonly used assay methods within the region. The establishment of reference intervals should follow criteria set by international bodies[11, 63]. Furthermore, the reporting of gestational thyroid function tests should be assay and pregnancy specific and clinicians should be alert to the potential for method related differences. For laboratories that lack gestation specific data the use of arbitrary cut-off points is now discouraged, and best practice in the circumstance would be to use reference intervals derived from a population with similar assay platform and comparable characteristics in terms of ethnicity and iodine nutrition. If non-pregnant reference intervals must be used, then clinicians need to be aware of the limitations of such an approach. Clinical studies investigating the impact of thyroid dysfunction should avoid outcome analyses based on fixed cut-offs and use comparable measures of population percentiles or multiples of medians as has previously been suggested[59].

Our study has some limitations. Because our review covers a 20-year period, it is likely that assay methods would have changed with time and some of the older studies may not reflect current methods. ~~W~~in addition, we were only able to evaluate the most commonly used assay platforms and as such the variation in other assay methods is unknown. ~~Furthermore~~ Also, some of the observed variation may reflect differences in laboratory quality standards as well as unmeasured confounders such as iodine nutrition. Lack of iodine nutrition data in most studies meant that we could not formally assess the impact of iodine status on reference intervals. For example, the study by Dhatt *et al* in a mixed-ethnic United Arab Emirate population, reported unequivocally raised TSH values suggesting unrecognised iodine deficiency or thyroid dysfunction in their cohort[15]. Lastly, we were unable to conduct a conventional meta-analysis of the lower and upper reference limits since most studies did not provide data distribution measures for these limits such as standard deviation or 95% confidence intervals. Instead, we adopted a pragmatic approach in which each study was represented as a single unweighted data point and medians for the lower and upper reference limits were compared using non-parametric methods. While this approach provides crude estimates of inter-method differences, it might have lacked the sensitivity to detect more subtle variation.

Our study's strength is that it is the first systematic review to focus on assay dependent differences in thyroid hormone reference intervals in pregnancy. We have used stringent inclusion criteria to systematically select relevant studies and to summarise a large body of data spanning 20 years. Lastly, we have probed the validity of current guideline recommendations and highlight practical challenges facing laboratories without gestation-specific reference interval data.

In conclusion we show wide variation in thyroid hormone reference intervals both within and across assay methods. We found no consistent relationship between the non-pregnant and pregnant reference intervals to permit extrapolation of pregnancy intervals from non-pregnant intervals. Future guidelines should acknowledge the limitations of current approaches, and efforts should now be invested in deriving gestation reference intervals that are assay and population specific.

REFERENCES

1. Carlé A, Pedersen IB, Knudsen N, et al. Epidemiology of subtypes of hyperthyroidism in Denmark: a population-based study. *Eur J Endocrinol*. 2011; 164: 801-9.

2. Taylor PN, Albrecht D, Scholz A, et al. Global epidemiology of hyperthyroidism and hypothyroidism. *Nat Rev Endocrinol*. 2018; 14: 301-16.

3. Chan S and Boelaert K. Optimal management of hypothyroidism, hypothyroxinaemia and euthyroid TPO antibody positivity preconception and in pregnancy. *Clin Endocrinol (Oxf)*. 2015; 82: 313-26.

4. Krassas GE, Poppe K and Glinoer D. Thyroid Function and Human Reproductive Health. *Endocrine Reviews*. 2010; 31: 702-55.

5. Abalovich M, Gutierrez S, Alcaraz G, Maccallini G, Garcia A and Levalle O. Overt and subclinical hypothyroidism complicating pregnancy. *Thyroid*. 2002; 12: 63-8.

6. Taylor PN, Minassian C, Rehman A, et al. TSH levels and risk of miscarriage in women on long-term levothyroxine: a community-based study. *J Clin Endocrinol Metab*. 2014; 99: 3895-902.

7. Maraka S, Mwangi R, McCoy RG, et al. Thyroid hormone treatment among pregnant women with subclinical hypothyroidism: US national assessment. *Bmj*. 2017; 356: i6865.

8. Glinoer D. The regulation of thyroid function in pregnancy: pathways of endocrine adaptation from physiology to pathology. *Endocr Rev*. 1997; 18: 404-33.

9. Okosieme OE, Khan I and Taylor PN. Preconception management of thyroid dysfunction. *Clin Endocrinol (Oxf)*. 2018; 89: 269-79.

10. Barth JH, Luvai A, Jassam N, et al. Comparison of method-related reference intervals for thyroid hormones: studies from a prospective reference population and a literature review. *Ann Clin Biochem*. 2018; 55: 107-12.

11. Thienpont LM, Van Uytfanghe K, Van Houcke S, et al. A Progress Report of the IFCC Committee for Standardization of Thyroid Function Tests. *Eur Thyroid J*. 2014; 3: 109-16.

12. Alexander EK, Pearce EN, Brent GA, et al. 2017 Guidelines of the American Thyroid Association for the Diagnosis and Management of Thyroid Disease During Pregnancy and the Postpartum. *Thyroid*. 2017; 27: 315-89.

13. Akarsu S, Akbiyik F, Karaismailoglu E and Dikmen ZG. Gestation specific reference intervals for thyroid function tests in pregnancy. *Clinical chemistry and laboratory medicine : CCLM / FESCC*. 2016.

14. Bocos-Terraz JP, Izquierdo-Alvarez S, Bancalero-Flores JL, et al. Thyroid hormones according to gestational age in pregnant Spanish women. *BMC research notes*. 2009; 2: 237.

15. Dhatt GS, Jayasundaram R, Wareth LA, et al. Thyrotrophin and free thyroxine trimester-specific reference intervals in a mixed ethnic pregnant population in the United Arab Emirates. *Clinica chimica acta; international journal of clinical chemistry*. 2006; 370: 147-51.

16. Fan JX, Yang S, Qian W, Shi FT and Huang HF. Comparison of the Reference Intervals Used for the Evaluation of Maternal Thyroid Function During Pregnancy Using Sequential and Nonsequential Methods. *Chinese medical journal*. 2016; 129: 785-91.

17. Ho CKM, Tan ETH, Ng MJ, et al. Gestational age-specific reference intervals for serum thyroid hormone levels in a multi-ethnic population. *Clinical chemistry and laboratory medicine : CCLM / FESCC*. 2017; 55: 1777-88.

18. Liu J, Yu X, Xia M, et al. Development of gestation-specific reference intervals for thyroid hormones in normal pregnant Northeast Chinese women: What is the rational division of gestation stages for establishing reference intervals for pregnancy women? *Clinical biochemistry*. 2017; 50: 309-17.

19. Männistö T, Surcel HM, Ruokonen A, et al. Early pregnancy reference intervals of thyroid hormone concentrations in a thyroid antibody-negative pregnant population. *Thyroid*. 2011; 21: 291-8.

20. Andersen SL, Andersen S, Carlé A, et al. Pregnancy Week-Specific Reference Ranges for Thyrotropin and Free Thyroxine in the North Denmark Region Pregnancy Cohort. *Thyroid*. 2019; 29: 430-8.

21. Benhadi N, Wiersinga WM, Reitsma JB, Vrijkotte TG, van der Wal MF and Bonsel GJ. Ethnic differences in TSH but not in free T4 concentrations or TPO antibodies during pregnancy. *Clin Endocrinol (Oxf)*. 2007; 66: 765-70.

22. Derakhshan A, Shu H, Broeren MAC, et al. Reference Ranges and Determinants of Thyroid Function During Early Pregnancy: The SELMA Study. *J Clin Endocrinol Metab*. 2018; 103: 3548-56.

23. Donovan LE, Metcalfe A, Chin A, et al. A Practical Approach for the Verification and Determination of Site- and Trimester-Specific Reference Intervals for Thyroid Function Tests in Pregnancy. *Thyroid*. 2019; 29: 412-20.

24. Duan Y, Peng L, Cui Y and Jiang Y. Reference Intervals for Thyroid Function and the Negative Correlation between FT4 and HbA1c in Pregnant Women of West China. *Clinical laboratory*. 2015; 61: 777-83.

25. Ekinci EI, Lu ZX, Sikaris K, et al. Longitudinal assessment of thyroid function in pregnancy. *Ann Clin Biochem*. 2013; 50: 595-602.

26. Han L, Zheng W, Zhai Y, et al. Reference intervals of trimester-specific thyroid stimulating hormone and free thyroxine in Chinese women established by experimental and statistical methods. *J Clin Lab Anal*. 2018; 32: e22344.

27. Jiang YX, Sun WJ, Zhang Y, et al. Thyroid function of twin-pregnant women in early pregnancy. *Chinese medical journal*. 2019; 132: 2033-8.

28. Khalid AS, Marchocki Z, Hayes K, et al. Establishing trimester-specific maternal thyroid function reference intervals. *Ann Clin Biochem*. 2014; 51: 277-83.

29. Kianpour M, Aminorroaya A, Amini M, et al. Reference Intervals for Thyroid Hormones During the First Trimester of Gestation: A Report from an Area with a Sufficient Iodine Level. *Hormone and metabolic research = Hormon- und Stoffwechselforschung = Hormones et metabolisme*. 2019; 51: 165-71.

30. Kostecka-Matyja M, Fedorowicz A, Bar-Andziak E, et al. Reference Values for TSH and Free Thyroid Hormones in Healthy Pregnant Women in Poland: A Prospective, Multicenter Study. *Eur Thyroid J*. 2017; 6: 82-8.

31. Li C, Shan Z, Mao J, et al. Assessment of thyroid function during first-trimester pregnancy: what is the rational upper limit of serum TSH during the first trimester in Chinese pregnant women? *J Clin Endocrinol Metab*. 2014; 99: 73-9.

32. Lin L, Zhang XL and Long Y. Analysis of thyroid peroxidase antibody in early pregnancy. *Genetics and molecular research : GMR*. 2014; 13: 5107-14.

33. Moon HW, Chung HJ, Park CM, Hur M and Yun YM. Establishment of trimester-specific reference intervals for thyroid hormones in Korean pregnant women. *Annals of laboratory medicine*. 2015; 35: 198-204.

34. Morais N, Assis ASA, Corcino CM, et al. Recent recommendations from ATA guidelines to define the upper reference range for serum TSH in the first trimester match reference ranges for pregnant women in Rio de Janeiro. *Archives of endocrinology and metabolism*. 2018; 62: 386-91.

35. Mosso L, Martínez A, Rojas MP, et al. Early pregnancy thyroid hormone reference ranges in Chilean women: the influence of body mass index. *Clin Endocrinol (Oxf)*. 2016; 85: 942-8.

36. Ollero MD, Toni M, Pineda JJ, Martínez JP, Espada M and Anda E. Thyroid Function Reference Values in Healthy Iodine-Sufficient Pregnant Women and Influence of Thyroid Nodules on Thyrotropin and Free Thyroxine Values. *Thyroid*. 2019; 29: 421-9.

37. Pearce EN, Oken E, Gillman MW, et al. Association of first-trimester thyroid function test values with thyroperoxidase antibody status, smoking, and multivitamin use. *Endocr Pract*. 2008; 14: 33-9.

38. Pop VJ, Ormindean V, Mocan A, et al. Maternal cognitive function during pregnancy in relation to hypo- and hyperthyroxinemia. *Clin Endocrinol (Oxf)*. 2019; 91: 824-33.

39. Quinn FA, Reyes-Mendez MA, Nicholson L, Compean LP and Tavera ML. Thyroid function and thyroid autoimmunity in apparently healthy pregnant and non-pregnant Mexican women. *Clinical chemistry and laboratory medicine : CCLM / FESCC*. 2014; 52: 1305-11.

40. Šálek T, Dhaifalah I, Langova D and Havalová J. Maternal thyroid-stimulating hormone reference ranges for first trimester screening from 11 to 14 weeks of gestation. *J Clin Lab Anal*. 2018; 32: e22405.

41. Shen FX, Xie ZW, Lu SM, Aw TC and Zhu B. Gestational thyroid reference intervals in antibody-negative Chinese women. *Clinical biochemistry*. 2014; 47: 673-5.

42. Springer D, Bartos V and Zima T. Reference intervals for thyroid markers in early pregnancy determined by 7 different analytical systems. *Scandinavian journal of clinical and laboratory investigation*. 2014; 74: 95-101.

43. Springer D, Zima T and Limanova Z. Reference intervals in evaluation of maternal thyroid function during the first trimester of pregnancy. *Eur J Endocrinol*. 2009; 160: 791-7.

44. Stricker R, Echenard M, Eberhart R, et al. Evaluation of maternal thyroid function during pregnancy: the importance of using gestational age-specific reference intervals. *Eur J Endocrinol*. 2007; 157: 509-14.

45. Sun R and Xia J. The Reference Intervals of Thyroid Hormones for Pregnant Women in Zhejiang Province. *Laboratory medicine*. 2017; 49: 5-10.

46. Veltri F, Belhomme J, Kleynen P, et al. Maternal thyroid parameters in pregnant women with different ethnic backgrounds: Do ethnicity-specific reference ranges improve the diagnosis of subclinical hypothyroidism? *Clin Endocrinol (Oxf)*. 2017; 86: 830-6.

47. Yan YQ, Dong ZL, Dong L, et al. Trimester- and method-specific reference intervals for thyroid tests in pregnant Chinese women: methodology, euthyroid definition and iodine status can influence the setting of reference intervals. *Clin Endocrinol (Oxf)*. 2011; 74: 262-9.

48. Yang X, Meng Y, Zhang Y, et al. Thyroid function reference ranges during pregnancy in a large Chinese population and comparison with current guidelines. *Chinese medical journal*. 2019; 132: 505-11.

49. Yu B, Wang QW, Huang RP, et al. Establishment of self-sequential longitudinal reference intervals of maternal thyroid function during pregnancy. *Experimental biology and medicine (Maywood, NJ)*. 2010; 235: 1212-5.

50. Zhang D, Cai K, Wang G, et al. Trimester-specific reference ranges for thyroid hormones in pregnant women. *Medicine*. 2019; 98: e14245.

51. Zhang J, Li W, Chen QB, et al. Establishment of trimester-specific thyroid stimulating hormone and free thyroxine reference interval in pregnant Chinese

women using the Beckman Coulter UniCel™ DxI 600. *Clinical chemistry and laboratory medicine : CCLM / FESCC*. 2015; 53: 1409-14.

52. Zhang X, Yao B, Li C, et al. Reference Intervals of Thyroid Function During Pregnancy: Self-Sequential Longitudinal Study Versus Cross-Sectional Study. *Thyroid*. 2016; 26: 1786-93.

53. Zhou Q, Zhang Y, Zhou J, et al. Analysis of detection results of thyroid function-related indexes in pregnant women and establishment of the reference interval. *Exp Ther Med*. 2019; 17: 1749-55.

54. Wong SL, Webster GM, Venners S and Mattman A. Second trimester thyroid-stimulating hormone, total and free thyroxine reference intervals for the Beckman Coulter Access® 2 platform. *Clinica chimica acta; international journal of clinical chemistry*. 2014; 428: 96-8.

55. Kim HJ, Cho YY, Kim SW, et al. Reference intervals of thyroid hormones during pregnancy in Korea, an iodine-replete area. *The Korean journal of internal medicine*. 2018; 33: 552-60.

56. Kalaria T, Sanders A, Fenn J, et al. The diagnosis and management of subclinical hypothyroidism is assay dependent implications for clinical practice. *Clin Endocrinol (Oxf)*. 2021.

57. MacKenzie F. UK NEQAS for thyroid hormones. https://birminghamquality.org.uk/assets/in-house/doc/eqa/thy-428.pdf 2018. Accessed 12th May 2021.

58. McNeil AR and Stanford PE. Reporting Thyroid Function Tests in Pregnancy. *The Clinical biochemist Reviews / Australian Association of Clinical Biochemists*. 2015; 36: 109-26.

59. Medici M, Korevaar TI, Visser WE, Visser TJ and Peeters RP. Thyroid function in pregnancy: what is normal? *Clin Chem*. 2015; 61: 704-13.

60. Gao X, Li Y, Li J, et al. Gestational TSH and FT4 Reference Intervals in Chinese Women: A Systematic Review and Meta-Analysis. *Frontiers in endocrinology*. 2018; 9: 432.

61. Korevaar TI, Muetzel R, Medici M, et al. Association of maternal thyroid function during early pregnancy with offspring IQ and brain morphology in childhood: a population-based prospective cohort study. *Lancet Diabetes Endocrinol*. 2016; 4: 35-43.

62. Hales C, Taylor PN, Channon S, et al. Controlled Antenatal Thyroid Screening II: Effect of Treating Maternal Suboptimal Thyroid Function on Child Behavior. *J Clin Endocrinol Metab*. 2020; 105.

63. Demers LM and Spencer CA. Laboratory medicine practice guidelines: laboratory support for the diagnosis and monitoring of thyroid disease. *Clin Endocrinol (Oxf)*. 2003; 58: 138-40.

TABLES AND FIGURES

Figure 1: Study selection flow chart

Figure 2: 1st trimester TSH and FT4 reference ranges

Figure 3: 2nd trimester TSH and FT4 reference ranges

Figure 4: 3rd trimester TSH and FT4 reference ranges

Figure 5: TSH lower and upper limits by assay method

Each circle represents the lower or upper limit reported in each study.

Figure 6: Non-pregnant minus 1st trimester (NP–T1) lower and upper reference limits

Legend: Circles represent data points from each study. The non-pregnant data was based on the manufacturer provided reference range for the corresponding non-pregnant population, as reported in the study. The dashed vertical lines in panel (a) (0–0.4) and panel (b) (0–0.5) represent the expected NP–T1 difference based on guideline recommendations for the lower and upper TSH limits, respectively.

Supplementary table 1: Study characteristics

Supplementary table 2: Inter-method comparisons for TSH lower and upper limits

Supplementary figure 1: TSH and FT4 lower and upper limits by ethnicity

Legend: Circles represent data points from each study. Studies in subjects of Chinese ethnicity are represented by white circles while studies in Caucasian subjects are represented by grey circles.

Figure 1: Study selection flow chart

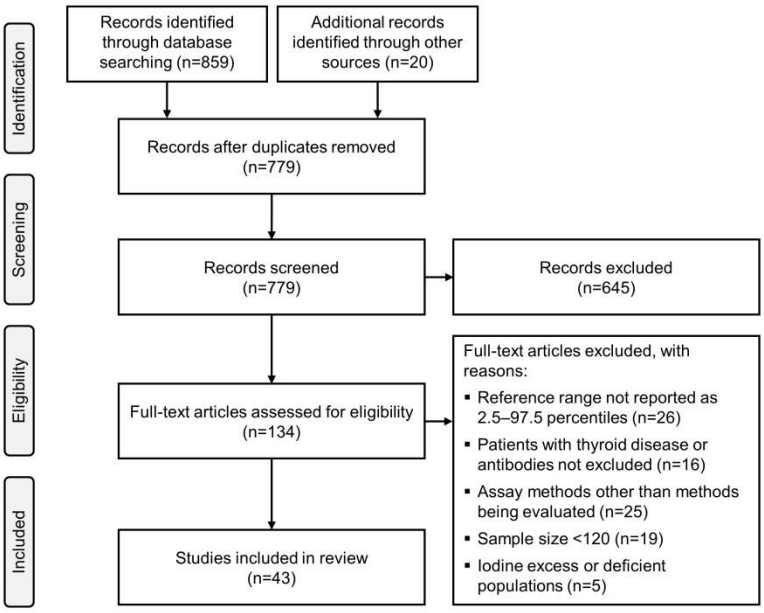


Figure 1: Study selection flow chart

254x190mm (300 x 300 DPI)

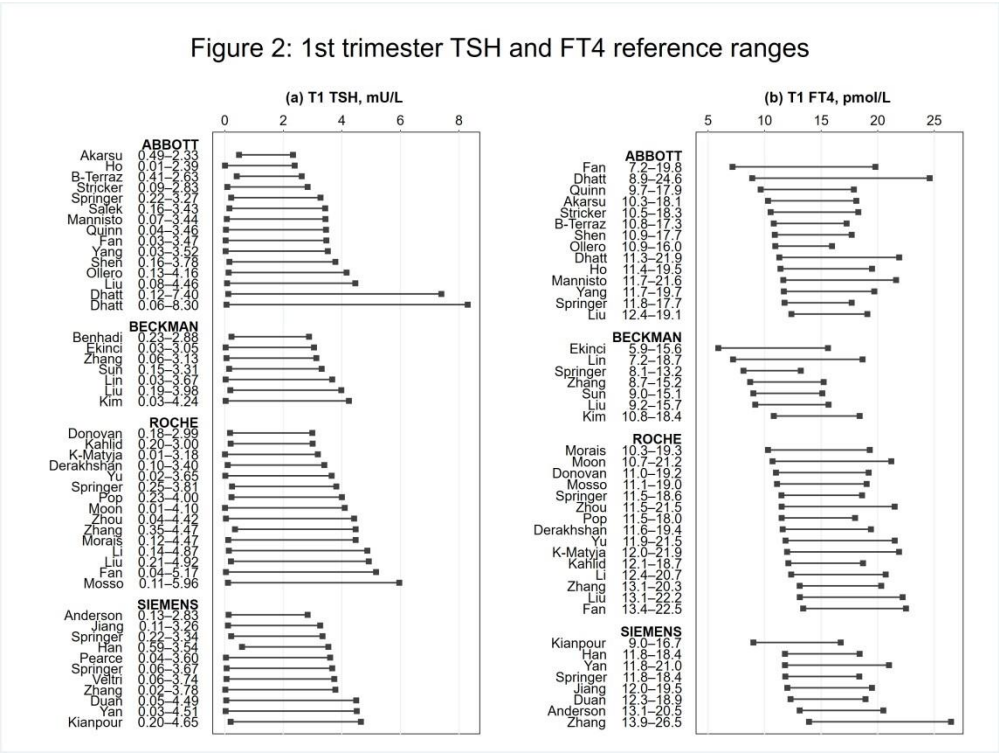


Figure 2: 1st trimester TSH and FT4 reference ranges

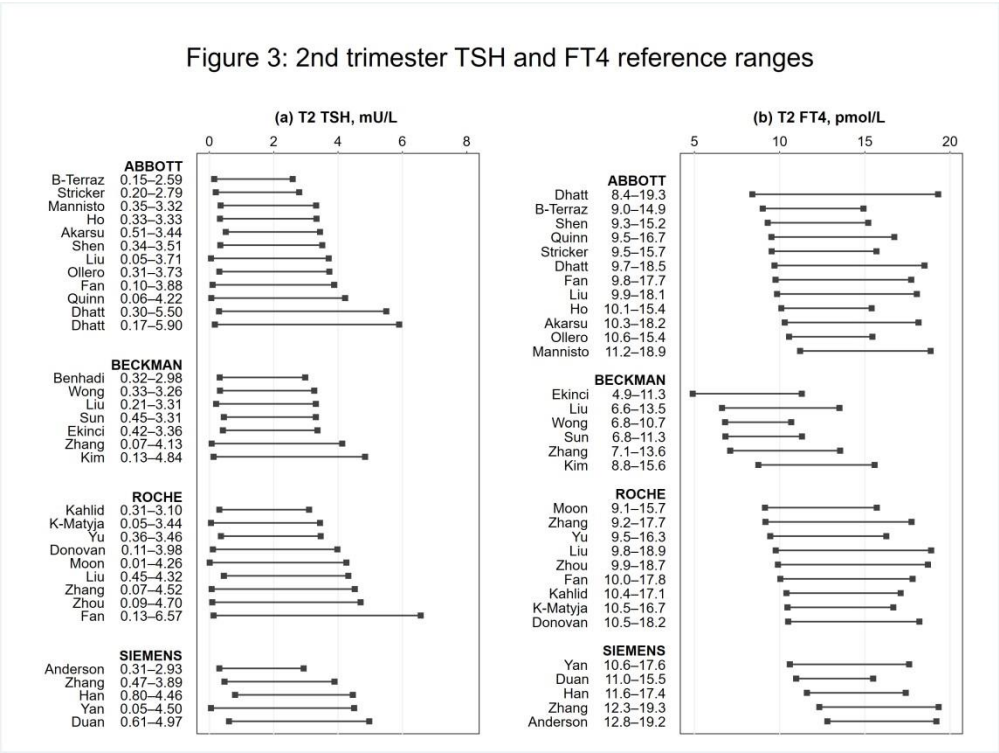


Figure 3: 2nd trimester TSH and FT4 reference ranges

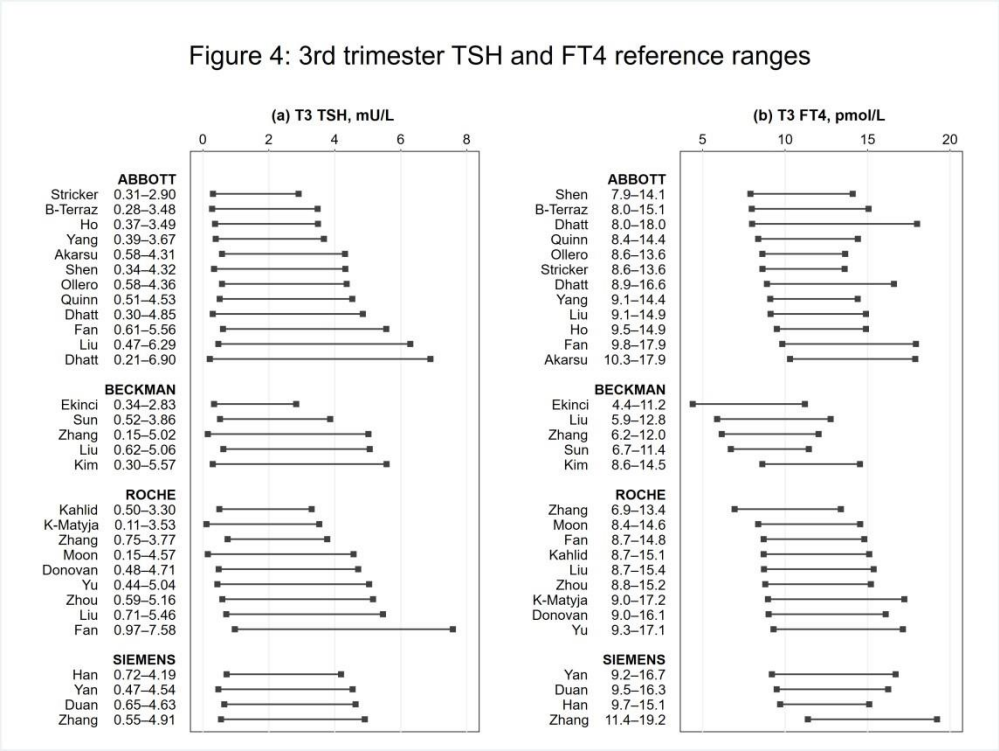


Figure 4: 3rd trimester TSH and FT4 reference ranges

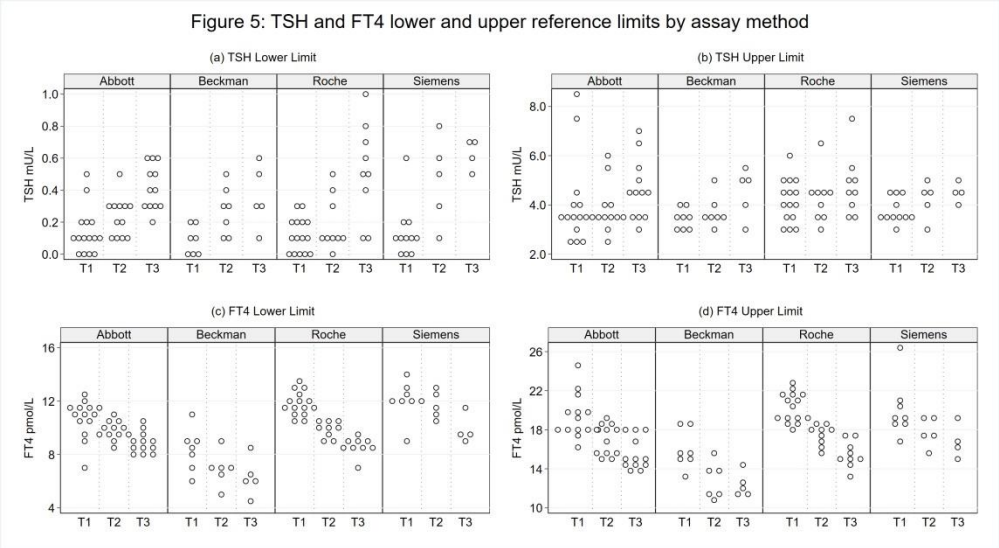


Figure 5: TSH lower and upper limits by assay method
Each circle represents the lower or upper limit reported in each study.

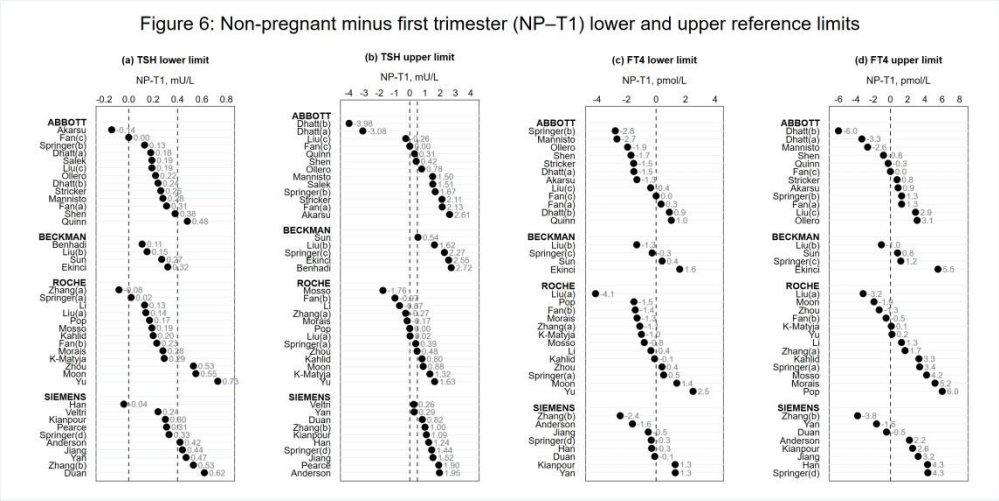
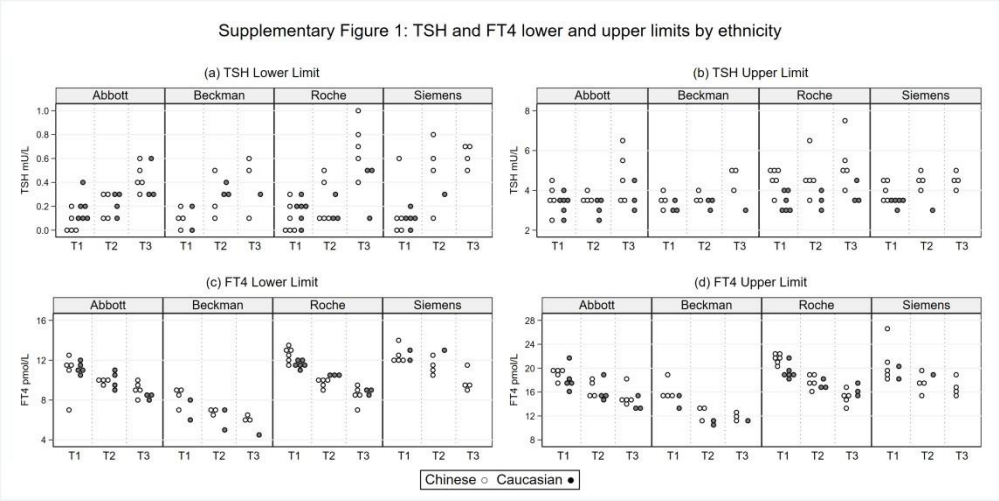


Figure 6: Non-pregnant minus 1st trimester (NP–T1) lower and upper reference limits

Legend: Circles represent data points from each study. The non-pregnant data was based on the manufacturer provided reference range for the corresponding non-pregnant population, as reported in the study. The dashed vertical lines in panel (a) (0–0.4) and panel (b) (0–0.5) represent the expected NP–T1 difference based on guideline recommendations for the lower and upper TSH limits, respectively.



Supplementary figure 1: TSH and FT4 lower and upper limits by ethnicity
Legend: Circles represent data points from each study. Studies in subjects of Chinese ethnicity are represented by white circles while studies in Caucasian subjects are represented by grey circles.

Supplementary table 1: Study Characteristics

Author, year	Country	Number	Age, years	Ethnicity	Quality score
ABBOTT					
Akarsu, 2016	Turkey	2460	31 (18-45) [a]	Turkish	7
B-Terraz, 2009	Spain	1007	31 (15-45)	White Spanish, 85%	6
Dhatt, 2006	UAE	1140	28 (16-51) [a]	Arab, 76%, Asian, 24%	8
Fan, 2016	China	647	30 (28-33)	Chinese	8
Ho, 2017	Singapore	560	NS	Chinese, Malay, Indian	8
Liu, 2017	China	947	28 (20-40)	Chinese	6
Mannisto, 2011	Finland	5043	28 (16-47)	Finnish	9
Ollero, 2019	Spain	291	33 (4.1)	Caucasian (94%)	8
Quinn, 2014	Mexico	557	25 (12-45)	Mexican	7
Salek, 2018	Czech	10592	29 (26-33)	Caucasian	7
Shen, 2014	China	1191	29 (17-47) [a]	Chinese	6
Springer, 2014	Czech	216	31 (19-42)	Caucasian	7
Stricker, 2007	Switzerland	1812	31 (18-44)	Swiss	7
Yang, 2019	China	41.634	30 (24-38)	Chinese	8
BECKMAN					
Benhadi, 2007	Netherlands	3146	31 (<20-45) [a]	Dutch, 79%	9
Ekinci, 2013	Australia	130	31 (4.7)	Not stated	9
Kim et al, 2018	Korea	417	32 (3.0)	Korean	9
Lin, 2014	China	471	29 (21-41)	Chinese	6
Liu, 2017	China	947	28 (20-40)	Chinese	6
Springer, 2014	Czech	216	31 (19-42)	Caucasian	7
Sun, 2017	China	6961	NS	Chinese	7
Wong, 2014	Canada	133	34 (25-43) [a]	Caucasian (82%)	8
Zhang, 2015	China	2743	28 (21-41)	Chinese	8
ROCHE					
Derakhshan, 2018	Sweden	2314	31 (4.8)	White Swedish, 98%	8
Donovan, 2019	Canada	416	32 (5.0)	Born in Canada	8
Fan, 2016	China	647	30 (28-33)	Chinese	8
Kahlid, 2013	Ireland	351	30 (17-45)	Caucasian (95%)	7
K-Matyja, 2017	Poland	172	35 (27-47)	Polish	9
Li, 2014	China	1024	28 (19-47) [a]	Chinese	7
Liu, 2017	China	947	28 (20-40)	Chinese	6
Moon, 2015	South Korea	465	32 (NS)	Korean	7
Morais, 2018	Brazil	225	28 (18-35)	Brazilian	7
Mosso, 2016	Chile	647	25 (6.6)	Chilean	8
Pop, 2019	Netherlands	1903	31 (3.5)	Dutch	7
Springer, 2014	Czech	216	31 (19-42)	Caucasian	7
Yu, 2010	China	301	24 (5.3)	Chinese	7
Zhang, 2016	China	957	29 (19-40)	Chinese	7
Zhou, 2018	China	20318	NS (16-48)	Chinese	6
SIEMENS					
Anderson, 2018	Denmark	10337	29 (16-51)	Born in Denmark	8
Duan, 2015	China	2433	25-35	Chinese	7
Han, 2018	China	477	20-40	Chinese	9
Jiang, 2019	China	480	31 (28-33)	Chinese	8
Kianpour, 2019	Iran	185	29 (15-45)	Iranian	8
Pearce, 2008	USA	585	33 (4.6)	White (77%)	7
Springer, 2009	Czech	4337	31 (NS)	Caucasian (99%)	7
Springer, 2014	Czech	216	31 (19-42)	Caucasian	7
Veltri, 2017	Belgium	1459	30 (5.9)	N-Afr, SSA, Caucasian	9
Yan, 2011	China	505	27 (18-40) [a]	Chinese	8
Zhang, 2019	China	805	27 (18-40)	Chinese	7

Age is presented as median (range), mean (SD), or mean (range) [a], NS, not stated, N-Afr, North African, SSA, Sub-Saharan African, T1, T2, T3, 1st, 2nd, 3rd trimester.

Supplementary table 2: Inter-method comparisons for TSH lower and upper limits

	ABBOTT	BECKMAN	ROCHE	SIEMENS	P value
T1					
TSH LRR, mU/L	0.09 (0.04–0.16)	0.06 (0.03–0.19)	0.12 (0.04–0.21)	0.06 (0.04–0.20)	0.94
TSH URR, mU/L	3.46 (2.83–4.16)	3.32 (3.09–3.83)	4.10 (3.40–4.87)	3.67 (3.34–4.49)	0.16
FT4 LRR, pmol/L	10.92 (10.30–11.65)	8.72 (7.21–9.17)	11.60 (11.10–12.35)	11.92 (11.80–12.70)	<0.001[a]
FT4 URR, pmol/L	18.69 (17.70–19.80)	15.60 (15.10–18.40)	20.31 (19.02–21.51)	19.20 (18.38–20.75)	0.002[a]
T2					
TSH LRR, mU/L	0.25 (0.13–0.34)	0.32 (0.13–0.42)	0.11 (0.07–0.31)	0.47 (0.31–0.61)	0.22
TSH URR, mU/L	3.61 (3.33–4.05)	3.31 (3.26–4.13)	4.26 (3.46–4.52)	4.46 (3.89–4.50)	0.26
FT4 LRR, pmol/L	9.73 (9.41–10.20)	6.81 (6.62–7.10)	9.90 (9.45–10.40)	11.60 (10.97–12.3)	<0.001 [a]
FT4 URR, pmol/L	17.23 (15.42–18.32)	12.41 (11.30–13.55)	17.74 (16.67–18.20)	17.60 (17.40–19.20)	0.004 [a]
T3					
TSH LRR, mU/L	0.38 (0.30–0.54)	0.34 (0.30–0.52)	0.50 (0.44–0.71)	0.60 (0.51–0.69)	0.22
TSH URR, mU/L	4.34 (3.58–5.21)	5.02 (3.86–5.06)	4.71 (3.77–5.16)	4.58 (4.37–4.77)	0.91
FT4 LRR, pmol/L	8.77 (8.18–9.31)	6.02 (5.14–6.44)	8.72 (8.70–8.96)	9.59 (9.34–10.54)	0.002 [a]
FT4 URR, pmol/L	14.90 (14.25–17.25)	11.73 (11.32–12.39)	15.20 (14.80–16.10)	16.48 (15.68–17.95)	0.007 [a]
NP-T1					
TSH LRR, mU/L	0.22 (0.18–0.28)	0.21 (0.13–0.29)	0.20 (0.14–0.29)	0.38 (0.30–0.47)	0.13
TSH URR, mU/L	0.78 (0 00–1.70)	2.30 (1.60–2.50)	0.02 (-0.27–0.80)	1.20 (0.82–1.50)	0.01 [b]
FT4 LRR, pmol/L	-1.40 (-1.80–0.17)	0.07 (-0.79–1.00)	-0.80 (-1.30–0.40)	-0.32 (-1.10–0.60)	0.39
FT4 URR, pmol/L	0.38 (-1.70–1.30)	1.00 (-0.11–3.40)	1.30 (-0.50–3.40)	2.40 (-1.00–3.80)	0.53

Figures are medians(IQR). P values derived by Kruskal Wallis test with the Bonferroni correction applied for multiple group comparisons.
a, Beckman v Abbott, Roche, or Siemens. b, Roche v Abbott, Beckman, or Siemens

PRISMA 2009 Checklist

1Section/topic	#	Checklist item	Reported on page #
TITLE			
Title	1	Identify the report as a systematic review, meta-analysis, or both.	1
ABSTRACT			
Structured summary	2	Provide a structured summary including, as applicable: background; objectives; data sources; study eligibility criteria, participants, and interventions; study appraisal and synthesis methods; results; limitations; conclusions and implications of key findings; systematic review registration number.	2
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of what is already known.	3-4
Objectives	4	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS).	4
METHODS			
Protocol and registration	5	Indicate if a review protocol exists, if and where it can be accessed (e.g., Web address), and, if available, provide registration information including registration number.	No review protocol, not registered
Eligibility criteria	6	Specify study characteristics (e.g., PICOS, length of follow-up) and report characteristics (e.g., years considered, language, publication status) used as criteria for eligibility, giving rationale.	4-5
Information sources	7	Describe all information sources (e.g., databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched.	4-5
Search	8	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated.	Appendix
Study selection	9	State the process for selecting studies (i.e., screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis).	4-5
Data collection process	10	Describe method of data extraction from reports (e.g., piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators.	5
Data items	11	List and define all variables for which data were sought (e.g., PICOS, funding sources) and any assumptions and simplifications made.	5
Risk of bias in individual studies	12	Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis.	5
Summary measures	13	State the principal summary measures (e.g., risk ratio, difference in means).	5-6

PRISMA 2009 Checklist

Synthesis of results	14	Describe the methods of handling data and combining results of studies, if done, including measures of consistency (e.g., I^2) for each meta-analysis.	5-6

Section/topic	#	Checklist item	Reported on page #
Risk of bias across studies	45 46		15 Specify any assessment of risk of bias that may affect the cumulative evidence (e.g., publication bias, selective reporting within studies).
Additional analyses			16 Describe methods of additional analyses (e.g., sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified.
RESULTS			
Study selection			17 Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram.
Study characteristics			18 For each study, present characteristics for which data were extracted (e.g., study size, PICOS,
Risk of bias within studies			
Results of individual studies			
Synthesis of results			
Risk of bias across studies			
Additional analysis			
DISCUSSION			
Summary of evidence			
Limitations			
Conclusions			
FUNDING			
Funding			

PRISMA 2009 Checklist

	follow- up period) and provide the citations.	each main outcome; consider their relevance to key groups (e.g., healthcare providers, users, and policy makers).	5
19	Present data on risk of bias of each study and, if available, any outcome level assessment (see item 12).	25 Discuss limitations at study and outcome level (e.g., risk of bias), and at review-level (e.g., incomplete retrieval of identified research, reporting bias).	5
		26 Provide a general interpretation of the results in the context of other evidence, and implications for future research.	6
20	For all outcomes considered (benefits or harms), present, for each study: (a) simple summary data for each intervention group (b) effect estimates and confidence intervals, ideally with a forest plot.	27 Describe sources of funding for the systematic review and other support (e.g., supply of data); role of funders for the systematic review.	20 (Supp Table 1)
21	Present results of each meta-analysis done, including confidence intervals and measures of consistency.		20 (Supp Table 1) Figures 2-5 21 (Supp Table 2) 20 (Supp Table 1) Figure 6
22	Present results of any assessment of risk of bias across studies (see Item 15).		10-11 12 11
23	Give results of additional analyses, if done (e.g., sensitivity or subgroup analyses, meta-regression [see Item 16]).		Not funded
24	Summarize the main findings including the strength of evidence for		

Annals of Clinical Biochemistry

PRISMA 2009 Checklist

From: Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med 6(7): e1000097. doi:10.1371/journal.pmed1000097

For more information, visit: **www.prisma-statement.org**.