Title: Defining gestational thyroid dysfunction through modified non-pregnancy reference intervals: an individual participant meta-analysis

Keywords: Thyroid Gland, Thyroid Function Tests, Reference values, Pregnancy, Thyrotropin, Thyroxine

Authors: Joris A.J. Osinga, MD[1,2], Scott M. Nelson, MRCOG, PhD[3], John P. Walsh, MB, PhD[4,5], Ghalia Ashoor, MD[6], Glenn E Palomaki, PhD[7], Abel López-Bermejo, MD, PhD[8,9], Judit Bassols, MD[10], Ashraf Aminorroaya, MD[11], Maarten A.C. Broeren, PhD[12], Liangmiao Chen, MD, PhD[13], Xuemian Lu, MD, PhD[13], Suzanne J. Brown[4], Flora Veltri, MD, PhD[14], Kun Huang, PhD[15], Tuija Männistö, MD, PhD[16], Marina Vafeiadi, PhD[17], Peter N. Taylor, FRCP, PhD[18], Fang-Biao Tao, MD, PhD[15], Lida Chatzi, MD, PhD[19], Maryam Kianpour, PhD[11], Eila Suvanto, MD, PhD[20], Elena N. Grineva, MD, PhD[21], Kypros H Nicolaides, MD, PhD[22], Mary E. D'Alton, MD[23], Kris G. Poppe, MD, PhD[14], Erik Alexander, MD, PhD[24], Ulla Feldt-Rasmussen, MD, DMSc[25], Sofie Bliddal, MD, PhD[25], Polina V. Popova, MD, PhD[21], Layal Chaker, MD, PhD[1,2,26], W. Edward Visser, MD, PhD[1,2], Robin P. Peeters, MD, PhD[1,2], Arash Derakhshan, MD[1,2], Tanja G.M. Vrijkotte, PhD[27], Victor J.M. Pop, MD, PhD[28], Tim I. M. Korevaar, MD, PhD[1,2]

Affiliations

1. Department of Internal Medicine, Erasmus University Medical Center, 3000 CA Rotterdam, the Netherlands.
2. Academic Center for Thyroid Diseases, Erasmus University Medical Center, 3000 CA, Rotterdam, the Netherlands.
3. School of Medicine, Dentistry and Nursing, University of Glasgow, G12 8QQ Glasgow, UK.
4. Department of Endocrinology and Diabetes, Sir Charles Gairdner Hospital, Nedlands 6009, Western Australia, Australia.
5. Medical School, University of Western Australia, Crawley 6009, Western Australia, Australia.
6. Harris Birthright Research Center for Fetal Medicine, King's College Hospital, SE5 9RS London, United Kingdom.
7. Department of Pathology and Laboratory Medicine, Women & Infants Hospital and Alpert Medical School at Brown University, RI 02903 Providence, Rhode Island.
8. Pediatric Endocrinology Research Group, Girona Biomedical Research Institute (IDIBGI), Dr. Josep Trueta Hospital, 17007 Girona, Spain.
9. Departament de Ciències Mèdiques. Universitat de Girona, 17003, Girona, Spain.

10. Maternal-Fetal Metabolic Research Group, Girona Biomedical Research Institute (IDIBGI), Dr. Josep Trueta Hospital, 17007 Girona, Spain.
11. Isfahan Endocrine and Metabolism Research Center, Isfahan University of Medical Sciences, 81745-33871 Isfahan, Iran.
12. Laboratory of Clinical Chemistry and Haematology, Máxima Medical Centre, 5504 DB Veldhoven, Netherlands.
13. Department of Endocrinology and Rui'an Center of the Chinese-American Research Institute for Diabetic Complications, Third Affiliated Hospital of Wenzhou Medical University, 325035 Wenzhou, China.
14. Endocrine Unit, Centre Hospitalier Universitaire Saint-Pierre, Université Libre de Bruxelles (ULB), Brussels, Belgium.
15. Department of Maternal, Child and Adolescent Health, Scientific Research Center in Preventive Medicine; School of Public Health; Anhui Medical University, 230032 Anhui, China.
16. NordLab, Oulu and Translational Medicine Research Unit, University of Oulu, 90570 Oulu, Finland.
17. Department of Social Medicine, School of Medicine, University of Crete, 710 03 Heraklion, Crete, Greece.
18. Thyroid Research Group, Systems Immunity Research Institute, Cardiff University School of Medicine, CF10 3EU Cardiff, UK.
19. Department of Preventive Medicine, Keck School of Medicine, University of Southern California, 90089 CA, USA.
20. Department of Obstetrics and Gynecology and Medical Research Center Oulu, University of Oulu, 90570 Oulu, Finland.
21. Institute of Endocrinology, Almazov National Medical Research Centre, 197341 Saint Petersburg, Russia.
22. Department of Women and Children's Health, Faculty of Life Sciences and Medicine King's College London, SE5 9RS London, United Kingdom.
23. Department of Obstetrics and Gynecology, Columbia University Irving Medical Center, NY 10032 New York, USA.
24. Division of Endocrinology, Hypertension and Diabetes, Brigham and Women's Hospital, Harvard Medical School, Boston, 02115 MA, USA.
25. Department of Medical Endocrinology and Metabolism, Copenhagen University Hospital, Rigshospitalet, and Department od Clinical Medicine, Faculty of Health and clinical Sciences, Copenhagen University, 2100 Copenhagen, Denmark.
26. Department of Epidemiology, Erasmus University Medical Center, 3000 CA Rotterdam, the Netherlands.
27. Department of Public and Occupational Health, Amsterdam UMC, University of Amsterdam, Amsterdam Public Health Research Institute, 1081 HV Amsterdam, the Netherlands
28. Department of Medical and Clinical Psychology, Tilburg University, 5000 LE Tilburg, The Netherlands.

Corresponding author: Joris Osinga, Generation R, Postbus 2040, 3000 CA Rotterdam,

j.osinga@erasmusmc.nl, ORCID-ID: 0000-0003-2527-5150

Funding: Netherlands Organization for Scientific Research (grant 401.16.020) and a Vidi grant

(016.176.331) from the Netherlands Organization for Scientific Research to R.P.P.

Disclosures: P.T. reports a travel grant from Society for Endocrinology (leadership development award). 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. 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 and reports lecture fee from Merck, Darmstadt. S.B.'s research salary was sponsored by the Capital Region of Denmark's Research Foundation and the Novo Nordisk Foundation (ID 0077221). S.B. received a lecture fee from Merck and Novo Nordisk. All other authors declare no competing interests.

Abstract

Background: Establishing local trimester-specific reference intervals for gestational TSH and FT4 is often not feasible, necessitating alternative strategies. We aimed to systematically quantify the diagnostic performance of standardized modifications of center-specific non-pregnancy reference intervals as compared to trimester-specific reference intervals.

Methods: We included prospective cohorts participating in the Consortium on Thyroid and Pregnancy. After relevant exclusions, reference intervals were calculated per cohort in thyroperoxidase antibody-negative women. Modifications to the non-pregnancy reference intervals included an absolute modification (per 0.1 mU/L TSH or 1 pmol/L FT4), relative modification (in steps of 5%) and fixed limits (upper TSH limit between 3.0 to 4.5 mU/L and lower FT4 limit 5-15 pmol/L). We compared (sub)clinical

hypothyroidism prevalence, sensitivity and positive predictive value (PPV) of aforementioned methodologies with population-based trimester-specific reference intervals.

Results: The final study population comprised 52,496 participants in 18 cohorts. Optimal modifications of standard reference intervals to diagnose gestational overt hypothyroidism were -5% for the upper limit of TSH and +5% for the lower limit of FT4 (sensitivity 0.70, confidence interval [CI] 0.47-0.86; PPV 0.64, CI 0.54-0.74). For subclinical hypothyroidism, these were -20% for the upper limit of TSH and -15% for the lower limit of FT4 (sensitivity 0.91, CI 0.67-0.98; PPV 0.71, CI 0.58-0.80). Absolute and fixed modifications yielded similar results. Confidence intervals were wide, limiting generalizability.

Conclusion: We could not identify modifications of non-pregnancy TSH and FT4 reference intervals that would enable centers to adequately approximate trimester-specific reference intervals. Future efforts should be turned towards studying the meaningfulness of trimester-specific reference intervals and risk-based decision limits.

Introduction

Thyroid dysfunction during pregnancy is associated with a higher risk of miscarriage, preeclampsia, preterm birth, aberrant birthweight and lower offspring IQ [1-6]. Current international guidelines recommend defining gestational thyroid dysfunction according to population and pregnancy-specific TSH and FT4 reference intervals, to take into account thyroid physiology during pregnancy, as well as differences in TSH and FT4 determinants between populations and the use of different laboratory assays [7-9]. However, calculating such local reference intervals is generally not feasible for most centers [10,11]. In addition to the practical hurdles, most of the published reference intervals for TSH and FT4 are not in accordance with the current ATA guidelines, as we recently showed by providing an overview of

published TSH and FT4 reference intervals and methodologies, showing that most studies included used additional exclusion criteria based on health status, did not exclude TPOAb positive participants or used different percentile cutoffs [8]. This is in part because of changing guidelines and in part because many centers use additional exclusion criteria or apply different reference limit cut-offs [8]. These varying methodologies hamper the adoption of reference intervals from other centers, and as such, the vast majority of centers rely on non-pregnancy reference intervals for TSH with either a fixed limit approach (upper limit of 4.0 mU/L for TSH) or a subtraction approach (subtraction of 0.5 mU/L of the upper limit of TSH), while for FT4 varying local approaches are used including non-pregnancy reference intervals [12-14]. These second-tier strategies are considered inferior compared to locally defined reference intervals [15-17]. In a follow-up study, we showed that the use of a fixed upper TSH limit or the subtraction approach results in poor detection rates and high false positive rates for (subclinical) hypothyroidism in early pregnancy with highly variable diagnostic performance between populations (sensitivity 0.63-0.82, false discovery rate 0.11-0.35) [18].

In search of a method that is both easy to implement in clinical practice and would better identify women with an abnormal thyroid function during pregnancy, we set out to investigate if it is possible to modify the center-specific non-pregnancy TSH and FT4 reference intervals so that these are useful in pregnancy. Such an approach could make the establishment of local pregnancy-specific reference intervals obsolete while it takes account of the local assay and pre-existing laboratory harmonization efforts [19,20]. A useful diagnostic approach would need to fulfill some conditions: 1) the diagnostic performance should at least perform better than currently recommended alternative methods (TSH upper limit of 4.0 mU/L or subtraction of 0.5 mU/L) [12,13], and 2) the diagnostic performance should be reasonably consistent between populations.

In this individual participant meta-analysis, we aimed to modify the center-specific non-pregnancy reference intervals of TSH and FT4 in a standardized manner and study the sensitivity and the positive predictive value (PPV) as compared to center-specific gestational reference intervals as calculated in accordance with the current international guidelines.

Methods

The study inclusion and eligibility procedures are described in detail previously [18]. In short, eligible studies were those participating in the Consortium on Thyroid and Pregnancy (https://www.consortiumthyroidpregnancy.org). Exclusion criteria for participants were pre-pregnancy thyroid disease, pregnancy through in-vitro fertilization/ intracytoplasmic sperm injection (IVF/ICSI), use of thyroid (interfering) medication and multiple gestation. For this study, we followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses guidelines for Individual Patient Data and preregistered the study protocol (CRD42021270078), which can be found in the supplemental materials along with an outline of protocol deviations [21]. Study quality and risk of bias were assessed using the Newcastle-Ottawa scale (Supplemental materials [21]). All cohorts were approved by a local review board and acquired participant informed consent or had been granted exemption from it by the local Ethics Committee.

Defining gestational thyroid dysfunction

Non-pregnancy reference intervals were either published and/or provided by the principal investigator of the included cohorts and are assay-specific. We defined the trimesters as 0 to 13 weeks, >13 to 27 weeks and >27 weeks of gestation. For cohorts containing participants with repeated measurements, we used the first available sample for each trimester.

Reference intervals, thyroid dysfunction (overt and subclinical hypothyroidism) and diagnostic test properties were calculated separately for each cohort to account for inter-population differences. All reference intervals were calculated as the 2.5th to 97.5th percentiles in TPOAb-negative participants. Our primary aim was to optimize the diagnosis of thyroid dysfunction states for which treatment is indicated or should be considered based on current guidelines, and thus we limited analyses to overt and subclinical hypothyroidism [13]. A treatment indication was defined as either 1) overt hypothyroidism, 2) subclinical hypothyroidism with TSH>10 mU/L or 3) subclinical hypothyroidism with TPOAb positivity. A treatment consideration was defined as 1) TSH between 2.5 mU/L and the upper reference limit with concomitant TPOAb positivity or 2) subclinical hypothyroidism without TPOAb positivity [13]. Treatment of hyperthyroidism was outside the scope of this study, since gestational hyperthyroidism is often considered physiological and we do not have data available to differentiate between gestational transient thyrotoxicosis and Graves' hyperthyroidism[13]. The prevalence of thyroid dysfunction and diagnostic performance measures were calculated according to several methods; 1) a relative modification of the non-pregnancy upper limit of TSH varying from -5% to -40% in steps of 5%, with modifications to the lower limit of FT4 varying from -20% to +20% in steps of 5% (relative modification approach); 2) a subtraction from the non-pregnancy upper limit of TSH varying from -0.1 to -1.0 mU/L, with modification of the non-pregnancy lower limit of FT4 varying from -5 to +5 pmol/L (-0.39 to +0.39 ng/dL; absolute modification approach) and 3) using fixed upper limits for TSH, varying from 3.0 to 4.5 mU/L, and fixed lower limits for FT4, varying from 5-15 pmol/L (0.39-1.17 ng/dL; fixed limit approach). The choice for the range of modifications was based on previous recommendations (e.g. the fixed upper limit of 4.0mU/L for TSH and 0.5 subtraction from this limit) and the optimal diagnostic performance in this study, to keep the results organized. The results for each method were compared to the reference standard (trimester-specific reference intervals), as is currently advised in international guidelines [12,13].

Diagnostic performance measures

The diagnostic performance of each assessed combination is described using the sensitivity (equivalent to true positive rate, true positive rate among all with the disease according to the trimester-specific method) and the PPV (equivalent to 1-false discovery rate, true positives among all with a positive test result). Presenting the PPV, rather than the specificity, was preferred since the PPV is more informative with regard to false positives for outcomes with a low prevalence [22]. The aim was to maximize both diagnostic performance markers, which poses a challenge, since maximizing sensitivity and the PPV is often a trade-off.

The primary outcome was a single diagnostic performance measure, the F-score (also referred to as F1-score), which is a combined measure of PPV (also referred to as 'precision') and sensitivity (also referred to as 'recall') [23]. A higher F-score denotes a better overall diagnostic performance.

Prediction intervals and the I^2 statistic are presented to illustrate the expected inter-population variation in diagnostic performance and between-study heterogeneity [21,24]. Prediction intervals are an attempt to predict future individual values whereas confidence intervals give an indication of where the mean value lies. To facilitate comparison of diagnostic performance markers between methods, interactive heatmaps were constructed which can be found online [25].

Statistical analyses

Diagnostic performance measures were calculated using 2x2 contingency tables (confusion matrices) per cohort and pooled using random intercept logistic regression models utilizing maximum likelihood for modeling between-study heterogeneity. This approach was chosen since it outperforms conventional two-step inverse-variance approaches for sparse event datasets [26,27]. For each alternative approach, the sensitivity, PPV and F-scores were calculated and compared with the trimester-specific

approach. All analyses were performed using R statistical software version 4.2.2 [28], specifically using the package 'meta' [29], 'ggplot2' [30] and 'heatmaply' [31].

Results

After exclusions, the final study population comprised 52,496 participants included in 18 cohorts (Figure 1) of whom 8.6% were TPOAb positive (range across cohorts 5.7-17.1%; Supplemental table 1 [21]). The prevalence of thyroid function test abnormalities (in the first and second trimester, respectively) according to the trimester-specific approach was 0.5% and 0.3% for overt hypothyroidism and 3.4% and 3.2% for subclinical hypothyroidism. The inclusion process and maternal demographics are described in detail previously [18]. Cohort-specific prevalence of thyroid disease, reference limits, iodine status and assay information can be found in Supplemental tables 2-6 [21]. All figures are accompanied by supplemental tables [21] containing the diagnostic performance markers for each specific combination (Figure 2 is an explanatory example of the diagnostic markers presented). To facilitate comparison of diagnostic performance measures, an interactive version of the heatmaps including other diagnostic performance measures can be found online and is also referred to throughout, as an alternative to the supplemental tables [21] (https://www.consortiumthyroidpregnancy.org/heatmaps [25]).

Diagnostic performance of alternative approaches

Using the relative modification approach in the first trimester, the highest F-scores for overt hypothyroidism were achieved with a relative subtraction of 5% for the upper reference limit of TSH and a relative addition of 5% for the lower reference limit of FT4 (F-score 0.65; Figure 3A). The associated sensitivity was 0.70 (95% confidence interval [CI] 0.47-0.86; 95% prediction interval [PI] 0.06-0.99; I^2 64%), and the PPV was 0.64 (CI 0.54-0.74; PI 0.18-0.94; I^2 45%; Figure 3A, Supplemental table 7 [21], Interactive figures [25]). For subclinical hypothyroidism the highest F-scores were achieved with a relative

subtraction of 20% for the upper reference limit of TSH and a relative subtraction of 15% for the lower reference limit of FT4 (F-score 0.69; Figure 3B). Associated sensitivity was 0.91 (CI 0.67-0.98; PI 0.02-1.00; I^2 95%) and PPV was 0.71 (CI 0.58-0.80; PI 0.20-0.96; I^2 95%; Supplemental table 8 [21], Interactive figures [25]).

Using the absolute modification approach in the first trimester, the highest F-scores for overt hypothyroidism were achieved with a subtraction of either -0.1, -0.2 or -0.3 mU/L for the upper limit of TSH and an addition of +1 pmol/L to the lower limit of FT4 and (F-score 0.62; Figure 3C). Associated sensitivity (for upper limit TSH -0.2 mU/L) was 0.74 (CI 0.52-0.89; PI 0.08-0.99; I^2 66%) and PPV was 0.57 (CI 0.45-0.68; PI 0.24-0.84; I^2 39%; Supplemental table 9 [21], Interactive figures [25]). For subclinical hypothyroidism the highest F-scores were achieved with a subtraction of -0.8 mU/L from the upper limit of TSH and a subtraction of either -1, -2, -3, -4 or -5 pmol/L from the lower limit of FT4 (F-score 0.64; Figure 3D). Associated sensitivity (for lower limit FT4 -4 pmol/L) was 0.91 (CI 0.61-0.98; PI 0.01-1.00; I^2 95%) and PPV was 0.68 (CI 0.55-0.78; PI 0.20-0.95; I^2 95%; Supplemental table 10 [21], Interactive figures [25]).

Using the fixed limit approach in the first trimester, the highest F-scores for overt hypothyroidism were achieved with an upper limit of TSH of either 3.8, 3.9, 4.0, 4.1 and 4.4 mU/L and a lower limit of FT4 of 12 pmol/L (F-score 0.65; Figure 3E). Associated sensitivity (for upper limit TSH 4.0 mU/L) was 0.83 (CI 0.70-0.91; PI 0.41-0.97; I^2 0%) and PPV was 0.50 (CI0.32-0.68; PI 0.05-0.95; I^2 70%; Supplemental table 11 [21], Interactive figures [25]). For subclinical hypothyroidism the highest F- were achieved with an upper limit of TSH of 3.2 mU/L and a lower limit of FT4 of either 5, 6, 7 or 8 pmol/L (F-score 0.70; Figure 3F). Associated sensitivity (for lower limit FT4 8 pmol/L) was 0.99 (CI 0.88-1.00; PI 0.03-1.00; I^2 91%) and PPV was 0.66 (CI 0.51-0.79; PI 0.11-0.97; I^2 96%; Supplemental table 12 [21], Interactive figures [25]).

Additional analyses

In the second trimester, maximum F-scores were similar for the relative modification method, the absolute modification approach and the fixed limit approach (Supplemental figure 1A-F [21]). However, comparing the diagnostic performance measures of individual studies, the variability between studies was very high, as reflected by overlapping confidence intervals for all methods, based on the wide prediction intervals and based on high I^2 statistics for higher F-scores (Supplemental tables 13-18 [21]). The diagnostic performance of alternative methods to detect women for whom levothyroxine treatment is indicated and those for whom treatment should be considered, according to ATA guidelines, in the first trimester and second trimester were similar based on overlapping confidence intervals (Supplemental figures 2, 3; Supplemental tables 19-30 [21]).

Discussion

In this study, we systematically evaluated multiple standardized procedures to modify non-pregnancy TSH and FT4 reference intervals with the aim of diagnosing the same individuals as having an abnormal gestational thyroid function in line with the 'gold-standard' approach of center-specific and trimester-specific reference intervals. Despite our efforts, we were unable to identify a standardized procedure that achieved a satisfactory balance between sensitivity and PPV for gestational thyroid dysfunction without considerable variability across different populations. These results underscore the inherent challenge in balancing precise identification of gestational thyroid dysfunction with the practical limitations of applying these diagnostic strategies universally in clinical settings, and indicate that calculating local center and pregnancy-specific reference intervals for TSH and FT4 should still be considered as current best practice.

Current recommendations on gestational reference interval definitions for TSH and FT4 are time and resource consuming and are not feasible for most centers worldwide. The modification of non-pregnancy reference intervals for the use in pregnancy could overcome feasibility problems. However, in

the current study we show that the variability in TSH and FT4 distributions leads to unacceptable variation in diagnostic performance between cohorts. A possible explanation for this variation is that even the non-pregnancy TSH and FT4 reference intervals are not an adequate reflection of the distribution of thyroid function tests for a population if they are based on the manufacturer's recommendation rather than local laboratory-specific establishment of the intervals. Methods for determining reference intervals in pregnancy and outside pregnancy often differ, as current recommendations on the establishment of reference limits in pregnancy include the local population and are by definition a reflection of local TSH and FT4 distributions [12-14], while reference limits outside pregnancy are often supplied by the assay manufacturer, who mostly established reference intervals in selected, non-pregnant populations [32,33]. Global harmonization efforts for TSH and FT4 assays by the International Federation of Clinical Chemistry and Laboratory Medicine (IFCC) Committee for Standardization of Thyroid Function Tests (C-STFT) are ongoing to address this issue outside of pregnancy, which could lead to an attenuation of this mismatch [19,20].

We also show that for overt hypothyroidism and for subclinical hypothyroidism, different and sometimes opposing modifications of the reference limits of TSH and FT4 were needed to achieve maximum diagnostic performance. For instance, when reviewing the relative modifications needed to achieve the best diagnostic performance for overt hypothyroidism in the first trimester, we find that the best F-score of 0.65 is achieved with the upper limit of TSH -5% and the lower limit of FT4 +5% (Figure 2A) while the best F-score for subclinical hypothyroidism of 0.69 is achieved with the upper limit of TSH -20% and the lower limit of FT4 -15% (Figure 2B). We previously showed that the use of trimester-specific reference intervals for FT4 are most important for the correct diagnosis of overt hypothyroidism while for subclinical hypothyroidism the use of trimester-specific reference intervals for TSH are more important[18], which could explain the current results. This finding suggests that a uniform rule established to diagnose both overt and subclinical disease, would be good at diagnosing one, at the cost

of incorrectly diagnosing the other. We also observe that the trends in diagnostic performance for a treatment indication (Supplemental figure 2A, 2C, 2E [21]) mostly overlap with the trend in diagnostic performance for subclinical hypothyroidism (Figure 2B, 2D, 2F). This is because most women with a treatment indication present with subclinical hypothyroidism with TPOAb positivity (73.6%) rather than overt hypothyroidism (25.4%) or subclinical hypothyroidism with TSH>10 (1.1%; data not shown). Since the prevalence of subclinical hypothyroidism is much higher than of overt hypothyroidism, it can be expected that the best diagnostic performance of a test to detect a treatment indication is reached with the same modifications as for subclinical hypothyroidism. This concept is important for future recommendations on universal reference limits because diagnosing overt hypothyroidism, an entity with an evident treatment indication, is generally prioritized in diagnostic strategies for gestational thyroid dysfunction. However, failing to identify the more prevalent subclinical disease could also lead to decreased benefits of (selective) screening. While we found no method with an agreeable trade-off in terms of diagnostic performance, it is important to realize that the interpretation of diagnostic performance of a test depends on the prior probability of disease [34]. This is a highly relevant concept when thinking about differences between generalized population screening (with a low prior probability) versus high-risk case-based screening (with higher prior probabilities). For example, for a hypothetical diagnostic test with a sensitivity of 0.75 and a specificity of 0.99 (roughly equal to the tests assessed in our study), a pre-test probability of 3% would result in a post-positive test probability (or PPV) of 70% and a false discovery rate of 30%. Using the same sensitivity and specificity, a pre-test probability of 10% would results in a post-positive test probability of 89% with a false discovery rate of 11%. The current study population consists of population-based cohort studies as a reflection of the general population, which have a low prior probability of disease equal to the population prevalence and similar to a universal screening approach. One option to improve how alternative reference interval strategies could identify those with an abnormal thyroid function would be to increase the prior

probability of disease [34]. This can be achieved by optimizing the identification of high-risk subgroups and a risk-based screening approach, which could improve the accuracy of diagnostic strategies [35]. Thus, the implementation of universal screening will be inherently associated with the lowest prior probability of disease and the highest rates of both over and underdiagnosis, especially if alternative strategies are used to define thyroid function test abnormalities.

The heterogeneity between populations (as denoted by wide prediction intervals and high I^2 statistics) underline that calculating local center and pregnancy specific reference intervals for TSH and FT4 should still be considered as current best practice. However, other strategies for the improvement of the diagnosis of gestational thyroid dysfunction might prove more effective. The trimester-specific approach is currently accepted as the best diagnostic method for diagnosing thyroid dysfunction in pregnancy, but the pragmatic division of the gestational period in trimesters does not necessarily reflect the physiological changes of thyroid function tests during pregnancy [36-38]. Further studies are needed to assess which gestational period reference intervals should be based upon to optimally identify the women at increased risk of adverse events due to thyroid dysfunction, or if any form of standardization to gestational age should be abandoned altogether. Current reference interval definitions are based on outlying percentiles of TSH and FT4 distributions (2.5^{th} and 97.5^{th} percentiles), values above or below those cutoffs were later shown to be associated with adverse pregnancy outcomes [39]. With increasing data availability in the literature, the ideal way to establish reference values would be to turn this methodology around and base the cut-offs on the risk of adverse outcomes, similar to other fields [40,41]. Obvious adverse pregnancy events would be those associated with thyroid function tests in previous studies such as preterm birth and offspring IQ scores [3,4,6]. Since we did not identify an adequate or easily implementable methodology to approach trimester-specific reference intervals in the current study, our group will aim to establish risk-based decision limits.

In this study, we were able to leverage a large international dataset of multiple population-based prospective cohort studies to assess novel strategies for diagnosing thyroid dysfunction in pregnancy. The interpretation of the results of this study are limited to populations with sufficient or mild-to-moderate iodine deficiency since studies with excessive status were excluded and no studies were performed in an area of severe iodine deficiency. Additionally, multiple differences between the included study populations, including differences in iodine supplementation, assays and determinants of thyroid function tests, could have contributed to the variability in diagnostic performance of the non-pregnancy reference interval adaptations assessed in this study. Adaptations of non-pregnancy reference limits could be more accurate in specific populations, which we were not able to assess with sufficient power. Nonetheless, this study reflects common practice, as these factors naturally vary between populations. The results of the current study may not be optimally generalizable to present-day populations since the inclusion periods for the majority of included cohorts were between the year 2000 and 2015. It is likely that determinants of thyroid function and assay calibrations standards have changed over time [42]. It can however be expected that large inter-population differences, as demonstrated in this study, are still present to this day. Ongoing harmonization efforts by the IFCC could improve the diagnostic performance of alternative strategies and future studies could assess if a generalizable rule is more effective in cohorts established after the start of the harmonization efforts.

In conclusion, this is the first study to systematically quantify the diagnostic performance of standardized modifications of non-pregnancy TSH and FT4 reference intervals in pregnancy. We show that standardized modifications have poor overlap in diagnostic accuracy compared with cohort and trimester-specific reference intervals, resulting in considerable variation in diagnostic performance between populations. Future efforts should be turned towards studying the meaningfulness of trimester-specific, pregnancy-specific reference intervals and the establishment of risk-based decision limits.

Acknowledgements

The authors would like to gratefully acknowledge all participants, general practitioners, hospitals, and midwives for their important contribution to the establishment of the cohorts and the resulting works. Acknowledgements for individual cohorts are listed in the supplemental materials [21].

Data availability

The data that support the findings of this study are not publicly available due to local, national and international restrictions aimed to protect the privacy of research participants.

References

1. Derakhshan A, Peeters RP, Taylor PN, Bliddal S, Carty DM, Meems M, Vaidya B, Chen L, Knight BA, Ghafoor F, Popova PV, Mosso L, Oken E, Suvanto E, Hisada A, Yoshinaga J, Brown SJ, Bassols J, Auvinen J, Bramer WM, Lopez-Bermejo A, Dayan CM, French R, Boucai L, Vafeiadi M, Grineva EN, Pop VJM, Vrijkotte TG, Chatzi L, Sunyer J, Jimenez-Zabala A, Riano I, Rebagliato M, Lu X, Pirzada A, Mannisto T, Delles C, Feldt-Rasmussen U, Alexander EK, Nelson SM, Chaker L, Pearce EN, Guxens M, Steegers EAP, Walsh JP, Korevaar TIM. Association of maternal thyroid function with birthweight: a systematic review and individual-participant data meta-analysis. *Lancet Diabetes Endocrinol*. Jun 2020;8(6):501-510. doi:10.1016/S2213-8587(20)30061-9

2. Toloza FJK, Derakhshan A, Mannisto T, Bliddal S, Popova PV, Carty DM, Chen L, Taylor P, Mosso L, Oken E, Suvanto E, Itoh S, Kishi R, Bassols J, Auvinen J, Lopez-Bermejo A, Brown SJ, Boucai L, Hisada A, Yoshinaga J, Shilova E, Grineva EN, Vrijkotte TGM, Sunyer J, Jimenez-Zabala A, Riano-Galan I, Lopez-Espinosa MJ, Prokop LJ, Singh Ospina N, Brito JP, Rodriguez-Gutierrez R, Alexander EK, Chaker L, Pearce EN, Peeters RP, Feldt-Rasmussen U, Guxens M, Chatzi L, Delles C, Roeters van Lennep JE, Pop VJM, Lu X, Walsh JP, Nelson SM, Korevaar TIM, Maraka S. Association between maternal thyroid function and risk of gestational hypertension and pre-eclampsia: a systematic review and individual-participant data meta-analysis. *Lancet Diabetes Endocrinol*. Apr 2022;10(4):243-252. doi:10.1016/S2213-8587(22)00007-9

3. Levie D, Korevaar TIM, Bath SC, Dalmau-Bueno A, Murcia M, Espada M, Dineva M, Ibarluzea JM, Sunyer J, Tiemeier H, Rebagliato M, Rayman MP, Peeters RP, Guxens M. Thyroid Function in Early Pregnancy, Child IQ, and Autistic Traits: A Meta-Analysis of Individual Participant Data. *J Clin Endocrinol Metab*. Aug 1 2018;103(8):2967-2979. doi:10.1210/jc.2018-00224

4. Thompson W, Russell G, Baragwanath G, Matthews J, Vaidya B, Thompson-Coon J. Maternal thyroid hormone insufficiency during pregnancy and risk of neurodevelopmental disorders in offspring: A systematic review and meta-analysis. *Clin Endocrinol (Oxf)*. Apr 2018;88(4):575-584. doi:10.1111/cen.13550

5. Han Y, Gao X, Wang X, Zhang C, Gong B, Peng B, Li J, Liu A, Shan Z. A Systematic Review and Meta-Analysis Examining the Risk of Adverse Pregnancy and Neonatal Outcomes in Women with Isolated Hypothyroxinemia in Pregnancy. *Thyroid*. May 2023;33(5):603-614. doi:10.1089/thy.2022.0600
6. Korevaar TIM, Derakhshan A, Taylor PN, Meima M, Chen L, Bliddal S, Carty DM, Meems M, Vaidya B, Shields B, Ghafoor F, Popova PV, Mosso L, Oken E, Suvanto E, Hisada A, Yoshinaga J, Brown SJ, Bassols J, Auvinen J, Bramer WM, Lopez-Bermejo A, Dayan C, Boucai L, Vafeiadi M, Grineva EN, Tkachuck AS, Pop VJM, Vrijkotte TG, Guxens M, Chatzi L, Sunyer J, Jimenez-Zabala A, Riano I, Murcia M, Lu X, Mukhtar S, Delles C, Feldt-Rasmussen U, Nelson SM, Alexander EK, Chaker L, Mannisto T, Walsh JP, Pearce EN, Steegers EAP, Peeters RP. Association of thyroid function test abnormalities and thyroid autoimmunity with preterm birth: a systematic review and meta-analysis. *JAMA*. Aug 20 2019;322(7):632-641. doi:10.1001/jama.2019.10931
7. Krassas GE, Poppe K, Glinoer D. Thyroid function and human reproductive health. *Endocr Rev*. Oct 2010;31(5):702-755. doi:10.1210/er.2009-0041
8. Osinga JAJ, Derakhshan A, Palomaki GE, Ashoor G, Mannisto T, Maraka S, Chen L, Bliddal S, Lu X, Taylor PN, Vrijkotte TGM, Tao FB, Brown SJ, Ghafoor F, Poppe K, Veltri F, Chatzi L, Vaidya B, Broeren MAC, Shields BM, Itoh S, Mosso L, Popova PV, Anopova AD, Kishi R, Aminorroaya A, Kianpour M, Lopez-Bermejo A, Oken E, Pirzada A, Vafeiadi M, Bramer WM, Suvanto E, Yoshinaga J, Huang K, Bassols J, Boucai L, Feldt-Rasmussen U, Grineva EN, Pearce EN, Alexander EK, Pop VJM, Nelson SM, Walsh JP, Peeters RP, Chaker L, Nicolaides KH, D'Alton ME, Korevaar TIM. TSH and FT4 reference intervals in pregnancy: a systematic review and individual participant data meta-analysis. *J Clin Endocrinol Metab*. Sep 28 2022;107(10):2925-2933. doi:10.1210/clinem/dgac425
9. Springer D, Bartos V, Zima T. Reference intervals for thyroid markers in early pregnancy determined by 7 different analytical systems. *Scand J Clin Lab Invest*. Mar 2014;74(2):95-101. doi:10.3109/00365513.2013.860617
10. Negro R, Attanasio R, Papini E, Guglielmi R, Grimaldi F, Toscano V, Niculescu DA, Paun DL, Poiana C. A 2018 Italian and Romanian Survey on Subclinical Hypothyroidism in Pregnancy. *Eur Thyroid J*. Nov 2018;7(6):294-301. doi:etj-0007-0294 [pii]10.1159/000490944
11. Toloza FJK, Ospina NMS, Rodriguez-Gutierrez R, O'Keeffe DT, Brito JP, Montori VM, Maraka S. Practice Variation in the Care of Subclinical Hypothyroidism During Pregnancy: A National Survey of Physicians in the United States. *J Endocr Soc*. Oct 2019;3(10):1892-1906. doi:10.1210/js.2019-00196
12. 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*. Jun 2014;3(2):76-94. doi:etj-0003-0076 [pii]
13. Alexander EK, Pearce EN, Brent GA, Brown RS, Chen H, Dosiou C, Grobman WA, Laurberg P, Lazarus JH, Mandel SJ, Peeters RP, Sullivan S. 2017 Guidelines of the American Thyroid Association for the Diagnosis and Management of Thyroid Disease During Pregnancy and the Postpartum. *Thyroid*. Mar 2017;27(3):315-389. doi:10.1089/thy.2016.0457
14. Thyroid Disease in Pregnancy: ACOG Practice Bulletin, Number 223. *Obstet Gynecol*. Jun 2020;135(6):e261-e274. doi:10.1097/AOG.0000000000003893
15. Bliddal S, Feldt-Rasmussen U, Boas M, Faber J, Juul A, Larsen T, Precht DH. Gestational age-specific reference ranges from different laboratories misclassify pregnant women's thyroid status: comparison of two longitudinal prospective cohort studies. *European Journal of Endocrinology*. Feb 2014;170(2):329-39.
16. Liu J, Yu X, Xia M, Cai H, Cheng G, Wu L, Li Q, Zhang Y, Sheng M, Liu Y, Qin X. 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? *Clin Biochem*. Apr 2017;50(6):309-317. doi:S0009-9120(16)30630-0 [pii]10.1016/j.clinbiochem.2016.11.036

17. Mehran L, Amouzegar A, Delshad H, Askari S, Hedayati M, Amirshekari G, Azizi F. Trimester-specific reference ranges for thyroid hormones in iranian pregnant women. Article. *J Thyroid Res*. 2013;2013doi:10.1155/2013/651517
18. Osinga JAJ, Derakhshan A, Feldt-Rasmussen U, Huang K, Vrijkotte TGM, Mannisto T, Bassols J, Lopez-Bermejo A, Aminorroaya A, Vafeiadi M, Broeren MAC, Palomaki GE, Ashoor G, Chen L, Lu X, Taylor PN, Tao FB, Brown SJ, Sitoris G, Chatzi L, Vaidya B, Popova PV, Vasukova EA, Kianpour M, Suvanto E, Grineva EN, Hattersley A, Pop VJM, Nelson SM, Walsh JP, Nicolaides KH, D'Alton ME, Poppe KG, Chaker L, Bliddal S, Korevaar TIM. TSH and FT4 reference interval recommendations and prevalence of gestational thyroid dysfunction: quantification of current diagnostic approaches. *J Clin Endocrinol Metab*. Sep 22 2023;doi:10.1210/clinem/dgad564
19. Thienpont LM, Van Uytfanghe K, De Grande LAC, Reynders D, Das B, Faix JD, MacKenzie F, Decallonne B, Hishinuma A, Lapauw B, Taelman P, Van Crombrugge P, Van den Bruel A, Velkeniers B, Williams P, Tests ICfSoTF. Harmonization of serum thyroid-stimulating hormone measurements paves the way for the adoption of a more uniform reference interval. *Clin Chem*. Jul 2017;63(7):1248-1260. doi:10.1373/clinchem.2016.269456
20. Thienpont LM, Van Uytfanghe K, Van Houcke S, Das B, Faix JD, MacKenzie F, Quinn FA, Rottmann M, Van den Bruel A, Tests ICfSoTF. A progress report of the IFCC committee for standardization of thyroid function tests. *Eur Thyroid J*. Jun 2014;3(2):109-16. doi:10.1159/000358270
21. Osinga JAJ, Derakhshan A, Korevaar TIM. Data from: reference intervals. Consortium on thyroid and pregnancy. Updated 18-07-2023. https://www.consortiumthyroidpregnancy.org/referenceintervals
22. Lutgendorf MA, Stoll KA. Why 99% may not be as good as you think it is: limitations of screening for rare diseases. *J Matern Fetal Neonatal Med*. 2016;29(7):1187-9.
23. Goutte C, Gaussier E. A probabilistic interpretation of precision, recall and F-score, with implication for evaluation. Springer; 2005:345-359.
24. Riley RD, Higgins JP, Deeks JJ. Interpretation of random effects meta-analyses. *BMJ*. Feb 10 2011;342:d549. doi:bmj.d549 [pii]10.1136/bmj.d549
25. Osinga JAJ, Derakhshan A, Peeters RP, Korevaar TIM. Data from: Heatmaps. Consortium on Thyroid and Pregnancy. Updated 31-01-2024. https://www.consortiumthyroidpregnancy.org/heatmaps
26. Lin L, Chu H. Meta-analysis of proportions using generalized linear mixed models. *Epidemiology*. Sep 2020;31(5):713-717. doi:00001648-202009000-00016 [pii]10.1097/EDE.0000000000001232
27. Stijnen T, Hamza TH, Ozdemir P. Random effects meta-analysis of event outcome in the framework of the generalized linear mixed model with applications in sparse data. *Stat Med*. Dec 20 2010;29(29):3046-67. doi:10.1002/sim.4040
28. R Core Team R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2022. https://www.R-project.org/.
29. Balduzzi S, Rucker G, Schwarzer G. How to perform a meta-analysis with R: a practical tutorial. *Evid Based Ment Health*. Nov 2019;22(4):153-160. doi:ebmental-2019-300117 [pii]10.1136/ebmental-2019-300117
30. Wickham H. ggplot2: elegant graphics for data analysis. Springer-Verlag New York.
31. Galili T, O'Callaghan A, Sidi J, Sievert C. heatmaply: an R package for creating interactive cluster heatmaps for online publishing. *Bioinformatics*. May 1 2018;34(9):1600-1602. doi:4562328 [pii]btx657 [pii]10.1093/bioinformatics/btx657
32. Brochure. Roche Diagnostics GmbH. Reference Intervals for Children and Adults Elecsys Thyroid Tests 2009.
33. Brochure. Abbott Diagnostic Division. Architect system TSH ref 7K62. 2010.
34. Bours MJ. Bayes' rule in diagnosis. *J Clin Epidemiol*. Mar 2021;131:158-160. doi:S0895-4356(20)31225-7 [pii]10.1016/j.jclinepi.2020.12.021

35. Osinga JAJ, Liu Y, Männistö T, Vafeiadi M, Tao FB, Vaidya B, Vrijkotte TGM, Mosso L, Bassols J, López-Bermejo A, Boucai L, Aminorroaya A, Feldt-Rasmussen U, Hisada A, Yoshinaga J, Broeren MAC, Itoh S, Kishi R, Ashoor G, Chen L, Veltri F, Lu X, Taylor PN, Brown SJ, Chatzi L, Popova PV, Grineva EN, Ghafoor F, Pirzada A, Kianpour M, Oken E, Suvanto E, Hattersley A, Rebagliato M, Riaño-Galán I, Irizar A, Vrijheid M, Delgado-Saborit JM, Fernández-Somoano A, Santa-Marina L, Boelaert K, Brenta G, Dhillon-Smith R, Dosiou C, Eaton JL, Guan H, Lee SY, Maraka S, Morris-Wiseman LF, Nguyen CT, Shan Z, Guxens M, Pop VJM, Walsh JP, Nicolaides KH, D'Alton ME, Visser WE, Carty DM, Delles C, Nelson SM, Alexander EK, Chaker L, Palomaki GE, Peeters RP, Bliddal S, Huang K, Poppe KG, Pearce EN, Derakhshan A, Korevaar TIM. Risk Factors for Thyroid Dysfunction in Pregnancy: An Individual Participant Data Meta-Analysis. *Thyroid*. Mar 28 2024;

36. Korevaar TIM, Medici M, Visser TJ, Peeters RP. Thyroid disease in pregnancy: new insights in diagnosis and clinical management. *Nat Rev Endocrinol*. Oct 2017;13(10):610-622. doi:10.1038/nrendo.2017.93

37. Glinoer D, de Nayer P, Bourdoux P, Lemone M, Robyn C, van Steirteghem A, Kinthaert J, Lejeune B. Regulation of maternal thyroid during pregnancy. *J Clin Endocrinol Metab*. Aug 1990;71(2):276-87. doi:10.1210/jcem-71-2-276

38. Andersen SL, Andersen S, Carle A, Christensen PA, Handberg A, Karmisholt J, Knosgaard L, Kristensen SR, Bulow Pedersen I, Vestergaard P. Pregnancy week-specific reference ranges for thyrotropin and free thyroxine in the north Denmark region pregnancy cohort. *Thyroid*. Mar 2019;29(3):430-438. doi:10.1089/thy.2018.0628

39. van den Boogaard E, Vissenberg R, Land JA, van Wely M, van der Post JAM, Goddijn M, Bisschop PH. Significance of (sub) clinical thyroid dysfunction and thyroid autoimmunity before conception and in early pregnancy: a systematic review. *Hum Reprod Update*. Sep-Oct 2011;17(5):605-619. doi:10.1093/humupd/dmr024

40. Group HSCR, Metzger BE, Lowe LP, Dyer AR, Trimble ER, Chaovarindr U, Coustan DR, Hadden DR, McCance DR, Hod M, McIntyre HD, Oats JJ, Persson B, Rogers MS, Sacks DA. Hyperglycemia and adverse pregnancy outcomes. *N Engl J Med*. May 8 2008;358(19):1991-2002. doi:358/19/1991 [pii]10.1056/NEJMoa0707943

41. Xu Y, Derakhshan A, Hysaj O, Wildisen L, Ittermann T, Pingitore A, Abolhassani N, Medici M, Kiemeney L, Riksen NP, Dullaart RPF, Trompet S, Dorr M, Brown SJ, Schmidt B, Fuhrer-Sakel D, Vanderpump MPJ, Muendlein A, Drexel H, Fink HA, Ikram MK, Kavousi M, Rhee CM, Bensenor IM, Azizi F, Hankey GJ, Iacoviello M, Imaizumi M, Ceresini G, Ferrucci L, Sgarbi JA, Bauer DC, Wareham N, Boelaert K, Bakker SJL, Jukema JW, Vaes B, Iervasi G, Yeap BB, Westendorp RGJ, Korevaar TIM, Volzke H, Razvi S, Gussekloo J, Walsh JP, Cappola AR, Rodondi N, Peeters RP, Chaker L, Thyroid Studies C. The optimal healthy ranges of thyroid function defined by the risk of cardiovascular disease and mortality: systematic review and individual participant data meta-analysis. *Lancet Diabetes Endocrinol*. Oct 2023;11(10):743-754. doi:S2213-8587(23)00227-9 [pii]10.1016/S2213-8587(23)00227-9

42. Van Uytfanghe K, Ehrenkranz J, Halsall D, Hoff K, Loh TP, Spencer CA, Kohrle J. Thyroid Stimulating Hormone and Thyroid Hormones (Triiodothyronine and Thyroxine): An American Thyroid Association-Commissioned Review of Current Clinical and Laboratory Status. *Thyroid*. Sep 2023;33(9):1013-1028. doi:10.1089/thy.2023.0169 [pii]

10.1089/thy.2023.0169

Figure legends

Figure 1 – Flowchart of included cohorts and participants

Figure 2 - Diagnostic performance of modified non-pregnancy reference invals for overt hypothyroidism using relative modification

Figure 2 legend: Diagnostic performance for relative modifications of non-pregnancy reference intervals for the diagnosis of overt hypothyroidism, presented as F-scores. The zoomed in section presents additional diagnostic performance markers for selected modifications, of which an interactive version can be found online (https://www.consortiumthyroidpregnancy.org/heatmaps).

Figure 3 – Diagnostic performance of modified non-pregnancy reference intervals for overt and subclinical hypothyroidism

Figure 3 legend – Diagnostic performance of modified non-pregnancy reference intervals are presented using a relative modification (A, B), absolute modifications (C, D) and fixed limits (E, F) for overt and subclinical hypothyroidism, respectively, of which an interactive version can be found online (https://www.consortiumthyroidpregnancy.org/heatmaps).

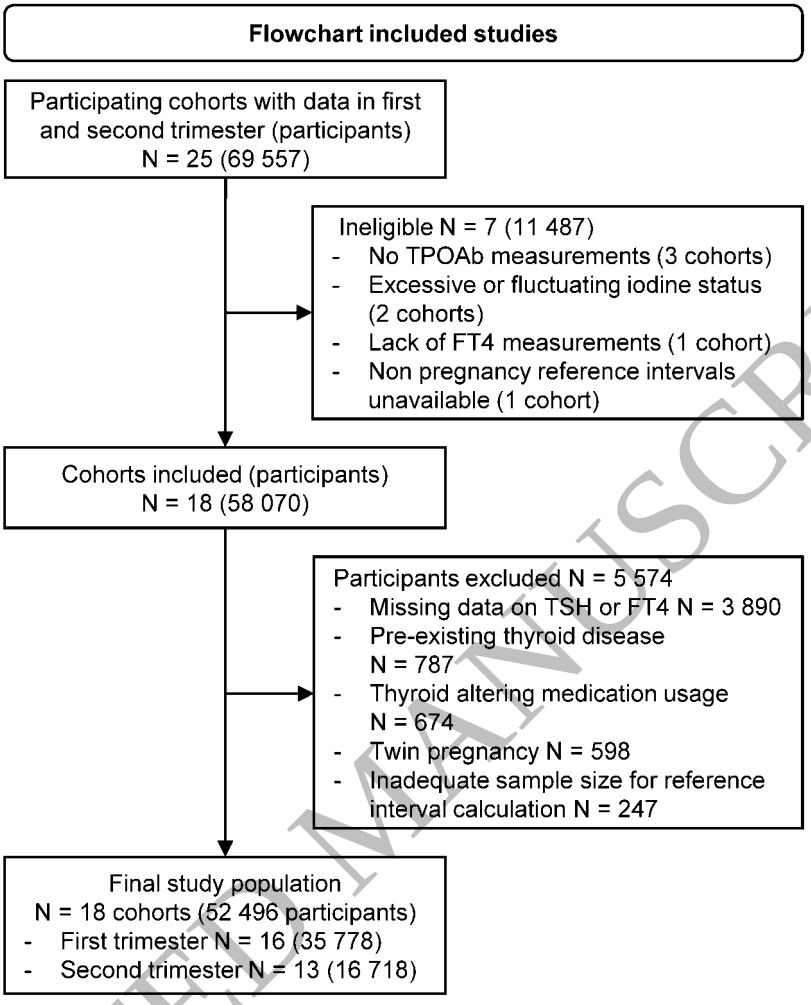
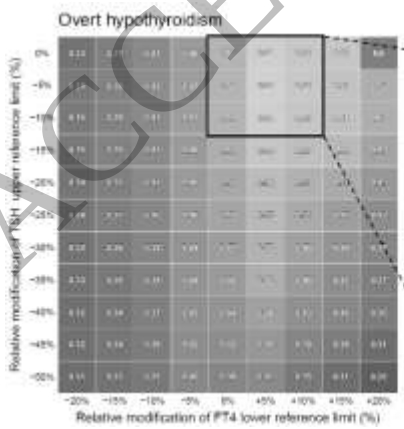


Figure 1
111x137 mm (x DPI)



F-Score	Modification		Sensitivity				Positive Predictive Value			
	ULTSH	*LLFT4*		*Confidence interval*	*Prediction interval*	I^2 *Statistic*		*Confidence interval*	*Prediction interval*	I^2 *Statistic*
0.54	None	None	0.49	0.33-0.65	0.08-0.91	59%	0.65	0.45-0.81	0.13-0.96	26%
0.54	-5%	None	0.53	0.32-0.73	0.05-0.96	48%	0.65	0.46-0.81	0.15-0.95	17%
0.54	-10%	None	0.54	0.34-0.73	0.05-0.96	44%	0.63	0.44-0.79	0.14-0.95	30%
0.64	None	+5%	0.65	0.46-0.80	0.10-0.97	67%	0.65	0.54-0.74	0.17-0.94	48%
0.65	-5%	+5%	0.70	0.47-0.86	0.06-0.99	64%	0.64	0.54-0.74	0.18-0.94	45%
0.64	-10%	+5%	0.71	0.49-0.86	0.07-0.99	64%	0.62	0.46-0.76	0.17-0.93	55%
0.61	None	+10%	0.75	0.50-0.91	0.05-0.99	51%	0.56	0.45-0.67	0.25-0.83	35%
0.61	-5%	+10%	0.83	0.53-0.95	0.03-1.00	44%	0.55	0.44-0.66	0.25-0.82	33%
0.60	-10%	+10%	0.84	0.55-0.95	0.03-1.00	46%	0.53	0.42-0.65	0.22-0.82	39%

Figure 2
313x107 mm (x DPI)

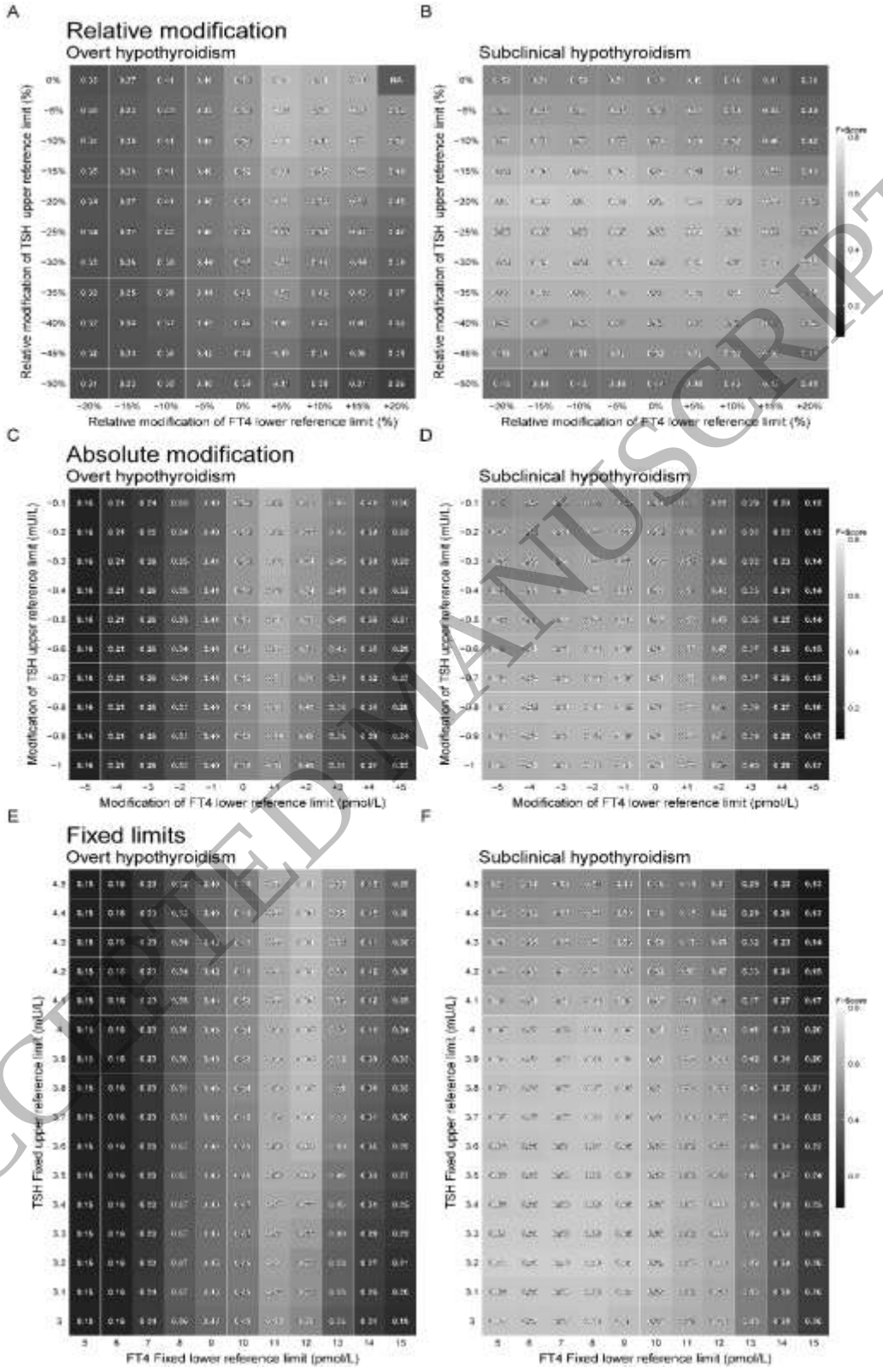


Figure 3
328x559 mm (x DPI)