Gestational age-specific reference ranges from different laboratories misclassify pregnant women's thyroid status: comparison of two longitudinal prospective cohort studies

Sofie Bliddal, Ulla Feldt-Rasmussen, Malene Boas^{1,*}, Jens Faber², Anders Juul³,
Torben Larsen⁴ and Dorte Hansen Precht⁵

Department of Endocrinology, Section 2132, Copenhagen University Hospital, Rigshospitalet, Blegdamsvej 9, 2100 Copenhagen, Denmark, ¹Department of Pediatrics, Copenhagen University Hospital, Hillerød, Denmark, ²Department of Endocrinology, Copenhagen University Hospital, Herlev, Denmark, ³Department of Growth and Reproduction, Faculty of Health and Medical Sciences, Copenhagen University Hospital, Rigshospitalet, University of Copenhagen, Copenhagen, Denmark, ⁴Department of Fetal Medicine, Copenhagen University Hospital, Holbaek, Denmark and ⁵Department of Child Psychiatry, Copenhagen University Hospital, Roskilde, Denmark
*(M Boas is now at Department of Growth and Reproduction, Faculty of Health and Medical Sciences, Copenhagen University Hospital, Rigshospitalet, University of Copenhagen, Copenhagen, Denmark)

Correspondence
should be addressed
to S Bliddal
Email
sofiebliddal@gmail.com

Abstract

Objectives: Correct interpretation of thyroid status during pregnancy is vital to secure fetal development. Pregnancy-related changes in maternal thyroid status necessitate the use of gestational age-specific reference ranges. In this study, we investigated between-laboratory reproducibility of thyroid reference ranges in pregnant women.

Design: Comparison of two longitudinal prospective cohort studies including 255 (cohort 1) and 101 (cohort 2) healthy antibody-negative Danish pregnant women attending prenatal care at Copenhagen University Hospital.

Methods: Different immunoassays were used to measure thyroid hormone levels in the two cohorts. Thyroid hormone reference ranges were established for every 5 weeks of gestation. Differences between cohorts were explored through mixed-model repeated measures regression analyses. By applying reference ranges from one cohort to the other, the proportion of women who would be misclassified by doing so was investigated.

Results: TSH increased and free thyroxine (FT₄) decreased as pregnancy progressed. Results indicated highly significant differences between cohorts in free triiodothyronine ($F=21.3$, $P<0.001$) and FT₄ ($F=941$, $P<0.001$). TSH levels were comparable ($P=0.09$). Up to 90.3% of the women had FT₄ levels outside their laboratory's nonpregnant reference range, and up to 100% outside the other cohort's gestational-age-specific reference ranges. Z-score-based reference ranges markedly improved comparison between cohorts.

Conclusion: Even in the same region, the use of gestational-age-specific reference ranges from different laboratories led to misclassification. Up to 100% of maternal FT₄ levels fell outside the other cohort's reference range despite similar TSH levels. In clinical practice, thyroid testing of pregnant women without adding method specificity to gestational age-dependent reference ranges will compromise patient safety.

European Journal of
Endocrinology
(2014) 170, 329–339

Introduction

Thyroid dysfunction during pregnancy is a risk factor to both mother and child (1, 2, 3). The high prevalence of thyroid dysfunction in young women has led to increased support among endocrinologists for thyroid screening of all pregnant women (4, 5). Evaluation of thyroid function in pregnant women, without taking into account the pregnancy-related changes in maternal thyroid status, has been shown to lead to misclassification and misdiagnosis of thyroid disease of risk to obstetric outcome and fetal development (6).

During pregnancy, maternal thyroid status changes according to gestational age (7). Especially in the first trimester, the rise in the thyrotropic-acting pregnancy-related hormone, human chorionic gonadotropin, leads to a rise in free thyroxine (FT₄) and free triiodothyronine (FT₃), and by negative feedback to a suppression of thyroid-stimulating hormone (TSH). Furthermore, the maternal thyroid status is affected by an estrogen-induced increase in thyroxine-binding globulin, an altered iodine clearance in the kidneys, as well as a placental metabolism of thyroid hormones (7). These unequivocal changes in maternal thyroid status necessitate the need for gestational age-specific reference ranges.

In recent years, several studies have been published with the aim of providing thyroid hormone reference ranges for use in pregnant women (8, 9, 10, 11) (reviewed in (11, 12)). Most of the studies were retrospective, cross-sectional studies of maternal thyroid status in the first trimester. The use of such studies entirely depends on the reproducibility of reference ranges in different populations, especially when using different laboratory methods.

In this study, we examined the maternal thyroid status throughout pregnancy based on two prospective longitudinal cohort studies on Danish pregnant women. The aim of the study was to establish two independent sets of longitudinal gestational age-specific reference ranges, and subsequently evaluate the use of these when applied to another comparable cohort.

Subjects and methods

Subjects

This study was based on results from two prospective longitudinal cohort studies on 466 pregnant Danish women at Copenhagen University Hospital (Rigshospitalet and Herlev Hospital), Copenhagen, Denmark (Fig. 1).

Cohort 1 ► From August 1996 to May 1997, all Danish-speaking pregnant women ($n=2526$) attending prenatal care at Copenhagen University Hospital (Rigshospitalet) were invited to join a project regarding stress factors and pregnancy. A total of 1605 women (1623 pregnancies) filled in a health questionnaire including information regarding socioeconomic status, smoking habits, prescription medications, and parity. A random subset of 315 women (316 pregnancies) with no regular cigarette smoking or alcohol consumption nor drug abuse, no psychiatric illness, and no known chronic diseases requiring daily prescription medications, was consecutively enrolled. They were followed with questionnaires, ultrasound scans, and sequential blood sampling once each trimester and upon birth (including umbilical cord blood sampling of the newborn). Within 1 week *post partum*, neurological examinations of their neonates were performed. The examining doctors had no knowledge of the women's background or thyroid status.

Cohort 2 ► During the period of May 1999 to October 2001, women referred to prenatal care at Copenhagen University Hospital (Herlev) were invited to participate in a study regarding fetal growth during pregnancy (13). Among the women who actively volunteered to participate in the study, 151 pregnant women with no risk factors for preeclampsia were consecutively recruited. After informed consent, all the women were scheduled to have three ultrasound examinations and questionnaires, and blood samples drawn approximately every 4–6 weeks throughout pregnancy. Placental growth hormone and insulin-like growth factor 1 levels were analyzed (13). Subsequently, thyroid hormone levels were analyzed to compare intra-individual changes with inter-individual changes during pregnancy (10).

In both cohorts, informed consent was obtained from each woman, and the projects were approved by the Danish Data Protection Agency as well as the local branch of the Danish Ethics Committee (reference numbers 01-077/96 and KF 01 276357 respectively).

In this, only antibody-negative women with singleton pregnancies, without known thyroid disease and without other significant disease (preeclampsia, hypertension, and asthma treated with moderate/high-dose steroid) were included. Women who had achieved pregnancy through assisted reproductive technologies ($n=14$), or had taken hormone supplements during pregnancy ($n=4$), were excluded. Antibody positivity was defined as a thyroid peroxidase antibody level ≥ 30 U/ml and/or

Figure 1

Study flow diagram. Selection process of cohorts 1 and 2. ART, assisted reproductive technology.

a thyroglobulin antibody level ≥ 25 U/ml corresponding to the functional assay sensitivities. Although lower than most cutoffs used in clinical practice, we found this cutoff to be reasonable when attempting to establish normal reference ranges given that the clinical impact of antibody presence was still unresolved even in euthyroid pregnant women (14). Women in whom thyroid antibody analysis (thyroid peroxidase antibody and/or thyroglobulin antibody) had not been performed at least once during their pregnancy were excluded ($n=18$). Twenty-one (7.0%) women in cohort 1 and 15 (10.1%) women in cohort 2 were antibody positive and excluded from the calculation of reference ranges.

In both cohorts, intake of vitamins, prepregnancy height and weight, and smoking status were self-reported in questionnaires. Women who had reported smoking at any time during pregnancy were considered to be smoking.

Laboratory methods

In both cohorts, blood was drawn as nonfasting peripheral venous samples between midmorning and early afternoon. Samples were centrifuged and sera were stored until hormone analysis at -80°C in cohort 1 and -20°C in cohort 2.

In cohort 1, the 1235 AutoDelfia automatic fluoroimmunoassay system (Wallac, Turku, Finland) was used to analyze the following: TSH, FT_3 , total $\text{T}_3(\text{TT}_3)$, and FT_4 , and total $\text{T}_4(\text{TT}_4)$. The inter-assay coefficients of variance (CV) were 4.8, 2.2, and 2.2% at the concentrations of 0.05, 0.9, and 17.6 mU/l respectively for TSH; 10.7, 4.4, 3.2, and 1.6% at concentrations of 2.8, 4.7, 6.5, and 9.7 pmol/l respectively for FT_3 ; and 5.3, 3.7, and 3.1% for 9.3, 15.9, and 19.5 pmol/l respectively for FT_4 . Thyroid peroxidase antibodies (functional assay sensitivity 30 U/ml) were analyzed by RIA (DYNOtest, Brahms, Hennigsdorf, Germany).

In cohort 2, electrochemiluminescence immunoassays on the Roche Modular E170 (Roche) were used to analyze the following: TSH, FT_3 , TT_3 , and FT_4 , and TT_4 . The inter-assay CV were 8.7 and 8.4% at concentrations of 0.9 and 4.9 mU/l respectively for TSH; 6.0 and 8.1% at concentrations of 12 and 30 pmol/l respectively for FT_4 ; and 6.4 and 6.4% at concentrations of 5.3 and 15.0 pmol/l respectively for FT_3 . Thyroid peroxidase antibodies (functional assay sensitivity 28 U/ml) and thyroglobulin antibodies (functional assay sensitivity 25 U/ml) were analyzed by immunofluorescent assays on the Kryptor instrument (Brahms).

Statistical analysis

We used SPSS Statistics, version 20.0 (IBM) for all transformations, independent-sample *t*-tests, χ^2 tests, and mixed model analyses. TSH, FT₃, and FT₄ variables were log transformed to approach normal distributions. No outliers were deleted as these were considered true values rather than analytical errors. Differences in demographics were tested with independent-sample *t*-tests (maternal age, parity, BMI, weight, and height) or χ^2 (smoking status and vitamin intake).

After collecting the large number of blood samples in each cohort, we grouped the blood samples by every 5 weeks of gestation in order to specify the changes in maternal thyroid status during pregnancy. If a woman had more than one blood sample drawn within a 5-week interval, we used the mean of the results in the reference range calculation and mixed model analyses. In Microsoft Excel 2007, we calculated a 95% reference range based on the mean and s.d.s of the log-transformed values within each 5-week group, and then inversely transformed the ranges to achieve actual values.

We performed multilevel mixed model repeated measures regression analysis for each hormone to adjust for the following potential covariates: gestational week group, maternal smoking status, age at conception, prepregnancy BMI, BMI $\times$ BMI, parity, and use of vitamins during pregnancy (15, 16). Log-transformed TSH was also included as a covariate in the analysis of free thyroid hormone levels. In the model, the respective hormone levels were predicted from the cohort belonging as the fixed factor and the gestational age-groups as the repeated factor. We did not replace missing data by dummy variables or missing indicators.

Results

We examined maternal thyroid status throughout pregnancy based on two prospective longitudinal studies on Danish pregnant women (Fig. 1). In cohort 1, a total of 722 blood samples were obtained from 254 women (255 pregnancies) between 14 and 38 weeks of pregnancy. In cohort 2, 715 blood samples from 101 women were taken between 4 and 42 weeks of pregnancy. The demographics of the two cohorts differed significantly in terms of age, smoking, parity, prepregnancy BMI, prepregnancy weight, and self-reported intake of vitamins (Table 1).

In both cohorts, there was a small increase in TSH as pregnancy progressed (Fig. 2A). FT₃ (Fig. 2B) and FT₄ (Fig. 2C) decreased slightly after the first trimester. The changes in hormone levels with increasing gestational age

Table 1 Clinical characteristics of cohorts.

Characteristic	Cohort 1 (n=255)	Cohort 2 (n=101)
Mean maternal age at conception (s.d.), years	29.6 (3.6)	30.8 (4.6)*
Mean prepregnancy height (s.d.), cm	169.2 (6.2)	168.3 (6.0)
Mean prepregnancy weight (s.d.), kg	62.8 (9.1)	68.1 (14.4)*
Mean prepregnancy BMI (s.d.), kg/m ²	21.9 (2.9)	24.4 (4.6)*
Previous births (%), n	87 (34.3)	60 (60.6)*
Smoking (%), n	2 (0.8)	23 (30.3)*
Vitamin intake (%), n	244 (96.8)	82 (85.4)*

*Significant difference between cohorts ($P < 0.05$).

were all highly significant ($P < 0.001$) in the mixed model analysis (described in detail below).

We calculated longitudinal gestational age-specific reference ranges for each cohort and compared the ranges with those of the other cohort (Table 2). To adjust for demographic differences and the varying number of blood samples from each participant, we entered data into a model that controlled for the nonindependence of the participants' repeated visits. TSH, FT₃, and FT₄ levels, respectively, were predicted in a mixed model repeated measures regression analysis using SPSS Software as described in the Methods section. In order to minimize estimation bias, we did not replace missing data with dummy variables or indicators, leaving a total of 311 women in the model. As predicted, the difference in cohort (laboratory method) did not have a significant relationship with TSH ($F(1,297) = 2.9$, $P = 0.09$). As pregnancy progressed, there was a small but highly significant increase ($P < 0.001$) in TSH levels. None of the other covariates entered into the model had a significant effect. In FT₃, there was a significant effect of cohort ($F(1,276) = 21.3$, $P < 0.001$ after adjustment of covariates, of which gestational week group, maternal prepregnancy BMI, and maternal age had a significant effect on the model. Finally, there was a highly significant difference between cohorts (laboratory methods) in FT₄ ($F(1,271) = 941$, $P < 0.001$, with those in cohort 2 showing higher hormone levels than those in cohort 1, $\beta = -0.21$, s.e.m. = 0.01. This effect was found regardless of the significant influence of TSH and gestational week group. Maternal prepregnancy BMI and smoking had borderline significant effects on FT₄ ($P = 0.06$ and $P = 0.09$ respectively).

In cohort 1, the average amount of women who fell outside their own laboratory's nonpregnant reference range (Fig. 3A) were for TSH 9.4% (median 8.2%), FT₃ 77.4% (median 81.2), and FT₄ 87.8% (median

90.1%). The corresponding values for cohort 2 (Fig. 3B) include TSH 2.1% (median 2.2%), FT₃ 30.1% (median 34.2%), and FT₄ 58.4% (median 62.8%).

When applying the reference ranges from one cohort to the other, the number of women with abnormal FT₄

concentrations increased markedly (Fig. 3C and D). Across gestational age-groups, the average percentage of women in cohort 1 who would fall outside of the cohort 2 reference ranges were as follows: TSH 7.0% (median 5.1%), FT₃ 3.5% (median 4.1%), and FT₄ 96.3% (median 95.5%). When applying the cohort 2 reference range to cohort 1, the corresponding numbers were TSH 5.7% (median 5.8%), FT₃ 16.5% (median 17.4%), and FT₄ 95.6% (median 94.6%). As much as 100% of maternal FT₄ values within a gestational age-group were outside of the nonmethod specific reference range.

The two cohorts had rather large proportions of women with a TSH above 3.0 mU/l in all three trimesters (Fig. 2A). To account for this, we recalculated free thyroid hormone reference ranges including only women with a TSH of 0.1–2.5 mIU/l in the first trimester, 0.2–3.0 mIU/l in the second trimester, and 0.3–3.0 mIU/l in the third trimester. These TSH reference ranges were suggested by the American Thyroid Association (ATA) as standardized ranges applicable to all pregnant women (14). Doing so, there was still a significant difference between FT₃ and FT₄ for the two cohorts in all gestational age groups. By use of the reference ranges based on women with a TSH within the ATA-ranges, up to 18.2% (mean 10.6%) of the FT₃ values and 100% (mean 95.7%) of the FT₄ values were outside the other cohort's reference range. When applying the ATA-guideline range for TSH, a mean of 1.9% (up to 4.0%) in cohort 1 and 4.2% (up to 8.2%) in cohort 2 had abnormal TSH-values, despite the fact that these women were within the limits of their own cohort-specific reference range. These proportions were similar to the ones found by applying the TSH-reference range from the other cohort (2.9% (up to 8.0%) and 3.1% (up to 5.1%) respectively). In both cohorts, the TSH reference ranges

Figure 2

Thyroid status in two cohorts of Danish pregnant women according to gestational age. Box plots of (A) TSH, (B) FT₃, and (C) FT₄ concentrations in cohorts 1 and 2 according to 5 week periods of gestation. Boxes represent the 25th–75th percentiles. Horizontal dark line represent the median. Whiskers indicate the 95% CI of the mean. Outliers are illustrated by squares (values > 1.5 interquartile ranges away from the 25th and 75th percentiles) and dots (values > 3 interquartile ranges away from the 25th and 75th percentiles). Horizontal lines across plots represent the calculated mean values of the laboratories' non-pregnant reference ranges. **P* < 0.05. TSH, thyroid-stimulating hormone; FT₃, free triiodothyronine; FT₄, free thyroxine.

Clinical Study	S Bliddal and others	Thyroid status misclassification in pregnancy	170:2	334
----------------	----------------------	---	-------	-----

Table 2 95% reference ranges for TSH, FT₃, and FT₄ in Danish pregnant women according to gestational age (weeks). Analyses were performed in two different laboratories using different methods: in cohort 1, EEG Wallac 1235 AutoDelfia automatic fluoroimmunoassay system, and in cohort 2, Roche Modular E170 electrochemiluminescence immunoassays. Results were based on log-transformed values, which were then inversed to yield the displayed results.

Weeks	Cohort 1				Cohort 2			
	2.5%	97.5%	Mean	n	2.5%	97.5%	Mean	n
TSH (mU/l)								
0–10	–	–	–	–	0.39	3.42	1.15	17
10–15	–	–	–	–	0.46	4.11	1.38	44
15–20	0.43	3.95	1.30	207	0.40	5.05	1.42	86
20–25	0.49	3.37	1.28	44	0.54	3.90	1.45	92
25–30	0.49	3.04	1.22	216	0.51	3.52	1.34	86
30–35	0.39	4.31	1.29	50	0.53	3.57	1.37	90
35–40	0.52	3.37	1.32	191	0.71	4.15	1.72	79
40+	–	–	–	–	0.89	5.80	2.27	16
FT ₃ (pmol/l)								
0–10	–	–	–	–	3.96	5.97	4.86	17
10–15	–	–	–	–	3.61	6.10	4.69	44
15–20	3.44	5.09	4.19	207	3.58	5.76	4.54	86
20–25	3.47	5.00	4.17	44	3.36	5.68	4.37	92
25–30	3.20	4.95	3.99	216	3.38	5.62	4.36	86
30–35	3.28	4.86	3.99	50	3.38	5.44	4.29	90
35–40	3.27	4.91	4.01	191	3.38	5.35	4.25	79
40+	–	–	–	–	3.40	5.32	4.25	17
FT ₄ (pmol/l)								
0–10	–	–	–	–	13.92	20.31	16.82	18
10–15	–	–	–	–	12.35	19.29	15.44	44
15–20	7.62	11.56	9.39	206	11.78	18.01	14.56	86
20–25	6.96	11.62	8.99	44	11.18	16.77	13.69	92
25–30	6.80	10.94	8.62	216	10.58	16.91	13.38	86
30–35	6.87	10.52	8.50	50	10.84	16.62	13.42	90
35–40	6.62	10.95	8.51	191	10.39	16.74	13.19	79
40+	–	–	–	–	10.98	16.53	13.47	16

TSH, thyroid-stimulating hormone; FT₃, free triiodothyronine; FT₄, free thyroxine.

had both higher lower and upper limits (Table 2) than the ones suggested by the ATA.

Z-scores were calculated for each woman based upon her own laboratory's nonpregnant-reference range. The mean and s.d. were calculated assuming that the reference ranges were based on 95% CIs of the mean. This was only done for the free hormone values, as such an approach would not be suitable for a skewed distribution like the one for TSH values. The women outside the other cohorts' z-score based FT₄ reference range fell from a mean of 96.3 to 14.3% in cohort 1 and 95.6 to 24.3% in cohort 2 (Fig. 3E and F). The Z-score reference ranges for FT₄ are provided in Table 3.

Discussion

Consistent with other studies, we found that thyroid status changes with gestational age. While the development in thyroid hormone levels was similar between

cohorts, the actual hormone levels were not. Using the gestational age-specific reference ranges obtained from another comparable cohort to interpret pregnant women's thyroid status misclassified up to 100% of FT₄ values. Although this study was not a methodological study comparing two assays applied to one set of blood samples (diagnostic accuracy study), it does provide an insight into the limited use of previously published reference ranges for pregnant women.

The two cohorts in this study consisted of women living in the Copenhagen area of Denmark, which is traditionally a mildly iodine-deficient area (17). The cohorts were gathered only 2 years apart, although with the latter cohort overlapping the transition of the Danish Iodine fortification program from voluntary to mandatory (17). Although no urine samples were collected in either cohort to test for urinary iodine, a study by Ristic-Medic *et al.* (18) showed that the most reliable biomarker of iodine status in pregnancy is TSH. TSH levels did not differ

Figure 3

Percentages of women whose thyroid status (TSH, FT₃, and FT₄ respectively) would be misclassified by use of a predetermined reference range. (A) Women in cohort 1 outside of own laboratory's nonpregnant reference range. (B) Women in cohort 2 outside of own laboratory's nonpregnant reference range. (C) Women in cohort 1 outside of cohort 2 gestational age-specific reference range. (D) Women in cohort 2 outside

of cohort 1 gestational age-specific reference range.

(E) Women in cohort 1 outside of cohort 2 Z-score-reference range based on cohort 2 nonpregnant reference range.

(F) Women in cohort 2 outside of cohort 1 Z-score-reference range based upon cohort 1 nonpregnant reference range.

TSH, thyroid-stimulating hormone; FT₃, free triiodothyronine; FT₄, free thyroxine.

Clinical Study	S Bliddal and others	Thyroid status misclassification in pregnancy	170:2	336
----------------	----------------------	---	-------	-----

Table 3 95% Z-score reference ranges for free T₄ (FT₄) levels in Danish pregnant women according to gestational age (weeks). Z-scores were based on the calculated mean and s.d. of the respective laboratories' non-pregnant reference range for FT₄.

Weeks	Cohort 1				Cohort 2			
	2.5%	97.5%	Mean	n	2.5%	97.5%	Mean	n
FT ₄ (Z-score)								
10–15	–	–	–	–	–2.9	0.3	–1.3	44
15–20	–3.6	–1.3	–2.4	206	–3.1	–0.3	–1.7	86
20–25	–4.0	–1.3	–2.7	44	–3.3	–0.9	–2.1	92
25–30	–4.1	–1.7	–2.9	216	–3.6	–0.9	–2.2	86
30–35	–4.0	–1.9	–3.0	50	–3.5	–1.0	–2.2	90
35–40	–4.2	–1.7	–3.0	191	–3.8	–0.9	–2.3	79
40+	–	–	–	–	–3.4	–1.0	–2.2	86

significantly between cohorts, and the differences in FT₄ concentrations remained the same when only including women with a TSH-level within the narrow range recommended in the ATA-guidelines. We therefore believe that it is reasonable to assume that the differences in FT₄ concentrations could not be attributed to differences in iodine status.

Differences between the cohorts in demographics did not alter the conclusions when included in the mixed model analysis. Also, while the cohorts differed markedly in smoking status, Shields *et al.* (19) showed that smoking status did not affect FT₄ concentrations. We therefore believe our results to show that even in women from the same geographical area, using the same methodological approach to calculating reference ranges, such reference ranges were highly laboratory-dependent and not applicable outside of its own clinical setting. This was especially true with regards to the analysis of free thyroid hormones.

Although TSH remains the gold standard for thyroid status monitoring during pregnancy (20), it is far from sufficient in most clinical settings (21), and the clinician must often rely on free hormone analyses to monitor a pregnant woman's thyroid status. In a systematic review (22), the present authors showed that in pregnant women treated with antithyroid drugs, FT₄ was a better indicator of maternal and fetal hypothyroidism than TSH. Also, recent guidelines from ATA (14) recommends that in hyperthyroid pregnant women, FT₄ should be kept in the upper normal reference range of FT₄ (regardless of the knowledge of pregnancy-related changes in both thyroid physiology and FT₄-assay results). Further, the importance of free thyroid hormones for fetal development and obstetric outcome has been demonstrated in recent years (23, 24, 25). Henrichs *et al.* (26) found a significant delay in communication skills in children of women who had a

low FT₄ during the first trimester of pregnancy. Making the important distinction between maternal subclinical and clinical hypothyroidism, as well as diagnosing hypothyroxinemia, requires reliable testing of free thyroid hormones.

Regardless of the method used to estimate FT₄ in pregnant women, this is susceptible to the extreme rise in thyroxine-binding globulin and thus the altered serum T₄ binding (22, 27). As recently reviewed by Thienpont *et al.* (28), the gold standard methodology for determination of thyroid hormones is liquid chromatography (LC)/tandem mass spectrometry. However, due to the high costs of this method, RIA is still the most commonly used method in thyroid hormone analysis (28). The difficulty in obtaining consistent RIA-based measures of free thyroid hormones between laboratories has led to a current discussion of whether to recommend analysis of total T₄ or FT₄ index instead of FT₄ (4, 29, 30, 31). The main benefit of total T₄ analyses is a lower interassay variability, which in turn facilitates the use of universal guidelines and reference ranges. As mentioned by Thienpont *et al.* (28), one can use the nonpregnant reference range for total T₄ multiplied by 1.5 throughout pregnancy. However, the interindividual variability in total T₄ is high, resulting in a broader reference range with greater diagnostic imprecision (10, 32). The latter can partly be overcome by comparing test results with former results from the same woman (10).

On the basis of our results, we suggest to work towards a standardization of FT₄ levels. Standardization of highly method-dependent analyses is known from, i.e. the evaluation of prothrombin time by use of the international normalized ratio (33). With Z-scores calculated for each woman based on her own laboratory's nonpregnant reference range, we showed a marked improvement in the comparability of FT₄ levels between

the two cohorts. In our cohorts, the proportionate reaction of each method to the pregnancy-related changes seemed to be of equal size. Therefore, the use of Z-scores enabled a comparison between otherwise completely different absolute FT₄ values in women with similar TSH-levels, from the same region, and at an equal stage of pregnancy.

This study has great implications for the use of thyroid reference ranges in pregnant women. Universal thyroid function screening during pregnancy has recently been officially endorsed by members of the Endocrine Society (4). Little doubt remains that such screening will detect many women with unknown thyroid dysfunction (5, 34, 35). However, our results show that implementing screening programs will have a grave impact if doing so without prior establishment of method- and region-specific gestational age-related reference ranges.

Conclusion

Incorrect interpretation of maternal thyroid status is a great risk factor in the care of pregnant women. Such medical errors can alter treatment strategies to negatively impact the fetal development and obstetrical outcome. Maternal thyroid status must be interpreted in the context of the dynamic pregnancy-related changes in thyroid hormone production and metabolism. The impracticability and costs of establishing method-specific reference ranges for pregnant women make the use of pre-established reference ranges highly appealing. This study provided longitudinal gestational age-specific reference ranges from two cohorts based on different immunoassays. We demonstrated that while this provides a reliable option in the interpretation of TSH, using predetermined reference ranges to interpret free thyroid hormone values will misclassify up to 100% of pregnant women – even within populations from the same region and using the same methodological approach to establish the reference ranges. However, our results showed that a feasible approach to overcome such differences could be to implement Z-score-standardization of FT₄-levels across laboratories. While the importance of trimester-specific reference ranges has been widely stressed, the importance of method specificity of such reference ranges is largely neglected when publishing reference ranges from different populations – this study illustrates the problems of doing so. Hopefully, in the future, FT₄ methods will be compared with and calibrated against the ‘gold standard’ method, LC/tandem mass spectrometry.

Regardless of the purpose of thyroid function testing during pregnancy (implementation of screening

strategies, monitoring of treatment for thyroid disorders, etc.), the care taker must be aware of the analytical pitfalls when doing so. In the care of pregnant women, patient safety will be compromised if method specificity is not added to gestational age-dependent reference ranges.

Declaration of interest

The authors declare that there is no conflict of interest that could be perceived as prejudicing the impartiality of the research reported.

Funding

The funding sources had no influence on the design and conduct of the study; the analysis and interpretation of the study data; drafting, revision or approval of the manuscript; or decision to submit the manuscript for publication. The work related to the studies and preparation of this article was supported by the Musikforlæggerne Agnes og Knut Mørks Foundation (2000, 2012); the Danish Council for Independent Research: Medical Sciences (1994, 1997); the Ministry of Social Affairs (1995); the Axel Muusfeldt's Foundation (2013); the Videnskabsminister Erna Hamilton Foundation (2012); the Director Ib Henriksen Foundation (1996, 2012); the Snedkermester Sophus Jacobsen og hustru Astrid Jacobsens Foundation (2013); the Faculty of Medical Science's Foundation (2013); the Frimodt-Heineke Foundation (2012); the Torben and Alice Frimodt's Foundation (2012); the Augustinus Foundation (2000); the Danish Medical Association's Foundation (2002); Doctor Søren Segel and Johanne Wiibroe Segel's Foundation (2002); the Illum Foundation (1997); the A P Møller Foundation for the Advancement of Medical Science (2000, 2001, 2012); the Arvid Nilsson Foundation (2011); the Gangsted Foundation (1995); Helsefonden (1999); the Elsass Foundation (2001); the Familien Hede Nielsens Foundation (2013); and the Copenhagen University Foundation (2013).

Author contribution statement

D H Precht was in charge of the study design, clinical examinations, and data gathering in cohort 1, and supervised the statistical analyses. A Juul, T Larsen, and M Boas were in charge of data gathering in cohort 2. J Faber was in charge of the analyses of the thyroid autoantibodies in cohort 1. S Bliddal did the statistical analyses and prepared the manuscript. All authors contributed to the discussion and review of the manuscript.

Acknowledgements

Annegrete Pedersen (Wallac) is thanked for her help in analyzing the hormones in cohort 1. Inger Wätjen (Frederiksberg Hospital) is thanked for her help in analyzing the thyroid peroxidase antibodies in cohort 1. Oriana Aragón (Yale University) is thanked for providing excellent statistical guidance.

References

- 1 Yazbeck CF & Sullivan SD. Thyroid disorders during pregnancy. *Medical Clinics of North America* 2012 **96** 235–256. (doi:10.1016/j.mcna.2012.01.004)
- 2 Karakosta P, Alegakis D, Georgiou V, Roumeliotaki T, Fthenou E, Vassilaki M, Boumpas D, Castanas E, Kogevinas M & Chatzi L. Thyroid dysfunction and autoantibodies in early pregnancy are associated with increased risk of gestational diabetes and adverse birth outcomes.

- Journal of Clinical Endocrinology and Metabolism* 2012 **97** 4464–4472. (doi:10.1210/jc.2012-2540)
- 3 Mannisto T, Vaarasmaki M, Pouta A, Hartikainen AL, Ruokonen A, Surcel HM, Bloigu A, Jarvelin MR & Suvanto-Luukkonen E. Perinatal outcome of children born to mothers with thyroid dysfunction or antibodies: a prospective population-based cohort study. *Journal of Clinical Endocrinology and Metabolism* 2009 **94** 772–779. (doi:10.1210/jc.2008-1520)
 - 4 De Groot L, Abalovich M, Alexander EK, Amino N, Barbour L, Cobin RH, Eastman CJ, Lazarus JH, Luton D, Mandel SJ *et al.* Management of thyroid dysfunction during pregnancy and postpartum: an Endocrine Society clinical practice guideline. *Journal of Clinical Endocrinology and Metabolism* 2012 **97** 2543–2565. (doi:10.1210/jc.2011-2803)
 - 5 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. (doi:10.1210/jc.2009-2009)
 - 6 Stricker R, Echenard M, Eberhart R, Chevailler MC, Perez V, Quinn FA & Stricker R. Evaluation of maternal thyroid function during pregnancy: the importance of using gestational age-specific reference intervals. *European Journal of Endocrinology* 2007 **157** 509–514. (doi:10.1530/EJE-07-0249)
 - 7 Glinoe D. The regulation of thyroid function in pregnancy: pathways of endocrine adaptation from physiology to pathology. *Endocrine Reviews* 1997 **18** 404–433. (doi:10.1210/er.18.3.404)
 - 8 Dashe JS, Casey BM, Wells CE, McIntire DD, Byrd EW, Leveno KJ & Cunningham FG. Thyroid-stimulating hormone in singleton and twin pregnancy: importance of gestational age-specific reference ranges. *Obstetrics and Gynecology* 2005 **106** 753–757. (doi:10.1097/01.AOG.0000175836.41390.73)
 - 9 Springer D, Zima T & Limanova Z. Reference intervals in evaluation of maternal thyroid function during the first trimester of pregnancy. *European Journal of Endocrinology* 2009 **160** 791–797. (doi:10.1530/EJE-08-0890)
 - 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. (doi:10.1530/EJE-09-0579)
 - 11 Soldin OP, Soldin D & Sastoque M. Gestation-specific thyroxine and thyroid stimulating hormone levels in the United States and worldwide. *Therapeutic Drug Monitoring* 2007 **29** 553–559. (doi:10.1097/FTD.0b013e31815709ac)
 - 12 Feldt-Rasmussen U, Bliddal Mortensen AS, Rasmussen AK, Boas M, Hilsted L & Main K. Challenges in interpretation of thyroid function tests in pregnant women with autoimmune thyroid disease. *Journal of Thyroid Research* 2011 **2011** 598712. (doi:10.4061/2011/598712)
 - 13 Chellakooty M, Vangsgaard K, Larsen T, Scheike T, Falck-Larsen J, Legarth J, Andersson AM, Main KM, Skakkebaek NE & Juul A. A longitudinal study of intrauterine growth and the placental growth hormone (GH)–insulin-like growth factor I axis in maternal circulation: association between placental GH and fetal growth. *Journal of Clinical Endocrinology and Metabolism* 2004 **89** 384–391. (doi:10.1210/jc.2003-030282)
 - 14 Stagnaro-Green A, Abalovich M, Alexander E, Azizi F, Mestman J, Negro R, Nixon A, Pearce EN, Soldin OP, Sullivan S *et al.* Guidelines of the American Thyroid Association for the diagnosis and management of thyroid disease during pregnancy and postpartum. *Thyroid* 2011 **21** 1081–1125. (doi:10.1089/thy.2011.0087)
 - 15 Haddow JE, Craig WY, Palomaki GE, Neveux LM, Lambert-Messerlian G, Canick JA, Malone FD & D'Alton ME. Impact of adjusting for the reciprocal relationship between maternal weight and free thyroxine during early pregnancy. *Thyroid* 2013 **23** 225–230. (doi:10.1089/thy.2012.0440)
 - 16 Mannisto T, Surcel HM, Ruokonen A, Vaarasmaki M, Pouta A, Bloigu A, Jarvelin MR, Hartikainen AL & Suvanto E. Early pregnancy reference intervals of thyroid hormone concentrations in a thyroid antibody-negative pregnant population. *Thyroid* 2011 **21** 291–298. (doi:10.1089/thy.2010.0337)
 - 17 Laurberg P, Jorgensen T, Perrild H, Ovesen L, Knudsen N, Pedersen IB, Rasmussen LB, Carle A & Vejbjerg P. The Danish investigation on iodine intake and thyroid disease, DanThyr: status and perspectives. *European Journal of Endocrinology* 2006 **155** 219–228. (doi:10.1530/eje.1.02210)
 - 18 Ristic-Medic D, Piskackova Z, Hooper L, Ruprich J, Casgrain A, Ashton K, Pavlovic M & Glibetic M. Methods of assessment of iodine status in humans: a systematic review. *American Journal of Clinical Nutrition* 2009 **89** 2052S–2069S. (doi:10.3945/ajcn.2009.27230H)
 - 19 Shields B, Hill A, Bilous M, Knight B, Hattersley AT, Bilous RW & Vaidya B. Cigarette smoking during pregnancy is associated with alterations in maternal and fetal thyroid function. *Journal of Clinical Endocrinology and Metabolism* 2009 **94** 570–574. (doi:10.1210/jc.2008-0380)
 - 20 Glinoe D & Spencer CA. Serum TSH determinations in pregnancy: how, when and why? *Nature Reviews. Endocrinology* 2010 **6** 526–529. (doi:10.1038/nrendo.2010.91)
 - 21 Ghassabian A & Tiemeier H. Is measurement of maternal serum TSH sufficient screening in early pregnancy? A case for more randomized trials. *Clinical Endocrinology* 2012 **77** 802–805. (doi:10.1111/j.1365-2265.2012.04494.x)
 - 22 Bliddal S, Rasmussen AK, Sundberg K, Brocks V & Feldt-Rasmussen U. Antithyroid drug-induced fetal goitrous hypothyroidism. *Nature Reviews. Endocrinology* 2011 **7** 396–406. (doi:10.1038/nrendo.2011.34)
 - 23 Cleary-Goldman J, Malone FD, Lambert-Messerlian G, Sullivan L, Canick J, Porter TF, Luthy D, Gross S, Bianchi DW & D'Alton ME. Maternal thyroid hypofunction and pregnancy outcome. *Obstetrics and Gynecology* 2008 **112** 85–92. (doi:10.1097/AOG.0b013e3181788dd7)
 - 24 Finken MJ, van Eijdsen M, Loomans EM, Vrijkotte TG & Rotteveel J. Maternal hypothyroxinemia in early pregnancy predicts reduced performance in reaction time tests in 5- to 6-year-old offspring. *Journal of Clinical Endocrinology and Metabolism* 2013 **98** 1417–1426. (doi:10.1210/jc.2012-3389)
 - 25 Casey BM, Dashe JS, Spong CY, McIntire DD, Leveno KJ & Cunningham GF. Perinatal significance of isolated maternal hypothyroxinemia identified in the first half of pregnancy. *Obstetrics and Gynecology* 2007 **109** 1129–1135. (doi:10.1097/01.AOG.0000262054.03531.24)
 - 26 Henrichs J, Bongers-Schokking JJ, Schenk JJ, Ghassabian A, Schmidt HG, Visser TJ, Hooijkaas H, de Muinck Keizer-Schrama SM, Hofman A, Jaddoe VV *et al.* Maternal thyroid function during early pregnancy and cognitive functioning in early childhood: The Generation R Study. *Journal of Clinical Endocrinology and Metabolism* 2010 **95** 4227–4234. (doi:10.1210/jc.2010-0415)
 - 27 Wang R, Nelson JC, Weiss RM & Wilcox RB. Accuracy of free thyroxine measurements across natural ranges of thyroxine binding to serum proteins. *Thyroid* 2000 **10** 31–39. (doi:10.1089/thy.2000.10.31)
 - 28 Thienpont LM, Van Uytfaange K, Poppe K & Velkeniers B. Determination of free thyroid hormones. *Best Practice & Research. Clinical Endocrinology & Metabolism* 2013 **27** 689–700. (doi:10.1016/j.beem.2013.05.012)
 - 29 Fereidoun A, Ladan M, Atieh A, Hossein D, Maryam T, Sahar A & Hedayati M. Establishment of the trimester-specific reference range for free thyroxine index. *Thyroid* 2013 **23** 354–359. (doi:10.1089/thy.2012.0407)
 - 30 Midgley JE & Hoermann R. Measurement of total rather than free thyroxine in pregnancy: the diagnostic implications. *Thyroid* 2013 **23** 259–261. (doi:10.1089/thy.2012.0469)
 - 31 Soldin OP. When thyroidologists agree to disagree: comments on the Endocrine Society pregnancy and thyroid disease clinical practice

Clinical Study	S Bliddal and others	Thyroid status misclassification in pregnancy	170:2	339
----------------	----------------------	---	-------	-----

- guideline. *Journal of Clinical Endocrinology and Metabolism* 2012 **97** 2632–2635. (doi:10.1210/jc.2012-2529)
- 32 Andersen S, Bruun NH, Pedersen KM & Laurberg P. Biologic variation is important for interpretation of thyroid function tests. *Thyroid* 2003 **13** 1069–1078. (doi:10.1089/105072503770867237)
- 33 Riley RS, Rowe D & Fisher LM. Clinical utilization of the international normalized ratio (INR). *Journal of Clinical Laboratory Analysis* 2000 **14** 101–114. (doi:10.1002/(SICI)1098-2825(2000)14:3<101::AID-JCLA4>3.0.CO;2-A)
- 34 Vaidya B, Anthony S, Bilous M, Shields B, Drury J, Hutchison S & Bilous R. Detection of thyroid dysfunction in early pregnancy: Universal screening or targeted high-risk case finding? *Journal of Clinical Endocrinology and Metabolism* 2007 **92** 203–207. (doi:10.1210/jc.2006-1748)
- 35 Horacek J, Spitalnikova S, Dlabalova B, Malirova E, Vizda J, Svilius I, Cepkova J, Mc Grath C & Maly J. Universal screening detects two-times more thyroid disorders in early pregnancy than targeted high-risk case finding. *European Journal of Endocrinology* 2010 **163** 645–650. (doi:10.1530/EJE-10-0516)

Received 16 August 2013

Revised version received 18 November 2013

Accepted 25 November 2013