The Journal of Clinical Endocrinology & Metabolism, 2022, **107**, 2925–2933
https://doi.org/10.1210/clinem/dgac425
Advance access publication 21 July 2022
Meta-Analysis

TSH and FT4 Reference Intervals in Pregnancy: A Systematic Review and Individual Participant Data Meta-Analysis

Joris A.J. Osinga,[1,2] Arash Derakhshan,[1,2] Glenn E. Palomaki,[3] Ghalia Ashoor,[4] Tuija Männistö,[5] Spyridoula Maraka,[6,7,8] Liangmiao Chen,[9] Sofie Bliddal,[10] Xuemian Lu,[9] Peter N. Taylor,[11] Tanja G.M. Vrijkotte,[12] Fang-Biao Tao,[13] Suzanne J. Brown,[14] Farkhanda Ghafoor,[15] Kris Poppe,[16] Flora Veltri,[16] Lida Chatzi,[17] Bijay Vaidya,[18] Maarten A.C. Broeren,[19] Beverley M Shields,[20] Sachiko Itoh,[21] Lorena Mosso,[22] Polina V. Popova,[23,24,25] Anna D. Anopova,[23] Reiko Kishi,[21] Ashraf Aminorroaya,[26] Maryam Kianpour,[26] Abel López-Bermejo,[27,28] Emily Oken,[29] Amna Pirzada,[30] Marina Vafeiadi,[31] Wichor M. Bramer,[32] Eila Suvanto,[33] Jun Yoshinaga,[34] Kun Huang,[35] Judit Bassols,[36] Laura Boucai,[37] Ulla Feldt-Rasmussen,[10] Elena N. Grineva,[23] Elizabeth N. Pearce,[38] Erik K. Alexander,[39] Victor J.M. Pop,[40] Scott M. Nelson,[41] John P. Walsh,[14,42] Robin P. Peeters,[1,2] Layal Chaker,[1,2,43] Kypros H. Nicolaides,[44] Mary E. D'Alton,[45] and Tim I.M. Korevaar[1,2]

[1]Department of Internal Medicine, Erasmus University Medical Center, Rotterdam, The Netherlands
[2]Academic Center for Thyroid Diseases, Erasmus University Medical Center, Rotterdam, The Netherlands
[3]Department of Pathology and Laboratory Medicine, Women & Infants Hospital and Alpert Medical School at Brown University, Providence, RI 02905, USA
[4]Harris Birthright Research Center for Fetal Medicine, King's College Hospital, London, UK
[5]Northern Finland Laboratory Center Nordlab and Medical Research Center Oulu, Oulu University Hospital and University of Oulu, Oulu, Finland
[6]Division of Endocrinology and Metabolism, University of Arkansas for Medical Sciences, Little Rock, AR 72205, USA
[7]Knowledge and Evaluation Research Unit, Division of Endocrinology, Diabetes, Metabolism and Nutrition, Department of Medicine, Mayo Clinic, Rochester, MN 55902, USA
[8]Central Arkansas Veterans Healthcare System, Little Rock, AR 72205, USA
[9]Department of Endocrinology and Rui'an Center of the Chinese-American Research Institute for Diabetic Complications, Third Affiliated Hospital of Wenzhou Medical University, Wenzhou, China
[10]Department of Medical Endocrinology and Metabolism, Copenhagen University Hospital, Rigshospitalet, and Department of Clinical Medicine, Faculty of Health and clinical Sciences, Copenhagen University, Copenhagen, Denmark
[11]Thyroid Research Group, Systems Immunity Research Institute, Cardiff University School of Medicine, Cardiff, UK
[12]Department of Public and Occupational Health, Amsterdam UMC, University of Amsterdam, Amsterdam Public Health Research Institute, Amsterdam, The Netherlands
[13]Department of Maternal, Child and Adolescent Health, School of Public Health, Anhui Medical University; Anhui Provincial Key Laboratory of Population Health & Aristogenics, Hefei, Anhui, China
[14]Department of Endocrinology and Diabetes, Sir Charles Gairdner Hospital, Nedlands, Western Australia, Australia
[15]Department of Research and Innovation, Shalamar Institute of Health Sciences, Lahore, Pakistan
[16]Endocrine Unit, Centre Hospitalier Universitaire Saint-Pierre, Université Libre de Bruxelles (ULB), Brussels, Belgium
[17]Department of Preventive Medicine, Keck School of Medicine, University of Southern California, CA 90089, USA
[18]Department of Endocrinology, Royal Devon and Exeter Hospital NHS Foundation Trust, University of Exeter Medical School, Exeter, UK
[19]Laboratory of Clinical Chemistry and Haematology, Máxima Medical Centre, Veldhoven, The Netherlands
[20]Department of Medical Statistics, University of Exeter Medical School, Exeter, UK
[21]Center for Environmental and Health Sciences, Hokkaido University, Sapporo, Japan
[22]Departments of Endocrinology, Pontificia Universidad Catolica de Chile, Santiago, Chile
[23]Institute of Endocrinology, Almazov National Medical Research Centre, Saint Petersburg, Russia
[24]Department of Internal Diseases and Endocrinology, St. Petersburg Pavlov State Medical University, Saint Petersburg, Russian Federation
[25]World-Class Research Center for Personalized Medicine, Almazov National Medical Research Centre, Saint Petersburg, Russia
[26]Isfahan Endocrine and Metabolism Research Center, Isfahan University of Medical Sciences, Isfahan, Iran
[27]Pediatric Endocrinology Research Group, Girona Biomedical Research Institute (IDIBGI), Dr. Josep Trueta Hospital, Girona, Spain
[28]Departament de Ciències Mèdiques, Universitat de Girona, Spain

Received: 28 February 2022. **Editorial Decision:** 6 July 2022. **Corrected and Typeset:** 26 July 2022

[29]Division of Chronic Disease Research Across the Lifecourse, Department of Population Medicine, Harvard Medical School, Boston, MA 02115, USA
[30]Shifa Institute of Medical Technology, Shifa International Hospital, Islamabad, Pakistan
[31]Department of Social Medicine, School of Medicine, University of Crete, Heraklion, Crete, Greece
[32]Medical Library, Erasmus University Medical Centre, Rotterdam, The Netherlands
[33]Department of Obstetrics and Gynecology and Medical Research Center Oulu, University of Oulu, Oulu, Finland
[34]Faculty of Life Sciences, Toyo University, Gunma, Japan
[35]Department of Maternal, Child and Adolescent Health, Scientific Research Center in Preventive Medicine; School of Public Health; Anhui Medical University, China
[36]Maternal-Fetal Metabolic Research Group, Girona Biomedical Research Institute (IDIBGI), Dr. Josep Trueta Hospital, Girona, Spain
[37]Department of Medicine, Division of Endocrinology, Memorial Sloan-Kettering Cancer Center, Weill Cornell University, New York, NY 10065, USA
[38]Section of Endocrinology, Diabetes, and Nutrition, Boston University School of Medicine, Boston, Massachusetts 02118, USA
[39]Division of Endocrinology, Hypertension and Diabetes, Brigham and Women's Hospital, Harvard Medical School, Boston, MA 0211, USA
[40]Department of Medical and Clinical Psychology, Tilburg University, Tilburg, The Netherlands
[41]School of Medicine, University of Glasgow, Glasgow, UK
[42]Medical School, University of Western Australia, Crawley, Western Australia, Australia
[43]Department of Epidemiology, Erasmus University Medical Center, Rotterdam, The Netherlands
[44]Department of Women and Children's Health, Faculty of Life Sciences and Medicine King's College London, London, UK; and
[45]Department of Obstetrics and Gynecology, Columbia University Irving Medical Center, New York 10032, USA

Correspondence: Joris Osinga, MD, Erasmus MC, Generation R, Wytemaweg 80, 3015 CN Rotterdam, The Netherlands. Email: j.osinga@erasmusmc.nl.

Abstract

Context: Interpretation of thyroid function tests during pregnancy is limited by the generalizability of reference intervals between cohorts due to inconsistent methodology.

Objective: (1) To provide an overview of published reference intervals for thyrotropin (TSH) and free thyroxine (FT4) in pregnancy, (2) to assess the consequences of common methodological between-study differences by combining raw data from different cohorts.

Methods: (1) Ovid MEDLINE, EMBASE, and Web of Science were searched until December 12, 2021. Studies were assessed in duplicate. (2) The individual participant data (IPD) meta-analysis was performed in participating cohorts in the Consortium on Thyroid and Pregnancy.

Results: (1) Large between-study methodological differences were identified, 11 of 102 included studies were in accordance with current guidelines; (2) 22 cohorts involving 63 198 participants were included in the meta-analysis. Not excluding thyroid peroxidase antibody–positive participants led to a rise in the upper limits of TSH in all cohorts, especially in the first (mean +17.4%; range +1.6 to +30.3%) and second trimester (mean +9.8%; range +0.6 to +32.3%). The use of the 95th percentile led to considerable changes in upper limits, varying from –10.8% to –21.8% for TSH and –1.2% to –13.2% for FT4. All other additional exclusion criteria changed reference interval cut-offs by a maximum of 3.5%. Applying these findings to the 102 studies included in the systematic review, 48 studies could be used in a clinical setting.

Conclusion: We provide an overview of clinically relevant reference intervals for TSH and FT4 in pregnancy. The results of the meta-analysis indicate that future studies can adopt a simplified study setup without additional exclusion criteria.

Key Words: thyroid, pregnancy, reference values, thyrotropin (TSH), free thyroxine (FT4)

Abbreviations: ATA, American Thyroid Association; BMI, body mass index; FT4, free thyroxine; IPD, individual patient data; IVF, in vitro fertilization; SGA, small for gestational age; TgAb, thyroglobulin antibody; TPOAb, thyroid peroxidase antibody; TSH, thyrotropin.

Adequate thyroid hormone availability during pregnancy is important for an uncomplicated pregnancy as well as optimal fetal growth and development. Thyroid function test abnormalities during pregnancy are associated with a higher risk of adverse pregnancy and child outcomes (1-4). However, identifying thyroid function abnormalities during pregnancy is complicated by changes in maternal physiology. Furthermore, there is no universal reference interval for thyrotropin (TSH) or free thyroxine (FT4) during pregnancy due to considerable differences between assays as well as population characteristics (5-7). Current guidelines from international thyroid or endocrine societies, including the most recent 2017 guidelines by the American Thyroid Association (ATA), recommend the use of population- and trimester-specific TSH and FT4 reference intervals as the gold standard, calculated in a population with no known thyroid disease, optimal iodine status, and negative thyroid peroxidase antibody (TPOAb) status (4, 8, 9). However, for many laboratories these are unavailable because calculating reference intervals from a local reference population is often not feasible. Another option recently provided in the ATA guidelines either is to use a fixed cut-off for the upper limit of TSH of 4.0 mU/L or to subtract 0.5 mU/L from the nonpregnancy upper reference limit of TSH in the first trimester (4). While the method of using a fixed upper limit for TSH may lead to considerable under- and overdiagnosis compared with the gold standard because of interpopulation and interassay differences (10), the method of subtracting an absolute number from the nonpregnancy reference interval has not been thoroughly researched.

The most recent addition to the ATA guidelines is the option to adopt reference intervals that were calculated in a center with a similar population and using the same assay, which is a step in between the gold standard and fixed upper TSH limit approach (4). However, identification of adoptable TSH and FT4 reference intervals is cumbersome due to a lack of overview of all published data regarding thyroid hormone reference intervals. Moreover, large methodological differences exist between studies as a result of new insights and changing guidelines (4, 8, 11). One example of this is the use of additional exclusion criteria on top of those recommended by the current ATA guidelines, most of which remain of unknown significance, such as thyroglobulin antibody (TgAb) positivity, conception by in vitro fertilization (IVF), pregnancy complications, and characteristics including preexisting diabetes mellitus, hypertension, aberrant body mass index (BMI), and active smoking. Although some of these factors are determinants of TSH and FT4 concentrations, only some, but not all, studies show that exclusion of women

according to these determinants affects TSH and FT4 reference intervals (12-17). Another example of between-study differences in methodology is the calculation of the reference intervals. Some studies define reference intervals by the 5th and 95th percentiles, whereas in routine laboratory practice, the 2.5th to 97.5th percentiles are typically used (18). While the methods of many studies are not in line with current international guidelines from thyroid or endocrine societies, it remains unknown to what extent this affects the generalizability of the calculated reference intervals.

To aid clinicians in adopting suitable reference intervals for their center and for the purpose of future research, the aims of our study were (1) to provide an overview of methodologically valid and clinically useful published gestational TSH and FT4 reference intervals; and (2) to perform a meta-analysis of individual participant data (IPD) from a consortium of cohorts to add gestational TSH and FT4 reference intervals to the literature, and utilize these data to assess the validity and clinical relevance of using additional exclusion criteria.

Materials and Methods

For the first study aim, we performed a systematic literature search to identify all available studies on TSH and/or FT4 reference intervals during pregnancy (Fig. 1). To address the second aim, we performed an IPD meta-analysis within the Consortium on Thyroid and Pregnancy.

Systematic Review

We searched Ovid MEDLINE, EMBASE and Web of Science databases from inception to December 12, 2021, without language restrictions (document 1 (24)). All studies with data on TSH and/or FT4 in an unselected (ie, without selection based on health indicators) population of pregnant women were included. For the main figures (Figs. 2-4), we reported reference intervals calculated using the 2.5th to 97.5th percentile and after exclusion of participants with pre-pregnancy thyroid disease, thyroid hormone altering medication use and/or TPOAb-positivity. We then excluded studies in which additional exclusion criteria of unknown significance were applied for selecting the reference population (eg, the exclusion of participants with non-thyroidal autoimmune disease). We also excluded studies in populations with severe iodine deficiency (defined as urinary iodine secretion <50 µg/L (19)) as assessed in the cohort or, if unavailable, as reported by the WHO or regional studies (20, 21). Mild-to-moderate iodine deficiency (50-149 µg/L) was not a criterion for exclusion, since TSH and FT4 reference intervals do not meaningfully differ from iodine-sufficient reference intervals (22, 23). Finally, studies in which reference intervals were not reported or in which the exact methods for reference interval determinations could not be extracted were excluded. Studies which published reference intervals which covered less than 2 gestational weeks were not included in the final overview, as these were deemed to be less useful to clinicians. If additional exclusion criteria for the reference population were used that could be assessed in the meta-analyses, studies were included if the additional exclusion criteria led to a maximum of 5% variability around the reference limit as compared to the gold standard. If additional exclusion criteria could not be assessed in the meta-analysis, or if the criteria led to more than 5% variability around the reference limits, the studies were excluded from the final overview (Figs. 2-4).

Possible studies for inclusion were independently assessed for suitability in duplicate (title and abstracts: A.D. and T.K.; full texts: A.D. and J.O.), and any disagreement was resolved by discussion with a third author (J.O. or T.K., respectively). Additional records were retrieved through citation searching. The study protocol was preregistered which can be found elsewhere along with protocol violations.

Individual Participant Data Meta-analysis

Eligible studies were those that participated in the Consortium on Thyroid and Pregnancy (https://www.consortiumthyroidpregnancy.org), an international research collaboration that aims to study gestational thyroid (dys) function physiology, determinants, and clinical risk profiles. Information on regional iodine status and assays used in each cohort can be found elsewhere (Tables 1 and 2 (24)). Reference intervals were calculated as the 2.5th to 97.5th interval after excluding participants with known thyroid disease, use of thyroid (interfering) medication, multiple gestation, and/or TPOAb positivity. To quantify the effects of between-study methodological differences of TSH and FT4 reference interval calculations, reference intervals were also calculated using 9 different methodologies in addition to the methodology described

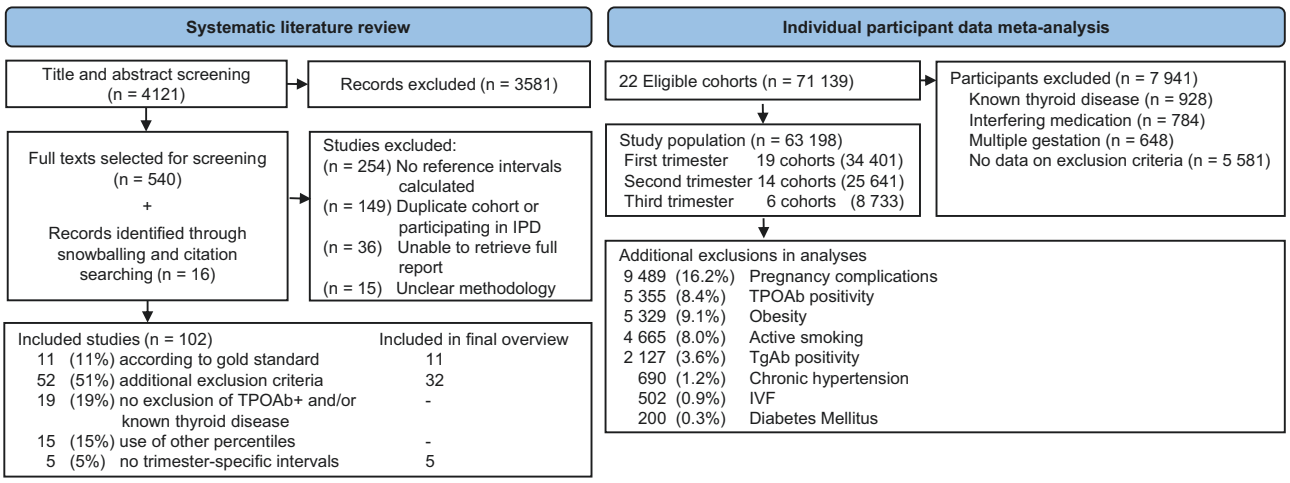

Figure 1. Study selection flowchart.

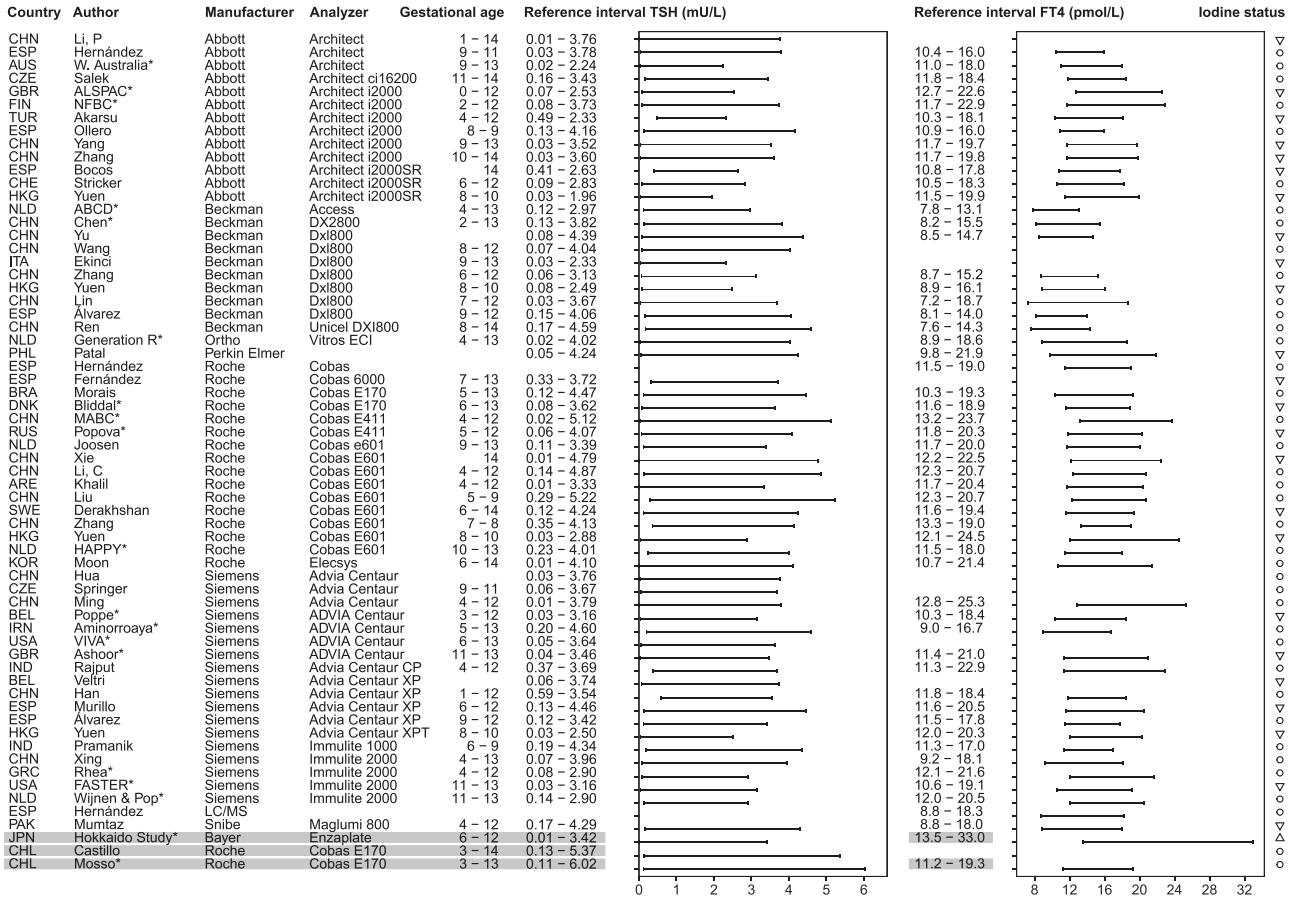

Figure 2. Overview studies systematic review trimester 1. *Reference intervals calculated with individual participant data in consortium. ○, iodine sufficiency; ▽, mild-to-moderate iodine deficiency; △, excessive iodine status. Reference intervals calculated in cohorts with fluctuating or excessive iodine status are listed in gray.

above: (1) without excluding TPOAb-positive participants, (2) using the 5th to 95th percentiles, and using additional exclusion criteria defined as (3) prepregnancy diabetes mellitus, (4) essential hypertension, (5) obesity, defined as a BMI > 30 kg/m^2 at time of the first study visit, (6) active smoking, (7) conception by IVF, (8) exclusion of TgAb positivity, or (9) pregnancy complications, defined as gestational diabetes, gestational hypertension, pre-eclampsia, preterm birth, and/or small for gestational age (SGA). We defined the trimesters as 0 to 13 weeks, 13 to 27 weeks, and >27 weeks of gestation.

Statistical Analyses

Trimester-specific reference intervals were calculated per cohort including participants with data on exclusion criteria. For cohorts containing participants with repeated measurements, we used the first available sample for each trimester. Outliers were removed if values were inaccurate (eg, TSH or FT4 above or below detection limit). All analyses were performed using IBM SPSS Statistics for Windows (version 25.0) and R 4.1.1 for Windows.

Results

Systematic Review (Aim 1)

We identified 4121 published reports, of which 540 were eligible for inclusion based on title and abstract screening. After reading full texts and adding 16 articles identified through snowballing, 102 articles were included in the systematic review (Fig. 1). Out of all included studies, 11 (11%) reported reference intervals calculated in accordance with the current ATA guidelines, 52 (51%) studies used additional exclusion criteria for the selection of the reference population (ie, participants with acute or chronic illnesses, pre-existing hypertension or diabetes mellitus, autoimmune disease, and some or all pregnancy complications defined above), 15 (15%) used percentile cut-offs other than the 2.5th to 97.5th percentiles (mostly the 5th to 95th percentiles (8)) (Fig. 1). An overview of selected TSH and FT4 reference intervals is presented in Figs. 2-4, which either adhere to the ATA guidelines (11 studies), or use additional exclusion criteria which had less than 5% effect on the reference limits according to our meta-analysis (32 studies, see below). Data (including future updates) can be downloaded from https://www.consortiumthyroidpregnancy.org/referenceintervals.

Individual Participant Data Meta-analysis (Aim 2)

At time of analysis, 71 139 participants in 22 cohorts participating in the Consortium on Thyroid and Pregnancy were eligible for inclusion. After exclusions, the final study population comprised 63 198 women (Fig. 1). Between all included cohorts, the upper limit for TSH ranged from 2.24 to 6.02 mU/L in the first trimester, from 2.67 to 6.15 mU/L in the second trimester and from 3.03 to 6.13 mU/L in the third trimester (third trimester data from n = 2 cohorts). The lower FT4 limit ranged from 7.8 to 13.2 pmol/L (0.61-1.03 ng/dL)

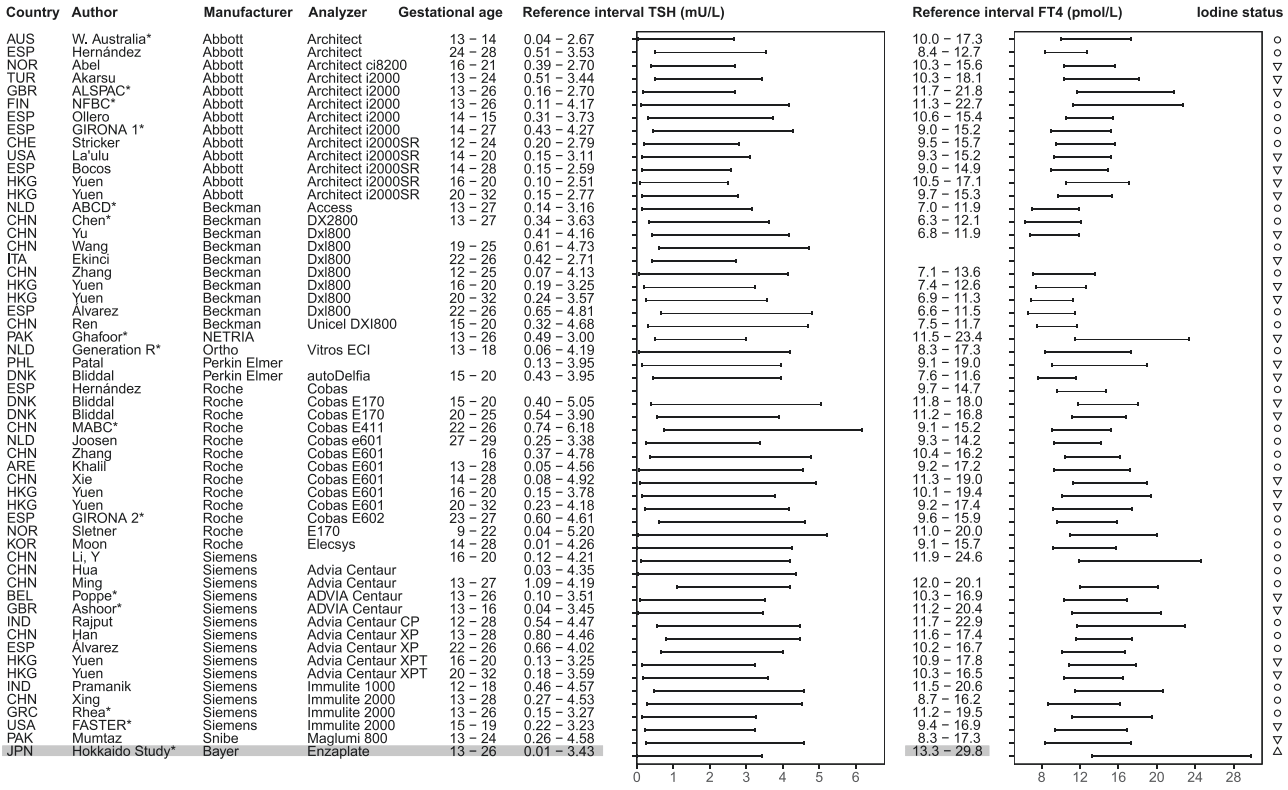

Figure 3. Overview studies systematic review trimester 2. *Reference intervals calculated with individual participant data in consortium. ○, iodine sufficiency; ▽, mild-to-moderate iodine deficiency; △, excessive iodine status. Reference intervals calculated in cohorts with fluctuating or excessive iodine status are listed in gray.

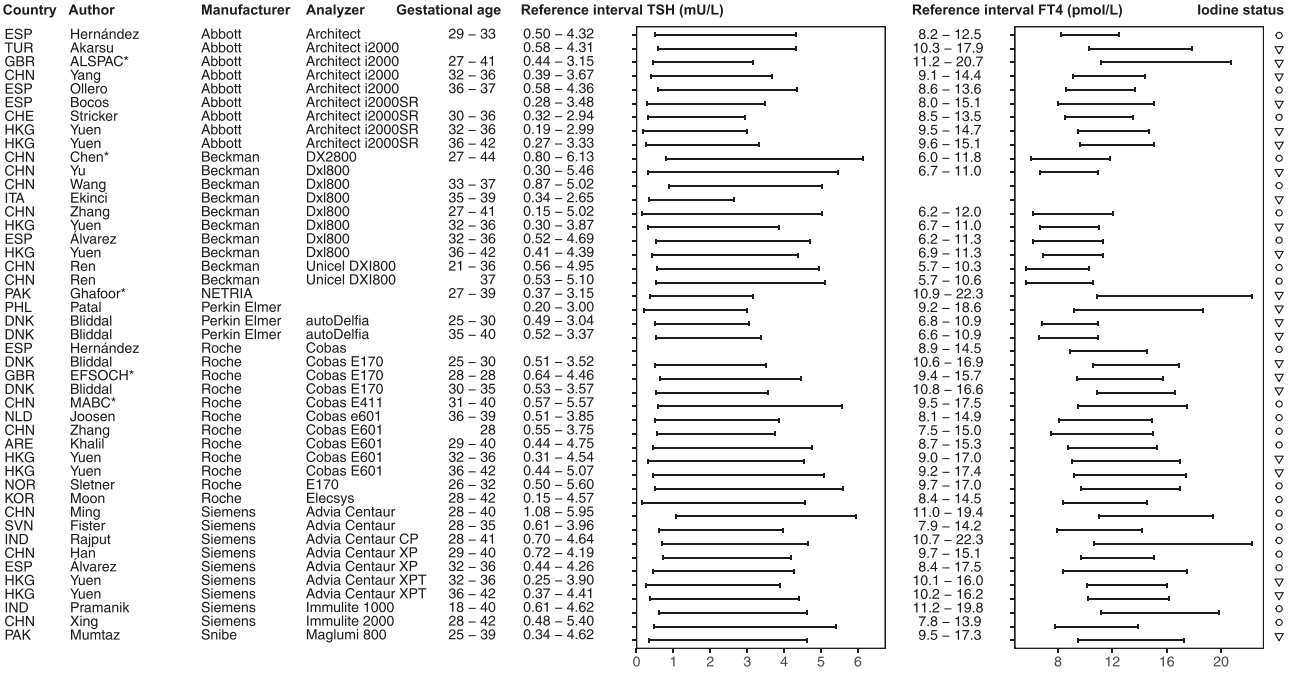

Figure 4. Overview studies systematic review trimester 3. *Reference intervals calculated in individual participant data in consortium; ○, iodine sufficiency; ▽, mild-to-moderate iodine deficiency.

in the first trimester, from 7.1 to 11.5 pmol/L (0.55-0.89 ng/dL) in the second trimester, and from 9.5 to 11.1 pmol/L (0.74-0.86 ng/dL) in the third trimester (third trimester data from n = 2 cohorts). The upper limit of FT4 ranged from 13.2 to 23.8 pmol/L (1.03-1.85 ng/dL) in the first trimester, from 11.8 to 22.7 pmol/L (0.92-1.76 ng/dL) in the second trimester, and 17.5 to 19.6 pmol/L (1.36-1.52 ng/dL) in the third trimester (third trimester data from n = 2 cohorts). Cohort-specific reference intervals can be found elsewhere (Tables 4-9 (24)).

Compared with the 2.5th to 97.5th population-based reference intervals, the use of the 5th to 95th percentiles resulted in an upper limit of TSH that was 0.63 mU/L (range 0.35-1.22) lower in the first trimester, 0.65 mU/L (range 0.29-0.83) lower in the second trimester, and 0.73 mU/L (range 0.70-0.73) lower in the third trimester (Fig. 5; Tables 4-6 (24)). For FT4, the lower limits were 0.3 to 1.3 pmol/L (0.02-0.10 ng/dL) higher in the first trimester, 0.0 to +1.0 pmol/L (0.0-0.08 ng/dL) in the second trimester, and +0.5 to +1.2 pmol/L (0.04-0.09 ng/dL) in the third trimester (Fig. 6). The upper limit of FT4 was –0.4 to –2.2 pmol/L lower among cohorts in the first trimester, –0.2 to –3.0 pmol/L in the second trimester, and –0.4 to –0.7 pmol/L in the third trimester (Fig. 7; Tables 7-9 (24)).

Not excluding TPOAb-positive participants led to an increase in the upper limit of TSH in all cohorts, with a mean increase of 0.65 mU/L (range 0.05-1.34) in the first trimester, 0.42 mU/L (range 0.02-0.83) in the second trimester, and 0.14 mU/L (range 0.08-0.65) in the third trimester (Fig. 5; Tables 4-6 (24)). No meaningful changes were observed in the lower and upper reference limits of FT4 (Figs. 6 and 7; Tables 7-9 (24)).

All other additional exclusion criteria, namely pre-existing diabetes mellitus or essential hypertension, obesity, active smoking, conception by IVF, TgAb positivity or any pregnancy complication, led to less than 3.5% variation around the upper limits of TSH and the lower and upper limits of FT4, without a clear trend toward an increase or decrease in the reference limits (Figs. 5-7).

Discussion

In this study, we present an overview of reference intervals for TSH and FT4 during pregnancy, exhibiting widely differing absolute values for TSH and FT4 reference intervals and considerable heterogeneity in the methods used for reference interval calculations. Furthermore, in an IPD meta-analysis, we showed that inclusion of TPOAb-positive women was associated with a considerable increase in the upper limits for TSH but not FT4 reference intervals, while the use of a 5th to 95th percentile range led to a considerable decrease in the upper limit for both TSH and FT4 reference intervals as well as an increase in the lower limit for the FT4 interval. On the other hand, none of the studied additional reference population selection criteria meaningfully affected TSH and FT4 reference intervals.

Systematic Review (Aim 1)

In our systematic review, we identified that out of all published studies reporting reference intervals in the literature, at least 48 out of 102 studies can be implemented in clinical settings. These studies published reference intervals which either adhere to the most recent ATA guidelines or used additional exclusion criteria which resulted in less than 5% change of the reference limit, as assessed in our meta-analysis. The considerable variation in methods is partly explained by progressive insight and changing guidelines over the years. However, well before this time, guidelines by various clinical chemistry societies for reference interval determinations were already

Additional exclusions	First trimester N cohorts (participants)	First trimester Pooled difference in upper limit[1] (mU/L)	First trimester Range of difference in cohorts	Second trimester N cohorts (participants)	Second trimester Pooled difference in upper limit[1] (mU/L)	Second trimester Range of difference in cohorts	Third trimester N cohorts (participants)	Third trimester Pooled difference in upper limit[1] (mU/L)	Third trimester Range of difference in cohorts
Gold standard[2]	11 (17.924)			9 (10.929)			2 (2.957)		
Without TPOAb+ exclusion	11 (19.723)	0.6	+0.05 to +1.34	9 (11.726)	0.4	+0.02 to +0.83	2 (3.216)	0.1	+0.08 to +0.65
5th - 95th percentile	11 (17.924)	-0.6	-1.22 to -0.35	9 (10.929)	-0.7	-0.83 to -0.29	2 (2.957)	-0.7	-0.73 to -0.70
Diabetes mellitus[3]	11 (17.795)	0.0	-0.01 to +0.01	9 (10.829)	0.0	-0.06 to +0.02	2 (2.950)	0.0	0 to 0
Essential hypertension	11 (17.521)	0.0	-0.08 to +0.07	9 (10.775)	0.0	-0.02 to +0.03	2 (2.943)	0.0	0 to +0.02
Obesity[4]	11 (16.467)	0.0	-0.15 to +0.11	9 (10.016)	0.0	-0.08 to +0.08	2 (2.918)	0.0	0 to +0.02
Active smoking	11 (16.078)	0.0	-0.06 to +0.11	9 (9.858)	0.0	-0.24 to +0.14	2 (2.880)	0.0	0 to +0.06
Gold standard[2]	17 (20.121)			13 (19.002)					
IVF[5]	17 (19.811)	0.0	-0.03 to +0.05	13 (18.826)	0.0	-0.02 to +0.04			
Gold standard[2]	10 (21.306)			8 (16.555)			2 (7.044)		
TgAb+	10 (20.264)	-0.1	-0.21 to +0.08	8 (15.855)	-0.1	-0.19 to +0.10	2 (6.808)	0.0	-0.08 to 0
Gold standard[2]	17 (22.533)			13 (20.306)			4 (7.260)		
Pregnancy complications[6]	17 (18.986)	0.0	-0.51 to +0.29	13 (17.303)	0.0	-0.02 to +0.18	4 (6.370)	0.0	-0.04 to +0.01

Figure 5. Results meta-analyses trimester 1. [1]Defined as the weighted average change in reference limits across cohorts. [2]Calculated using 2.5th to 97.5th percentiles, excluding prepregnancy thyroid disease, use of thyroid hormone–altering medication, and TPOAb positivity; the total number of participants in subgroups differs based on availability of data on additional exclusion criteria. [3]Defined as prepregnancy diabetes mellitus. [4]Defined as BMI > 30 kg/m^2 at time of intake. [5]Pregnancy by in vitro fertilization. [6]Defined as gestational diabetes, gestational hypertension, pre-eclampsia, preterm birth, and/or small for gestational age.

Additional exclusions	First trimester N cohorts (participants)	First trimester Pooled difference in lower limit[1] (pmol/L)	First trimester Range of difference in cohorts	Second trimester N cohorts (participants)	Second trimester Pooled difference in lower limit[1] (pmol/L)	Second trimester Range of difference in cohorts	Third trimester N cohorts (participants)	Third trimester Pooled difference in lower limit[1] (pmol/L)	Third trimester Range of difference in cohorts
Gold standard[2]	10 (17.435)			9 (10.962)			2 (2.955)		
Without TPOAb+ exclusion	10 (19.153)	-0.1	-0.3 to 0	9 (11.768)	0.0	-0.1 to 0	2 (3.215)	0.0	0 to 0
5th - 95th percentile	10 (17.435)	0.5	+0.3 to +1.3	9 (10.962)	0.5	0 to +1.0	2 (2.955)	0.6	+0.5 to +1.2
Diabetes mellitus[3]	10 (17.305)	0.0	0 to 0	9 (10.862)	0.0		2 (2.948)	0.0	0 to 0
Essential hypertension	10 (17.043)	0.0	0 to 0	9 (10.807)	0.0	-0.1 to +0.1	2 (2.942)	0.0	0 to 0
Obesity[4]	10 (16.038)	0.0	0 to +0.5	9 (10.046)	0.1	-0.1 to +0.2	2 (2.916)	0.0	0 to 0
Active smoking	10 (15.681)	0.0	0 to +0.1	9 (9.889)	0.0	-0.1 to +0.1	2 (2.879)	0.0	-0.2 to 0
Gold standard[2]	10 (20.238)			8 (19.051)					
IVF[5]	10 (19.927)	0.0	-0.3 to 0	8 (18.875)	0.0	-0.1 to 0			
Gold standard[2]	10 (21.370)			8 (16.753)			2 (7.177)		
TgAb+	10 (20.325)	0.0	-0.1 to +0.1	8 (15.969)	0.0	-0.1 to +0.1	2 (6.884)	0.0	0 to 0
Gold standard[2]	11 (21.956)			8 (20.320)			2 (7.260)		
Pregnancy complications[6]	11 (18.512)	0.0	0 to +0.3	8 (17.312)	0.1	-0.1 to +0.1	2 (6.370)	0.0	0 to 0

Figure 6. Results meta-analyses trimester 2. [1]Defined as the weighted average change in reference limits across cohorts. [2]Calculated using 2.5th to 97.5th percentiles, excluding pre-pregnancy thyroid disease, use of thyroid hormone altering medication, and TPOAb-positivity; the total number of participants in subgroups differs based on availability of data on additional exclusion criteria. [3]Defined as prepregnancy diabetes mellitus. [4]Defined as BMI >30 kg/m^2 at time of intake. [5]Pregnancy by in vitro fertilization. [6]Defined as gestational diabetes, gestational hypertension, pre-eclampsia, preterm birth, and/or small for gestational age.

Additional exclusions	First trimester N cohorts (participants)	First trimester Pooled difference in upper limit[1] (pmol/L)	First trimester Range of difference in cohorts	Second trimester N cohorts (participants)	Second trimester Pooled difference in upper limit[1] (pmol/L)	Second trimester Range of difference in cohorts	Third trimester N cohorts (participants)	Third trimester Pooled difference in upper limit[1] (pmol/L)	Third trimester Range of difference in cohorts
Gold standard[2]	10 (17.435)			9 (10.962)			2 (2.955)		
Without TPOAb+ exclusion	10 (19.153)	0.1	-0.3 to +0.8	9 (11.768)	0.0	-0.2 to +0.2	2 (3.215)	0.1	+0.1 to +0.1
5th - 95th percentile	10 (17.435)	-1.6	-2.2 to -0.4	9 (10.962)	-1.2	-3.0 to -0.2	2 (2.955)	-0.7	-0.7 to -0.4
Diabetes mellitus[3]	10 (17.305)	0.0	-0.1 to 0	9 (10.862)	0.0	0 to +0.1	2 (2.948)	0.0	0 to 0
Essential hypertension	10 (17.043)	0.0	0 to +0.2	9 (10.807)	0.0	-0.3 to +0.2	2 (2.942)	0.0	0 to +0.1
Obesity[4]	10 (16.038)	0.1	0 to +0.4	9 (10.046)	0.2	-0.2 to +1.1	2 (2.916)	0.0	0 to 0
Active smoking	10 (15.681)	0.0	-0.1 to +0.9	9 (9.889)	0.0	-0.9 to +0.2	2 (2.879)	0.0	0 to 0
Gold standard[2]	10 (20.238)			8 (19.051)					
IVF[5]	10 (19.927)	0.0	-0.1 to +0.6	8 (18.875)	0.0	-0.2 to +0.1			
Gold standard[2]	10 (21.370)			8 (16.753)			2 (7.177)		
TgAb+	10 (20.325)	0.0	-0.2 to +0.4	8 (15.969)	0.0	-0.3 to +0.1	2 (6.884)	0.0	0 to 0
Gold standard[2]	11 (21.956)			8 (20.320)			2 (7.260)		
Pregnancy complications[6]	11 (18.512)	0.1	-0.2 to +0.9	8 (17.312)	0.0	-0.2 to +1.1	2 (6.370)	0.1	0 to +0.1

Figure 7. Results meta-analyses trimester 3. [1]Defined as the weighted average change in reference limits across cohorts. [2]Calculated using 2.5th to 97.5th percentiles, excluding prepregnancy thyroid disease, use of thyroid hormone altering medication, and TPOAb-positivity; the total number of participants in subgroups differs based on availability of data on additional exclusion criteria. [3]Defined as prepregnancy diabetes mellitus. [4]Defined as BMI >30 kg/m^2 at time of intake. [5]Pregnancy by in vitro fertilization. [6]Defined as gestational diabetes, gestational hypertension, pre-eclampsia, preterm birth, and/or small for gestational age.

advising similar selection criteria, including but not limited to excluding TPOAb positivity, prepregnancy thyroid disease, thyroid-interfering medication use, and the use of a 2.5th to 97.5th percentile range (11). We identified that these key selection criteria were not met for 37% (20% lack of advised exclusion criteria +17% use of different percentiles) of all published studies, potentially causing relevant under- and overdiagnosis of gestational thyroid disease. This is particularly important in light of the recommendation in the current ATA guidelines that TSH and FT4 reference intervals can be adopted from similar patient populations using similar TSH assays. Interestingly, we also identified that 54% of all studies in the published literature used additional exclusion criteria for the selection of a reference population. To our knowledge, this is the first study to systematically quantify the effects of such exclusions on TSH and FT4 reference intervals during pregnancy in a multicohort setting, showing no meaningful effects besides decreasing the study population sample size. On the one hand, these results validate currently available studies that have used these additional exclusion criteria. On the other hand, our results also caution against the use of such methods because a decrease in the size of a study population due to unnecessary exclusions can affect the precision of reference interval calculations.

A notable observation from the results of the systematic review is the large variation in reference limits within the same assays, as it can be observed in Figs. 2-4. Furthermore, while no formal test for differences between reference intervals has been carried out, it can be observed that the Beckman assay has relatively lower FT4 reference intervals than the other assays. This demonstrates the impact of preanalytical factors as well as determinants of thyroid function tests on the variation between populations, and the importance of locally derived reference intervals for TSH and FT4 in pregnancy.

Regarding iodine status, iodine deficiency is associated with a higher risk of thyroid dysfunction, and inclusion of participants with iodine deficiency might lead to unreliable reference intervals. However, 2 recently published studies show no significant differences in reference limits when including mild to moderate iodine–insufficient participants (urinary iodine secretion 50-149 µg/L), which calls for a more lenient recommendation regarding the iodine status of the reference population (22, 23). On the other hand, chronic excessive iodine intake may lead to hypothyroidism, while a sudden increase in iodine intake may lead to transient hyperthyroidism (25). For instance, the Chilean population has a history of chronic high iodine intake following the iodine fortification program in 1979, which resulted in a reduction of fortification in 2000 (26). These fluctuations might explain the relatively high upper reference limits for TSH in the Chilean cohort in the IPD meta-analysis and the Chilean study included in the systematic review, although the inclusion periods were well over a decade past the reduction of the fortification dose. One Japanese cohort in the consortium had excessive iodine intake at the time of inclusion, which may explain the relatively high reference limits of FT4. Iodine status, in particular severe deficiency or excessive intake, influences TSH and FT4 concentrations considerably, and thyroid function test measurements are highly dependent on the timing and amount of iodine intake. Reference intervals calculated in a region known for an excessive or fluctuating iodine status are listed in gray in the final overview (Figs. 2-4) and should be interpreted with care. In addition, it is also important to realize that geographic determinants of thyroid function tests other than iodine status, such as ethnicity or genetics, exposure to endocrine disrupting chemicals, and harmonization of local laboratories, affect the variation in TSH and FT4 reference intervals. These should be taken into account when adopting reference intervals, most pragmatically by using locally derived reference intervals.

Individual Participant Data Meta-analysis (Aim 2)

One of the most prominent quantifications of the IPD analyses we performed is that not excluding TPOAb-positive women when calculating reference intervals considerably increases the upper reference limit of TSH, especially in the first (+0.65 mU/L) and second trimester (+0.42 mU/L). The larger effect in the first trimester coincides with the physiological peak of human chorionic gonadotropin and the higher prevalence of TPOAb positivity in early pregnancy. Between the cohorts in this study, 3% to 18% of all participants are TPOAb-positive, which is associated with an impaired thyroidal response to human chorionic gonadotropin stimulation and a higher risk of adverse pregnancy outcomes (27-29). It is noteworthy that 18% of all studies in the literature did not exclude TPOAb-positive participants. This specifically leads to underdiagnosis of TPOAb-positive subclinical hypothyroidism, for which levothyroxine treatment is recommended (4). The subsequent potential risk of under treatment emphasizes the importance of excluding TPOAb-positive women from the reference population. The additional exclusion of TgAb positivity on top of TPOAb positivity only led to a 3.5% mean decrease of the upper reference limit of TSH. Studies on the risk of adverse pregnancy outcomes in TgAb-positive women have been inconsistent (30-32), but the

lack of a relevant effect (defined as at least 5% variability around the reference limit) of excluding this group on the reference limits is in line with the results of a recent prospective Swedish cohort study (12). This was further established in a recent study by the Consortium on Thyroid and Pregnancy wherein TPOAbs were shown to have a stronger association with higher TSH concentrations than TgAbs, and where the association with TgAb concentrations lost significance when adjusting for TPOAb concentration (33). The results of the current meta-analysis confirm the validity of reference intervals from previous studies in which TgAb-positive women were additionally excluded, but also implicate no added value from the additional efforts and costs related to measurement of TgAbs for studies with the purpose of defining TSH and FT4 reference intervals during pregnancy.

Another important finding of the IPD is the substantial change in reference limits when using the 5th to 95th percentile range rather than the internationally advised 2.5th to 97.5th percentile range (18). This effect was consistent through all trimesters for the upper limit of TSH (for the first, second, and third trimester: –0.63, –0.65, and –0.73 mU/L, respectively), the lower limit of FT4 (+0.5, +0.5, +0.6 pmol/L, or +0.04, +0.04, +0.05 ng/dL) and in a declining trend for the upper limit of FT4 (–1.6, –1.2, –0.7 pmol/L, or –0.12, –0.09, –0.05 ng/dL). Especially TSH, and to a lesser extent FT4, concentrations in pregnancy follow a right skewed distribution, which explains the relatively larger absolute effect of percentile shifts on the upper reference limit. Narrowing the reference intervals by using reference cut-offs above the 2.5th and below the 97.5th percentile (eg, the 5th to 95th percentile as assessed in this study) will result in more women diagnosed with thyroid disease in clinical settings purely from a mathematical point of view. Moreover, since predominantly the upper rather than the lower reference limits of TSH and FT4 are affected, specifically more women will be diagnosed with subclinical hypothyroidism and overt hyperthyroidism. To our knowledge, no studies have systematically assessed whether women with TSH or FT4 concentrations between the 95th and 97.5th percentile are also at increased risk of adverse pregnancy events and whether they might benefit from treatment. Implementing reference intervals based on the 5th to 95th percentiles may therefore lead to considerable overdiagnosis and overtreatment in the clinical setting.

Interestingly, all other additional exclusion criteria (eg, preexisting diabetes mellitus, essential hypertension, obesity, active smoking, IVF, and all pregnancy complications) assessed in the IPD did not meaningfully change the reference limits. This is despite opposite results in some single-center studies and considering that most factors are known determinants of mean TSH and/or FT4 concentrations in a nonpregnant population (34). One interesting example of this is BMI. While a higher BMI is positively associated with TSH, and negatively with FT4 in nonpregnant individuals (35), the results in pregnancy are inconclusive (36-38), and the effect of excluding obese participants for calculating reference intervals is largely unknown. Two cohort studies from Finland and China demonstrated a positive correlation between BMI and TSH in pregnancy (16, 38); however, while a higher BMI was associated with a higher upper limit of TSH in the Finnish cohort, in the Chinese cohort a higher BMI was associated with a lower reference limit of TSH. The results of the current study show that additional exclusion of women with pregnancy complications on average led to less than 1.5% (range between cohorts –8% to +7%) change of TSH and FT4 reference limits. While the lack of clinically meaningful changes with these additional exclusions does not affect the validity of studies that incorporated them in the past, future studies should avoid these exclusion criteria as they may limit precision of study results by decreasing the total number of participants in a study population.

For this study, we were able to summarize key features of TSH and FT4 reference intervals in pregnant women and further elucidate and quantify the relevance of commonly encountered deviations from guideline methodology standards using a large dataset consisting of worldwide prospective cohort studies. The interpretation of the results of our systematic literature overview is limited by the details communicated in the original reports. Underreporting of relevant details in original work may have affected the generalizability of our results to the literature as a whole and may have resulted in exclusion of studies which published valid reference intervals. Furthermore, analyses focusing on exclusion criteria which are uncommon in a population, such as pre-existing diabetes in young women, are not likely to meaningfully change the reference limits, because only a small number of participants are excluded. These results may not be directly generalizable to populations which have a substantially higher prevalence of the analyzed exclusion criteria. Moreover, this study was limited by data available in the included cohorts and therefore it was not possible to assess the impact of all additional exclusion criteria found in the literature, for instance the additional exclusion of participants with a nonthyroidal autoimmune disease.

In conclusion, this systematic review and IPD meta-analysis provides an overview with available reference intervals which can be adopted in clinical settings, taking population and assay similarity into account. The importance of excluding TPOAb-positive participants and the use of proper percentiles as defined by international guidelines is emphasized, and future studies aiming to calculate reference intervals can adopt a simplified study setup in terms of exclusion criteria for the reference population.

Funding

Netherlands Organization for Scientific Research (grant 401.16.020).

Disclosures

P.T. reports a travel grant from Society for Endocrinology (leadership development award). E.O. reports grants from the National Institutes of Health. E.N.G. received speaker's fees and payment for expert testimony from Merck and consulting fees from Brunel Rus. T.G.M.V. reports grants from the Netherlands Organization for Health Research and Development. E.K.A. reports consultancy with Roche Diagnostics. L.C. received travel support by Pfizer. S.M.N. has received consultancy, speakers' fees, or travel support from Access Fertility, Beckman Coulter, Ferring Pharmaceuticals, Merck, Modern Fertility, Roche Diagnostics, and The Fertility Partnership. S.M.N. also reports payments for medical–legal work and investment in The Fertility Partnership. T.I.M.K. reports lectureship fees from Berlin-Chemie, Goodlife Healthcare, Institut Biochimique SA, Merck, and Quidel. U.F.R.'s research salary was sponsored by an unrestricted grant from Kirsten and Freddy Johansen's Fund. All other authors declare no competing interests.

Data Availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

1. Negro R, Stagnaro-Green A. Diagnosis and management of subclinical hypothyroidism in pregnancy. *BMJ*. 2014;349:g4929.
2. Korevaar TIM, *et al*; Consortium on Thyroid and Pregnancy—Study Group on Preterm Birth. Association of thyroid function test abnormalities and thyroid autoimmunity with preterm birth: a systematic review and meta-analysis. *JAMA*. 2019;322(7):632-641.
3. Derakhshan A, Peeters RP, Taylor PN, *et al*. Association of maternal thyroid function with birthweight: a systematic review and individual-participant data meta-analysis. *Lancet Diabetes Endocrinol*. 2020;8(6):501-510.
4. 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(3):315-389.
5. Hershman JM. The role of human chorionic gonadotropin as a thyroid stimulator in normal pregnancy. *J Clin Endocr Metab*. 2008;93(9):3305-3306.
6. Korevaar TIM, Medici M, Visser TJ, Peeters RP. Thyroid disease in pregnancy: new insights in diagnosis and clinical management. *Nat Rev Endocrinol*. 2017;13(10):610-622.
7. Krassas GE, Poppe K, Glinoer D. Thyroid function and human reproductive health. *Endocr Rev*. 2010;31(5):702-755.
8. Stagnaro-Green A, Abalovich M, Alexander E, *et al*. Guidelines of the American Thyroid Association for the diagnosis and management of thyroid disease during pregnancy and postpartum. *Thyroid*. 2011;21(10):1081-1125.
9. Lazarus J, Brown RS, Daumerie C, Hubalewska-Dydejczyk A, Negro R, Vaidya B. 2014 European thyroid association guidelines for the management of subclinical hypothyroidism in pregnancy and in children. *Eur Thyroid J*. 2014;3(2):76-94.
10. Dong AC, Stagnaro-Green A. Differences in diagnostic criteria mask the true prevalence of thyroid disease in pregnancy: a systematic review and meta-analysis. *Thyroid*. 2019;29(2):278-289.
11. 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):3-126.
12. 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(9):3548-3556.
13. 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*. 2017;86(6):830-836.
14. Korevaar TIM, Medici M, De Rijke YB, *et al*. Ethnic differences in maternal thyroid parameters during pregnancy: the generation r study. *J Clin Endocrinol Metab*. 2013;98(9):3678-3686.
15. Mosso L, Martinez 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(6):942-948.
16. Sheng Y, Huang D, Liu S, *et al*. Reference intervals of thyroid hormones and correlation of BMI with thyroid function in healthy Zhuang ethnic pregnant women. *Biomed Res Int*. 2018;2018:2032413.
17. Andersen SL, Knosgaard L, Handberg A, Vestergaard P, Andersen S. Maternal adiposity, smoking, and thyroid function in early pregnancy. *Endocr Connect*. 2021;10(9):1125-1133.
18. Ozarda Y, Sikaris K, Streichert T, Macri J; IFCC Committee on Reference intervals and Decision Limits (C-RIDL). Distinguishing reference intervals and clinical decision limits – a review by the IFCC Committee on Reference Intervals and Decision Limits. *Crit Rev Clin Lab Sci*. 2018;55(6):420-431.
19. Zimmermann MB. The adverse effects of mild-to-moderate iodine deficiency during pregnancy and childhood: a review. *Thyroid*. 2007;17(9):829-835.
20. Andersson M, Karumbunathan V, Zimmermann MB. Global iodine status in 2011 and trends over the past decade. *J Nutr*. 2012;142(4):744-50.
21. Andersson M, Takkouche B, Egli I, Allen HE, de Benoist B. Current global iodine status and progress over the last decade towards the elimination of iodine deficiency. *Bull World Health Organ*. 2005;83(7):518-25.
22. Abel MH, Korevaar TIM, Erlund I, *et al*. Iodine intake is associated with thyroid function in mild to moderately iodine deficient pregnant women. *Thyroid*. 2018;28(10):1359-1371.
23. Levie D, Derakhshan A, Shu H, *et al*. The association of maternal iodine status in early pregnancy with thyroid function in the Swedish Environmental Longitudinal, Mother and Child, Asthma and Allergy Study. *Thyroid*. 2019;29(11):1660-1668.
24. Osinga JAJ, Derakhshan A, Korevaar TIM, *et al*. Data from: Reference intervals. Consortium on Thyroid and Pregnancy. Deposited February 21, 2022. https://www.consortiumthyroidpregnancy.org/referenceintervals
25. Shi XG, Han C, Li CY, *et al*. Optimal and safe upper limits of iodine intake for early pregnancy in iodine-sufficient regions: a cross-sectional study of 7190 pregnant women in China. *J Clin Endocr Metab*. 2015;100(4):1630-1638.
26. Muzzo S, Burgueno M, Carvajal F, *et al*. [Iodine nutrition in school children of four census areas of Chile] Nutricion de yodo en escolares de cuatro zonas censorias de Chile. *Rev Med Chil*. 1997;125(11):1299-1304.
27. Dhillon-Smith RK, Coomarasamy A. TPO antibody positivity and adverse pregnancy outcomes. *Best Pract Res Clin Endocrinol Metab*. 2020;34(4):101433.
28. Korevaar TI, Steegers EA, Pop VJ, *et al*. Thyroid autoimmunity impairs the thyroidal response to human chorionic gonadotropin: two population-based prospective cohort studies. *J Clin Endocrinol Metab*. 2017;102(1):69-77.
29. De Leo S, Pearce EN. Autoimmune thyroid disease during pregnancy. *Lancet Diabetes Endocrinol*. 2018;6(7):575-586.
30. Hou Y, Liu A, Li J, *et al*. Different thyroidal responses to human chorionic gonadotropin under different thyroid peroxidase antibody and/or thyroglobulin antibody positivity conditions during the first half of pregnancy. *Thyroid*. 2019;29(4):577-585.
31. Ticconi C, Giuliani E, Veglia M, Pietropolli A, Piccione E, Di Simone N. Thyroid autoimmunity and recurrent miscarriage. *Am J Reprod Immunol*. 2011;66(6):452-459.
32. He X, Wang P, Wang Z, He X, Xu D, Wang B. Thyroid antibodies and risk of preterm delivery: a meta-analysis of prospective cohort studies. *Eur J Endocrinol*. 2012;167(4):455-64.
33. Bliddal S, Derakhshan A, Xiao Y, *et al*. Association of thyroid peroxidase antibodies and thyroglobulin antibodies with thyroid function in pregnancy: An individual participant data meta-analysis. *Thyroid*. 2022;828-840.
34. Chaker L, Korevaar TI, Medici M, *et al*. Thyroid function characteristics and determinants: the Rotterdam study. *Thyroid*. 2016;26(9):1195-1204.
35. Pearce EN. Thyroid hormone and obesity. *Curr Opin Endocrinol Diabetes Obes*. 2012;19(5):408-13.
36. Pop VJ, Biondi B, Wijnen HA, Kuppens SM, Lvader H. Maternal thyroid parameters, body mass index and subsequent weight gain during pregnancy in healthy euthyroid women. *Clin Endocrinol (Oxf)*. 2013;79(4):577-583.
37. Haddow JE, Craig WY, Palomaki GE, *et al*. Impact of adjusting for the reciprocal relationship between maternal weight and free thyroxine during early pregnancy. *Thyroid*. 2013;23(2):225-230.
38. Mannisto 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(3):291-298.