Aalborg Universitet

Classification of maternal thyroid function in early pregnancy using repeated blood samples

Knøsgaard, Louise; Andersen, Lars Stig; Hansen, Annebirthe Bo; Vestergaard, Peter; Andersen, Stine Linding

Link to publication from Aalborg University

RESEARCH

Classification of maternal thyroid function in early pregnancy using repeated blood samples

Louise Knøsgaard[1,2], Stig Andersen[2,3], Annebirthe Bo Hansen[1], Peter Vestergaard[2,4,5] and **Stine Linding Andersen[1,2]**

[1]Department of Clinical Biochemistry, Aalborg University Hospital, Aalborg, Denmark
[2]Department of Clinical Medicine, Aalborg University, Aalborg, Denmark
[3]Department of Geriatrics, Aalborg University Hospital, Aalborg, Denmark
[4]Department of Endocrinology, Aalborg University Hospital, Aalborg, Denmark
[5]Steno Diabetes Center North Jutland, Aalborg University Hospital, Aalborg, Denmark

Correspondence should be addressed to L Knøsgaard: **l.knoesgaard@rn.dk**

Abstract

Objective: The assessment of maternal thyroid function in early pregnancy is debated. It is well-established that pregnancy-specific reference ranges preferably should be used. We speculated if the use of repeated blood samples drawn in early pregnancy would influence the classification of maternal thyroid function.
Methods: Pregnant women with repeated early pregnancy blood samples were identified in the North Denmark Region Pregnancy Cohort. Each sample was used for the measurement of TSH, free T4 (fT4), thyroid peroxidase antibodies (TPO-Ab), and thyroglobulin antibodies (Tg-Ab) (ADVIA Centaur XPT, Siemens Healthineers). Method- and pregnancy week-specific reference ranges were used for the classification of maternal thyroid function.
Results: Among 1466 pregnancies included, 89 women had TSH above the upper reference limit in the first sample (median pregnancy week 8) and 44 (49.4%) of these similarly had high TSH in the second sample (median week 10). A total of 47 women had TSH below the lower reference limit in the first sample and 19 (40.4%) of these similarly had low TSH in the second sample. Regarding women classified with isolated changes in fT4 in the first sample, less than 20% were similarly classified as such in the second sample. The percentage agreement between the samples was dependent on the level of TSH in the first sample and the presence of TPO- and Tg-Ab.
Conclusion: In a large cohort of pregnant women, the classification of maternal thyroid function varied considerably with the use of repeated blood samples. Results emphasize a focus on the severity of thyroid function abnormalities in pregnant women.

Introduction

The diagnosis and treatment of thyroid disease during pregnancy are debated and uncertainties exist regarding the assessment of maternal thyroid function in early pregnancy (1, 2). During a normal pregnancy, several physiological changes in maternal thyroid function occur, and it is well-established that these changes necessitate pregnancy-specific reference ranges for thyroid function tests in pregnant women (3). The use of trimester-specific reference ranges is generally recommended (3, 4), but the assessment in the first trimester is a particular challenge because maternal TSH levels show considerable dynamics within the early pregnancy (5, 6, 7). Recently, we reported

the frequency of thyroid function abnormalities in the North Denmark Region Pregnancy Cohort (NDRPC) from a single early pregnancy blood sample using method- and pregnancy week-specific reference ranges (8). Furthermore, we found considerable variation in the classification of thyroid function abnormalities in the NDRPC when different analytical methods and thyroid function tests were used, even when the reference ranges were method- and pregnancy week-specific (7). A recent study reported variation in the classification of maternal thyroid function when blood samples were repeated in short time intervals within the same day (9) and added to the knowledge on physiological variability in thyroid function tests (10). These findings led us to speculate if maternal thyroid function abnormalities would persist with repeated blood samples within some weeks in early pregnancy and if the persistency would be related to the severity of thyroid function abnormalities. Such results may have implications for the clinical diagnosis of thyroid disease in pregnant women and for the assessment of maternal thyroid function in scientific outcome studies.

The NDRPC included a substantial number of pregnancies in which the pregnant woman had a blood sample drawn twice in the early pregnancy. For the present study, we identified these pregnancies and aimed to describe and compare the classification of maternal thyroid function in each sample still using method- and pregnancy week-specific reference ranges.

Materials and methods

Study design

This is a retrospective cohort study within the NDRPC. The NDRPC was established from 2011–2015 and includes a biobank of stored early pregnancy blood samples drawn as part of the prenatal screening program for chromosomal anomalies (6). After ethical approval, biochemical measurements of maternal thyroid function and thyroid autoantibodies were performed and results were linked to information in the Danish nationwide registers including the Medical Birth Register (11), the Danish National Hospital Register (12), the Danish National Prescription Register (13), and demographics available at Statistics Denmark. As previously described (6, 8), data linkage provided information on pregnancy week of blood sampling, maternal age, origin, pre-pregnancy BMI, parity, smoking status, and fetal gender, as well as hospital diagnoses and redeemed prescriptions of drugs. The study was approved by the North Denmark Region Committee

on Health Research Ethics (N-20150015) and the Danish Data Protection Agency (J. no. 2008-58-0028).

Study population

All pregnant women in Denmark are offered prenatal screening for chromosomal anomalies that includes a blood sample drawn in pregnancy weeks 9–14 followed by obstetric ultrasound in weeks 12–15 (14). In some cases, the blood sample drawn as part of the screening program is repeated at the time of obstetric ultrasound to ensure the correct timing of the sample within the recommended weeks of pregnancy. As a result, the NDRPC includes a subgroup of pregnant women with repeated blood sampling in the early pregnancy ($n = 1466$) in addition to the large group of women with a single blood sample from the early pregnancy ($n = 14,169$). Multiple pregnancies ($n = 321$) and women receiving medical treatment for thyroid disease in the pregnancy prior to blood sampling ($n = 196$) were not included. None of the women in the repeated samples cohort started medical treatment for thyroid disease in the pregnancy weeks between the repeated blood samples.

Thyroid function and autoimmunity

Serum residues from the blood samples were stored at −80°C until biochemical analyses of TSH, free T4 (fT4), thyroid peroxidase antibodies (TPO-Ab), and thyroglobulin antibodies (Tg-Ab). The analyses were performed in 2015–2016 in a routine hospital laboratory using ADVIA Centaur XPT (Siemens Healthineers) as previously described (6). Maternal thyroid function was classified in each sample using method- and pregnancy week-specific reference ranges for TSH and fT4 (defined as the 2.5th–97.5th percentiles) established within the NDRPC (6). As recommended (3), the reference ranges were established among TPO- and Tg-Ab negative women with no known thyroid or other autoimmune disease. Stratification by week of pregnancy was used to account for the physiological dynamics in maternal thyroid function within the first trimester, but specific adjustment for maternal human chorionic gonadotropin levels or other determinants was not performed. Overall, abnormal maternal thyroid function was defined when TSH or fT4 was outside the method- and pregnancy week-specific reference ranges. Biochemical hyperthyroidism or hypothyroidism was then defined by TSH below or above the reference ranges, respectively. Isolated changes in fT4 were defined by abnormal fT4 and TSH within the week-specific reference ranges. Further classification of overt

(abnormal TSH and fT4) and subclinical (isolated abnormal TSH) abnormalities was defined by the combined use of TSH and fT4. After the classification of maternal thyroid function in each blood sample, women with repeated blood sampling were grouped according to the combined classification of maternal thyroid function (normal vs abnormal) in samples 1 and 2. Thus, a group of women ($n = 1215$) had similar classification (samples agree) and the remaining ($n = 251$) had non-similar classification (samples disagree) across the repeated samples. Finally, results of maternal thyroid autoantibodies were included in the evaluation, and cut-off values of 60 U/mL given by the manufacturer for TPO- and Tg-Ab were applied.

Statistical analyses

Continuous variables were described by the median and interquartile range (IQR), whereas categorical variables were described by the number (n) and frequency (%). Percentage agreement in the classification of maternal thyroid function and thyroid autoimmunity between the first (sample 1) and second blood sample (sample 2) was calculated with a 95% CI using the binomial exact model with sample 1 as the reference sample. Mann–Whitney

U test and chi-squared test were used to compare continuous and categorical variables, respectively, between unpaired groups. Wilcoxon signed-rank test and McNemar's test were used for the comparison of paired continuous and categorical variables, respectively, between samples 1 and 2 within the repeated samples cohort. Statistical analyses were performed using Stata 16.1 (StataCorp LLC, College Station, TX, USA) with a 5% level of significance.

Results

Altogether, 15,635 singleton pregnancies were included in the study as part of the single sample cohort ($n = 14,169$) or the repeated samples cohort ($n = 1466$) (Table 1). Pregnancy week of blood sampling was within the recommended weeks in the single sample cohort, whereas sample 1 in the repeated samples cohort as expected was drawn too early (Table 1). Considering maternal characteristics, women with repeated samples tended to be younger, had slightly higher pre-pregnancy BMI, and were more often nulliparous and smoking (Table 1). On the other hand, maternal characteristics were similar among women with repeated samples when stratified in groups according to

Table 1 Maternal characteristics of women with either a single or repeated early pregnancy blood sample ($n = 15,635$).

	Single sample cohort		Repeated samples cohort					
			All		Samples agree[a]		Samples disagree[b]	
All pregnancies (*n*)	**14,169**		**1466**		**1215**		**251**	
Week of blood sampling (median, IQR)								
Sample 1	10	9–11	8	7–8	8	8–8	8	7–8
Sample 2	–	–	12	11–13	12	11–13	11	11–12
Age in years (median, IQR)	29.9	26.6–33.5	28.4	25.6–32.0	28.3	25.5–31.9	28.8	26.1–32.1
Origin[c] (*n*, %)								
Born in Denmark	12,523	88.6	1308	89.2	1083	89.1	225	89.6
Not born in Denmark	1609	11.4	158	10.8	132	10.9	26	10.4
All births[d] (*n*)	**12,948**		**1450**		**1202**		**248**	
Pre-pregnancy BMI in kg/m²[c] (median, IQR)	23.7	21.3–27.4	24.1	21.2–28.7	24.1	21.3–28.4	24.3	21.1–29.8
Parity[c] (*n*, %)								
Nulliparous	5939	46.1	742	51.3	626	52.3	116	47.0
Multiparous	6955	53.9	703	48.7	572	47.7	131	53.0
Smoking in pregnancy[c] (*n*, %)								
No smoking	11,417	88.6	1238	85.8	1028	85.9	210	85.0
Smoking	1465	11.4	205	14.2	168	14.1	37	15.0
Fetal gender[c] (*n*, %)								
Female	6326	48.9	708	48.9	587	48.8	121	49.0
Male	6615	51.1	741	51.1	615	51.2	126	51.0

[a]Similar classification of maternal thyroid function (normal or abnormal) in samples 1 and 2; [b]Non-similar classification of maternal thyroid function in samples 1 and 2; [c]Women with missing information not included; origin ($n = 37$), pre-pregnancy BMI ($n = 65$), parity ($n = 59$) smoking ($n = 73$), and fetal gender ($n = 8$); [d]Information on maternal pre-pregnancy BMI, parity, smoking, and fetal gender was only available for births (live births and stillbirths). IQR, interquartile range.

the combined classification of abnormal maternal thyroid function across the samples (Table 1). Median TSH was 1.09 (IQR: 0.64–1.65) mIU/L and fT4 was 16.0 (IQR: 14.7–17.3) pmol/L in the single sample cohort. In the repeated samples cohort, median TSH was 1.34 (IQR: 0.88–1.93) mIU/L and 1.12 (IQR: 0.69–1.74) mIU/L and fT4 was 15.8 (IQR: 14.6–17.1) pmol/L and 16.1 (IQR: 14.8–17.3) pmol/L in samples 1 and 2, respectively. When medians of TSH and fT4 in the repeated samples were compared, TSH was higher ($P < 0.001$) and fT4 was lower ($P < 0.001$) in sample 1 compared to sample 2.

Considering the classifications of maternal thyroid function abnormalities and markers of thyroid autoimmunity, the frequencies were overall similar in the cohorts (Table 2). Within the repeated samples cohort, a higher frequency of TPO- and/or Tg-Ab positive as well as Tg-Ab positive women in sample 1 was observed in comparison to sample 2 (Table 2). When the percentage agreement between samples 1 and 2 was assessed, the overall agreement was less than 40% meaning that only two in five women classified with abnormal thyroid function in sample 1 had persistent abnormal thyroid function in sample 2 (Table 2). More specifically, less than 50% of women classified with hyper- or hypothyroidism in sample 1 similarly had such thyroid function abnormality in sample 2, and the percentage agreement for the classification of isolated changes in fT4 in samples 1 and 2

was less than 20%, whereas, for the classification of being TPO- or Tg-Ab positive, the percentage agreement was nearly 90% (Table 2).

Characteristics and biochemical results for women in the repeated samples cohort who were classified with either hyperthyroidism ($n = 47$) or hypothyroidism ($n = 89$) in sample 1 were subsequently described when stratified by the combined classification in samples 1 and 2 (Table 3). Maternal age, pre-pregnancy BMI, and pregnancy week of blood sampling did not differ between agreement groups (samples agree vs samples disagree), but differences were observed regarding the biochemical results of thyroid function tests (Table 3). Thus, TSH was lower and fT4 was higher among women classified with hyperthyroidism in both samples, whereas TSH was higher and fT4 was lower among women classified with hypothyroidism in both samples (Table 3).

Subsequently, the percentage agreement between samples 1 and 2 was evaluated according to the level of TSH in sample 1 (Fig. 1). Notably, the percentage agreement was dependent on the level of TSH in a U-shaped manner (Fig. 1). For the classification of hyperthyroidism, the percentage agreement increased from 40 to 80% with decreasing TSH in sample 1 (Fig. 1). For the classification of hypothyroidism, the percentage agreement increased from 50 to 100% with increasing TSH in sample 1 (Fig. 1) and from nearly 70 to 100% when evaluated among women

Table 2 Frequencies of abnormal maternal thyroid function and thyroid autoimmunity in the blood samples and the percentage agreement across repeated samples.

| | Single sample cohort ($n = 14,169$) | | Repeated samples cohort ($n = 1466$) | | | | | | | |
| | | | Sample 1[a] | | Sample 2[a] | | Samples 1 and 2[b] | | Agreement | |
	n	%[c]	n	%[d]	n	%[d]	n	%[d]	%[e]	95% CI
Abnormal thyroid function[f]	**2072**	**14.6**	**222**	**15.1**	**203**	**13.9**	**87**	**5.9**	**39.2**	**32.7–45.9**
Hyperthyroidism	516	3.6	47	3.2	47	3.2	19	1.3	40.4	26.4–55.7
Overt	203	1.4	18	1.2	24	1.6	8	0.5	44.4	21.5–69.2
Subclinical	313	2.2	29	2.0	23	1.6	5	0.3	17.2	5.8–35.8
Hypothyroidism	843	6.0	89	6.1	80	5.5	44	3.0	49.4	38.7–60.2
Overt	139	1.0	11	0.8	10	0.7	5	0.3	45.5	16.7–76.6
Subclinical	704	5.0	78	5.3	70	4.8	32	2.2	41.0	30.0–52.7
Isolated high fT4	221	1.6	36	2.5	34	2.3	7	0.5	19.4	8.2–36.0
Isolated low fT4	492	3.5	50	3.4	42	2.9	9	0.6	18.0	8.6–31.4
TPO- and/or Tg-Ab positive	**2048**	**14.5**	**208**	**14.2**	**188**	**12.8**	**182**	**12.4**	**87.5**	**82.2–91.7**
TPO-Ab positive	1489	10.5	135	9.2	128	8.7	119	8.1	88.1	81.5–93.1
Tg-Ab positive	1441	10.2	156	10.6	140	9.5	138	9.4	88.5	82.4–93.0

[a]For each classification of maternal thyroid function, the frequencies in sample 1 and sample 2 were compared using McNemar's test: $P < 0.05$ for TPO- and/or Tg-Ab positive as well as Tg-Ab positive women; [b]Similar classification of abnormal thyroid function in samples 1 and 2 (samples agree); [c]Proportion of all within the single sample cohort; [d]Proportion of all within the repeated samples cohort; [e]Proportion of samples with similar classification in samples 1 and 2 classified as such in sample 1 (percentage agreement); [f]Defined as TSH or fT4 outside the method- and pregnancy week-specific reference ranges.

fT4, free T4; Tg-Ab, thyroglobulin antibodies; TPO-Ab, thyroid peroxidase antibodies.

Table 3 Characteristics and biochemical results for women with repeated samples classified with either hyperthyroidism (n = 47) or hypothyroidism (n = 89) in sample 1.

| | Hyperthyroidism | | | | | Hypothyroidism | | | | |
| | Samples agree[a] (n = 19) | | Samples disagree[b] (n = 28) | | P^c | Samples agree[b] (n = 44) | | Samples disagree[b] (n = 45) | | P^c |
	Median	IQR	Median	IQR		Median	IQR	Median	IQR	
Age (years)	30.1	27.1–31.1	30.4	27.6–33.8	0.648	28.8	25.4–31.7	28.9	26.2–31.5	0.758
Pre-pregnancy BMI (kg/m²)	23.9	21.6–25.0	22.2	20.0–26.4	0.485	26.6	21.8–30.8	24.1	20.6–32.4	0.449
Week of blood sampling										
Sample 1	8	7–9	8	8–8	0.409	8	8–9	8	7–8	0.765
Sample 2	11	11–13	12	10–12	0.784	12	11–12	12	11–13	0.306
TSH (mIU/L)										
Sample 1	0.08	0.01–0.18	0.21	0.14–0.27	0.001	4.49	4.00–6.17	3.66	3.38–4.31	<0.001
Sample 2	0.02	0.01–0.06	0.23	0.09–0.41	<0.001	3.97	3.26–4.88	2.17	1.89–2.49	<0.001
Δ[d]	0.03	0.00–0.10	−0.04	−0.17–0.11	0.045	0.94	−0.09–2.17	1.80	1.18–2.28	0.005
fT4 (pmol/L)										
Sample 1	21.1	17.8–22.5	18.9	17.1–20.0	0.022	14.2	13.3–15.4	15.6	14.5–16.8	0.001
Sample 2	20.7	18.6–22.4	17.1	16.1–17.9	<0.001	15.3	13.5–16.2	15.9	15.0–17.2	0.018
Δ[d]	1.05	−4.18–3.97	1.35	0.14–3.30	0.551	−0.73	−1.79–0.20	−0.24	−2.07–1.02	0.468

[a]Similar classification of maternal thyroid function in samples 1 and 2; [b]Non-similar classification of maternal thyroid function in samples 1 and 2; [c]P-values are results of comparison between groups within hyperthyroidism and hypothyroidism (samples agree vs samples disagree); [d]Computed as the difference between the two samples (Δ = sample1–sample2).
fT4, free T4; IQR, interquartile range.

who were TPO-Ab positive (Fig. 2A) or Tg-Ab positive (Fig. 2B) in sample 1. Regardless of the thyroid autoimmunity status, the percentage agreement was 100% when TSH in sample 1 was above 7 mIU/L (Figs 1 and 2).

Finally, the frequency of abnormal maternal thyroid function and the level of agreement with repeated samples were evaluated when the women were grouped according to their thyroid autoimmunity status (Table 4). Notably, the percentage agreement in the classification of abnormal maternal thyroid function was higher among antibody-positive women and especially among TPO-Ab positive women (Table 4). Considering maternal characteristics and levels of TSH and fT4, no differences were observed between agreement groups (samples agree vs samples disagree) for TPO-Ab negative (Supplementary Table 1, see section on supplementary materials given at the end of this article) or Tg-Ab negative women (Supplementary Table 2). However, in line with the tendency shown in Fig. 2, antibody-positive women with agreement in the classification of abnormal thyroid function had higher median level of TSH in both samples ($P < 0.05$) compared to the women with disagreement (Supplementary Tables 1 and 2).

Discussion

Principal findings

In a cohort of Danish pregnant women, we had the opportunity to investigate the classification of maternal thyroid function in early pregnancy when repeated blood samples within a few weeks were considered. Even though method- and pregnancy week-specific reference ranges were used, the classification of maternal thyroid function varied considerably with repeated blood samples. The percentage agreement between samples was low for all types of maternal thyroid function abnormalities, and it was a notable finding that the level of agreement was dependent on the degree of initial TSH deviation and the presence of thyroid autoantibodies.

Interpretation

The pregnancy-related physiological changes in maternal thyroid function necessitate the use of pregnancy-specific reference ranges for the interpretation of thyroid function tests (3). Recommendations on trimester-specific reference ranges are widespread (3, 4). However, weekly dynamics in TSH within the early pregnancy have been illustrated
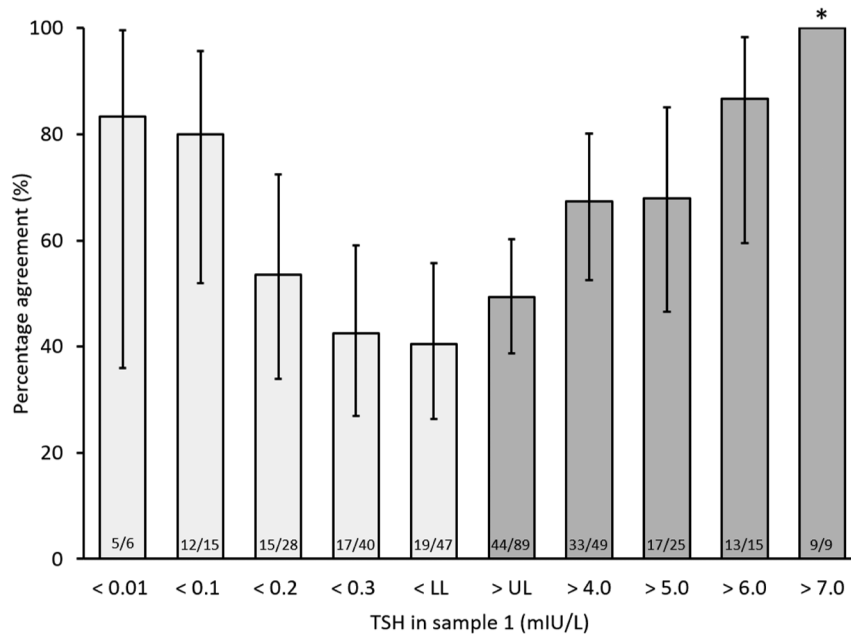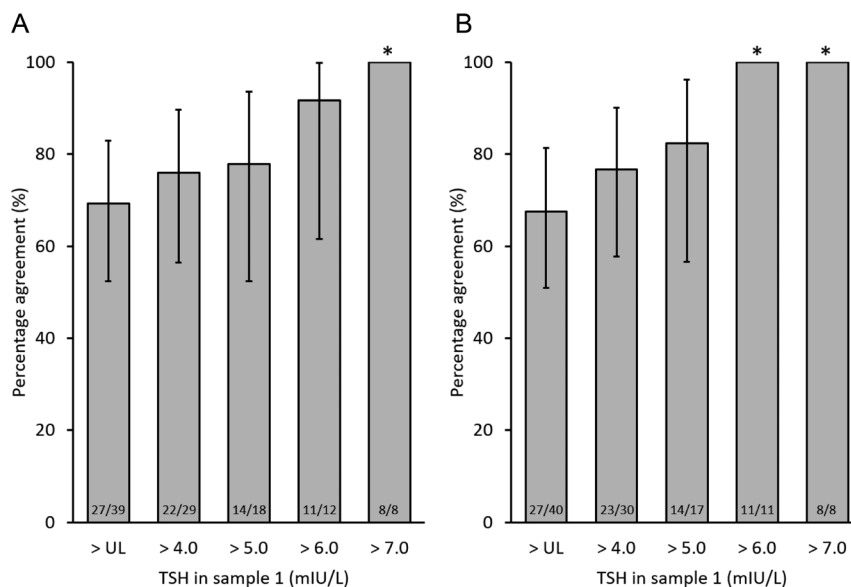
Figure 1
Percentage agreement (with 95% CIs) in the classification of abnormal maternal thyroid function between samples 1 and 2 stratified by level of TSH in sample 1. Light gray bars represent the percentage agreement in the classification of hyperthyroidism and dark gray bars represent the agreement in the classification of hypothyroidism. Numbers (*n/n*) represent samples with similar classification in sample 1 and 2 (samples agree) out of all samples classified as such in sample 1. LL, lower pregnancy week-specific reference limits; UL, upper pregnancy week-specific reference limits; *100% agreement, no 95% CI.

in different cohorts, suggesting that the use of uniform trimester-specific reference ranges may be too simple (5, 6, 7). In line with these results, we observed different levels of median TSH and fT4 between samples 1 and 2 in the repeated samples cohort and we accounted for these weekly dynamics in maternal thyroid function by the application of method- and pregnancy week-specific reference ranges established within the cohort (6). Furthermore, the indirect measurement of free thyroid hormones using automatic immunoassays is prone to flaws given the pregnancy-related alterations in binding proteins (15, 16) and stress the importance of method-

and pregnancy-specific reference ranges (3). Among a subgroup of women in the NDRPC, we recently measured a series of thyroid function tests using another analytical method (Cobas 8000, Roche Diagnostics), and pregnancy week-specific reference ranges for this analytical method were established (7). When the different method-specific reference ranges were used to classify maternal thyroid function, the classification varied with the analytical method and the type of thyroid function test (7). In accordance with the present study, the variation was most pronounced for the identification of isolated changes in fT4 (7). Thus, in our previous investigation and in the present

Figure 2
Percentage agreement (with 95% CIs) in the classification of TPO-Ab positive (A) or Tg-Ab-positive (B) hypothyroidism between samples 1 and 2 stratified by level of TSH in sample 1. Numbers (*n/n*) represent samples with TPO- or Tg-Ab positive hypothyroidism in samples 1 and 2 (samples agree) out of all samples classified as such in sample 1. UL, upper pregnancy week-specific reference limits; *100% agreement, no 95% CI.

Table 4 Frequencies of abnormal maternal thyroid function and thyroid autoimmunity in the blood samples and the percentage agreement across repeated samples when stratified by maternal status of thyroid autoimmunity.

	Single sample cohort			Repeated samples cohort									
	All	Abnormal thyroid function[a]		All	Abnormal thyroid function[a]							Agreement	
					Sample 1[b]		Sample 2[b]		Samples 1 and 2[c]				
	n	n	%[d]	n[e]	n	%[f]	n	%[f]	n	%[f]		%[g]	95% CI
TPO- and Tg-Ab negative	12,121	1459	12.0	1258	154	12.2	153	12.2	49	3.9		31.8	24.6–39.8
TPO- and/or Tg-Ab positive	2048	613	29.9	208	68	32.7	50	24.0	38	18.3		55.9	43.3–67.9
TPO-Ab negative	12,680	1572	12.4	1331	173	13.0	161	12.1	56	4.2		32.4	25.5–39.9
TPO-Ab positive	1489	500	33.6	135	49	36.3	42	31.1	31	23.0		63.3	48.3–76.6
Only TPO-Ab positive	607	149	24.6	52	15	28.9	11	21.2	8	15.4		53.3	26.6–78.7
Tg-Ab negative	12,728	1608	12.6	1310	169	12.9	164	12.5	57	4.4		33.7	26.6–41.4
Tg-Ab positive	1441	464	32.2	156	53	34.0	39	25.0	30	19.2		56.6	42.3–70.2
Only Tg-Ab positive	559	113	20.2	73	19	26.0	8	11.0	7	9.6		36.8	16.3–61.6

aDefined as TSH or fT4 outside the method- and pregnancy week-specific reference ranges; bFor each group of thyroid autoimmunity, the frequency of abnormal thyroid function in sample 1 was compared to sample 2 using McNemar's test. $P < 0.05$ for TPO- and/or Tg-Ab positive, Tg-Ab positive, as well as only Tg-Ab positive women; cClassification of abnormal thyroid function in samples 1 and 2 (samples agree); dFrequency of abnormal thyroid function among women with the specific status of thyroid autoimmunity within the single sample cohort; eFor women in the repeated samples cohort, the thyroid autoimmunity status was classified from the level of TPO- and/or Tg-Ab in sample 1; fFrequency of abnormal thyroid function among women with the specific status of thyroid autoimmunity within the repeated samples cohort; gProportion of samples with abnormal thyroid function in samples 1 and 2 classified as such in sample 1 (percentage agreement).
Tg-Ab, thyroglobulin antibodies; TPO-Ab, thyroid peroxidase antibodies.

study, it was a notable finding that the classification of maternal thyroid function varied considerably even when method- and pregnancy week-specific reference ranges were established and used to account for the expected methodological differences and the known physiological alterations in maternal thyroid function within the early pregnancy.

How to define a reference cohort for the establishment of pregnancy-specific reference ranges for thyroid function tests in pregnant women is a matter of debate, and genetic factors as well as a series of other determinants are considered (17, 18, 19). The method- and pregnancy week-specific reference ranges used in the present study (6) were established among TPO- and Tg-Ab negative pregnant women with no previous or current thyroid or other autoimmune disease in accordance with current guideline recommendations for the selection of a reference cohort (3). However, it should be acknowledged that other variables may be of importance in the assessment of reference ranges (17, 18, 19).

Sparse data are available on the impact of repeated blood sampling for the classification of maternal thyroid function in early pregnancy. On the other hand, most studies rely on the definition of exposure to maternal thyroid dysfunction from a single blood sample drawn in early pregnancy. Recently, Fan *et al.* reported the persistency of thyroid dysfunction from early to late pregnancy to be less than 25%, in a large study based on blood samples drawn as part of first-trimester screening and a late pregnancy follow-up (20). However, opposite to the retrospective design of our study using stored biobank samples, this study was prospective in which some women with abnormal thyroid function test results in early pregnancy started medical treatment and were excluded from the final analyses on persistency of thyroid dysfunction (20). In another recent study, Lewandowski *et al.* assessed the variation in maternal TSH within one day using a series of five blood samples drawn in the morning with 30-min intervals from 110 healthy first-trimester pregnant women (9). In line with our findings, the study demonstrated a considerable variation in TSH across the samples with a variation of up to 40% between the highest and lowest individual TSH (9). They further illustrated that the variation influenced the classification of maternal thyroid function (9). Notably, the disparity in the classification of maternal thyroid function across the samples was most pronounced in TPO-Ab negative women (9), which is in accordance with our findings in the present study.

Implication

Overt thyroid disease should be treated to prevent maternal and fetal complications (3). On the other hand, recommendations regarding slightly elevated TSH in pregnant women with or without the presence of TPO-Ab are less clear and treatment may be considered in some cases (3). Much focus regarding thyroid function in pregnant women has been on the role of smaller abnormalities in maternal thyroid function including TSH within the upper reference range and isolated changes in fT4 as well as the role of thyroid autoimmunity per se (2, 3). Although many observational studies have found associations between smaller abnormalities in maternal thyroid function and adverse pregnancy and child outcomes (3), these findings are in contrast to the results of large, randomized controlled trials (RCTs) (21, 22, 23, 24, 25, 26, 27, 28). Notably, the classification of maternal thyroid function in the vast majority of observational studies and in the RCTs relied on a single blood sample (21, 22, 23, 24, 25, 26, 27, 28). Considering the variation observed in our study and in the study by Lewandowsky *et al.* (9), one may speculate on the definition of exposure and randomization to treatment based on smaller abnormalities in maternal thyroid function and if this contributes to the finding of lack of treatment effect. Adding to this, the dependency on the level of TSH and the presence of TPO- and Tg-Ab observed in our study substantiate a clinical and scientific focus on marked thyroid function abnormalities in pregnant women (2).

Methodological comments

Our study was designed within a large birth cohort, and the inclusion of a substantial number of pregnant women with repeated blood samples from the early pregnancy was unique. Moreover, the selection of participants was performed from blood samples drawn as part of the Danish prenatal nationwide screening program for chromosomal anomalies in which the participation rate is high (14). The biochemical analyses of TSH, fT4, and the thyroid autoantibodies were performed retrospectively in the years following pregnancy termination and thereby unrelated to any clinical practice regarding diagnosis or monitoring of maternal thyroid disease during pregnancy. Furthermore, the main cause for the drawing of a repeated blood sample in relation to the screening program is to ensure that the timing of the blood sampling is within the recommended weeks of pregnancy (14). Thus, the selection of women with a repeated blood sample is likely non-differential according to the classification of maternal thyroid function, but

some differences were observed in maternal characteristics and in the frequencies of thyroid function abnormalities between the single sample and the repeated samples cohort and risk of selection bias cannot be excluded. On the other hand, no differences were observed in maternal characteristics between agreement groups (samples agree vs samples disagree) within the repeated samples cohort. Regarding thyroid function and autoimmunity, only differences in the frequencies of TPO- and/or Tg-Ab as well as Tg-Ab positive women were observed between samples 1 and 2 in the repeated samples cohort. We speculate if this reflects the use of antibody cut-offs that were not established among pregnant women specifically. Our study was based on consecutively collected blood samples stored in a biobank, and thyroid function tests as well as autoantibodies are known to be stable for decades after frozen storage (29). Regarding the individual dynamics in thyroid function tests between the first and the second sample, we acknowledge the statistical phenomenon of regression towards the mean (30). However, even though regression towards the mean may be present, our aim was not, and our data are not designed to add to this discussion. Furthermore, this should not change the disparity in the classification observed and the implication of our findings.

Conclusion

Accurate diagnosis of thyroid function abnormalities in pregnant women is important in clinical and scientific practice. When method- and pregnancy-specific reference ranges were established and used in a large cohort of pregnant women, the classification of abnormal thyroid function in early pregnancy varied considerably across repeated blood samples. The agreement was lowest for smaller thyroid function aberrations while increasing TSH and the presence of thyroid autoantibodies improved the agreement. Results substantiate a focus on the severity of thyroid dysfunction and call for further investigations on the method of thyroid function assessment in pregnant women.

References

1 Laurberg P, Andersen SL, Pedersen IB, Andersen S & Carlé A. Screening for overt thyroid disease in early pregnancy may be preferable to searching for small aberrations in thyroid function tests. *Clinical Endocrinology* 2013 **79** 297–304. (https://doi.org/10.1111/cen.12232)

2 Andersen SL & Andersen S. Turning to thyroid disease in pregnant women. *European Thyroid Journal* 2020 **9** 225–233. (https://doi.org/10.1159/000506228)

3 Alexander EK, Pearce EN, Brent GA, Brown RS, Chen H, Dosiou C, Grobman WA, Laurberg P, Lazarus JH, Mandel SJ, *et al.* 2017 guidelines of the American Thyroid Association for the diagnosis and management of thyroid disease during pregnancy and the postpartum. *Thyroid* 2017 **27** 315–389. (https://doi.org/10.1089/thy.2016.0457)

4 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. *European Thyroid Journal* 2014 **3** 76–94. (https://doi.org/10.1159/000362597)

5 Laurberg P, Andersen SL, Hindersson P, Nohr EA & Olsen J. Dynamics and predictors of serum TSH and fT4 reference limits in early pregnancy: a study within the Danish National Birth Cohort. *Journal of Clinical Endocrinology and Metabolism* 2016 **101** 2484–2492. (https://doi.org/10.1210/jc.2016-1387)

6 Andersen SL, Andersen S, Carlé A, Christensen PA, Handberg A, Karmisholt J, Knøsgaard L, Kristensen SR, Bülow Pedersen I & Vestergaard P. Pregnancy week-specific reference ranges for thyrotropin and free thyroxine in the North Denmark region pregnancy cohort. *Thyroid* 2019 **29** 430–438. (https://doi.org/10.1089/thy.2018.0628)

7 Andersen SL, Christensen PA, Knøsgaard L, Andersen S, Handberg A, Hansen AB & Vestergaard P. Classification of thyroid dysfunction in pregnant women differs by analytical method and type of thyroid function test. *Journal of Clinical Endocrinology and Metabolism* 2020 **105** e4012–e4022. (https://doi.org/10.1210/clinem/dgaa567)

8 Knøsgaard L, Andersen S, Hansen AB, Vestergaard P & Andersen SL. Thyroid function abnormalities and thyroid autoantibodies in Danish pregnant women. *Clinical Endocrinology* 2020 **93** 329–338. (https://doi.org/10.1111/cen.14147)

9 Lewandowski KC, Garnysz K, Horzelski W, Kawalec J, Budzen K, Grzesiak M & Lewinski A. Subclinical thyroid dysfunction in the first trimester of pregnancy: 'disease' versus physiological (pulsatile) variation in TSH concentrations. *Clinical Endocrinology* 2020 **93** 739–745. (https://doi.org/10.1111/cen.14256)

10 Boas M, Forman JL, Juul A, Feldt-Rasmussen U, Skakkebaek NE, Hilsted L, Chellakooty M, Larsen T, Larsen JF, Petersen JH, *et al.* Narrow intra-individual variation of maternal thyroid function in pregnancy based on a longitudinal study on 132 women. *European Journal of Endocrinology* 2009 **161** 903–910. (https://doi.org/10.1530/EJE-09-0579)

11 Knudsen LB & Olsen J. The Danish medical birth registry. *Danish Medical Bulletin* 1998 **45** 320–323.

12 Schmidt M, Schmidt SAJ, Sandegaard JL, Ehrenstein V, Pedersen L & Sørensen HT. The Danish National Patient Registry: a review of content, data quality, and research potential. *Clinical Epidemiology* 2015 **7** 449–490. (https://doi.org/10.2147/CLEP.S91125)

13 Kildemoes HW, Sørensen HT & Hallas J. The Danish National prescription registry. *Scandinavian Journal of Public Health* 2011 **39** (Supplement) 38–41. (https://doi.org/10.1177/1403494810394717)

14 Danish Health Authority. *Retningslinjer for Fosterdiagnostik.* Copenhagen: Danish Health Authority, 2017.

15 Lee RH, Spencer CA, Mestman JH, Miller EA, Petrovic I, Braverman LE & Goodwin TM. Free T4 immunoassays are flawed during pregnancy. *American Journal of Obstetrics and Gynecology* 2009 **200** 260.e1–260.e6. (https://doi.org/10.1016/j.ajog.2008.10.042)

16 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* 2014 **170** 329–339. (https://doi.org/10.1530/EJE-13-0672)

17 Korevaar TIM, Steegers EAP, Pop VJ, Broeren MA, Chaker L, de Rijke YB, Jaddoe VWV, Medici M, Visser TJ, Tiemeier H, *et al.* Thyroid autoimmunity impairs the thyroidal response to human chorionic gonadotropin: two population-based prospective cohort studies. *Journal of Clinical Endocrinology and Metabolism* 2017 **102** 69–77. (https://doi.org/10.1210/jc.2016-2942)

18 Medici M, Visser TJ & Peeters RP. Genetics of thyroid function. *Best Practice and Research: Clinical Endocrinology and Metabolism* 2017 **31** 129–142. (https://doi.org/10.1016/j.beem.2017.04.002)

19 Veltri F & Poppe K. Variables contributing to thyroid (Dys)function in pregnant women: more than thyroid antibodies? *European Thyroid Journal* 2018 **7** 120–128. (https://doi.org/10.1159/000488279)

20 Fan J, Zhang Y, Zhang C, Barjaktarovic M, Yang X, Peeters RP, Huang HF & Korevaar TIM. Persistency of thyroid dysfunction from early to late pregnancy. *Thyroid* 2019 **29** 1475–1484. (https://doi.org/10.1089/thy.2019.0115)

21 Negro R, Formoso G, Mangieri T, Pezzarossa A, Dazzi D & Hassan H. Levothyroxine treatment in euthyroid pregnant women with autoimmune thyroid disease: effects on obstetrical complications. *Journal of Clinical Endocrinology and Metabolism* 2006 **91** 2587–2591. (https://doi.org/10.1210/jc.2005-1603)

22 Negro R, Schwartz A, Gismondi R, Tinelli A, Mangieri T & Stagnaro-Green A. Universal screening versus case finding for detection and treatment of thyroid hormonal dysfunction during pregnancy. *Journal of Clinical Endocrinology and Metabolism* 2010 **95** 1699–1707. (https://doi.org/10.1210/jc.2009-2009)

23 Lazarus JH, Bestwick JP, Channon S, Paradice R, Maina A, Rees R, Chiusano E, John R, Guaraldo V, George LM, *et al.* Antenatal thyroid screening and childhood cognitive function. *New England Journal of Medicine* 2012 **366** 493–501. (https://doi.org/10.1056/NEJMoa1106104)

24 Casey BM, Thom EA, Peaceman AM, Varner MW, Sorokin Y, Hirtz DG, Reddy UM, Wapner RJ, Thorp Jr JM, Saade G, *et al.* Treatment of subclinical hypothyroidism or hypothyroxinemia in pregnancy. *New England Journal of Medicine* 2017 **376** 815–825. (https://doi.org/10.1056/NEJMoa1606205)

25 Nazarpour S, Ramezani FT, Simbar M, Tohidi M, Alavi Majd H & Azizi F. Effects of levothyroxine treatment on pregnancy outcomes in pregnant women with autoimmune thyroid disease. *European Journal of Endocrinology* 2017 **176** 253–265. (https://doi.org/10.1530/EJE-16-0548)

26 Nazarpour S, Ramezani FT, Simbar M, Tohidi M, Minooee S, Rahmati M & Azizi F. Effects of levothyroxine on pregnant women with subclinical hypothyroidism, negative for thyroid peroxidase antibodies. *Journal of Clinical Endocrinology and Metabolism* 2018 **103** 926–935. (https://doi.org/10.1210/jc.2017-01850)

27 Wang H, Gao H, Chi H, Zeng L, Xiao W, Wang Y, Li R, Liu P, Wang C, Tian Q, *et al.* Effect of levothyroxine on miscarriage among women with normal thyroid function and thyroid autoimmunity undergoing in vitro fertilization and embryo transfer: a randomized clinical trial. *JAMA* 2017 **318** 2190–2198. (https://doi.org/10.1001/jama.2017.18249)

28 Dhillon-Smith RK, Middleton LJ, Sunner KK, Cheed V, Baker K, Farrell-Carver S, Bender-Atik R, Agrawal R, Bhatia K, Edi-Osagie E, *et al.* Levothyroxine in women with thyroid peroxidase antibodies before conception. *New England Journal of Medicine* 2019 **380** 1316–1325. (https://doi.org/10.1056/NEJMoa1812537)

29 Männistö T, Suvanto E, Surcel HM & Ruokonen A. Thyroid hormones are stable even during prolonged frozen storage. *Clinical Chemistry and Laboratory Medicine* 2010 **48** 1669–1670; author reply 1671. (https://doi.org/10.1515/CCLM.2010.324)

30 Bland JM & Altman DG. Regression towards the mean. *BMJ* 1994 **308** 1499. (https://doi.org/10.1136/bmj.308.6942.1499)