Received Date : 23-May-2016

Revised Date : 01-Jul-2016

Accepted Date : 04-Jul-2016

Article type : 2 Original Article - Europe, excluding UK

Risk factors and a clinical prediction model for low maternal thyroid function during early pregnancy: two population-based prospective cohort studies.

Tim I M Korevaar MD [1,2,3], Daan Nieboer MSc [4], Peter H L T Bisschop MD [5], Mariette Goddijn MD [6], Marco Medici MD [2,3], Layal Chaker MD [2,3], Professor Yolanda B de Rijke [2,7], Professor Vincent W V Jaddoe [1,8,9], Professor Theo J Visser [2,3], Professor Ewout W Steyerberg [4], Professor Henning Tiemeier [8,10], Tanja G Vrijkotte PhD [11], Professor Robin P Peeters [2,3].

[1] The Generation R Study Group and Departments of [2] Internal Medicine, [3] Rotterdam Thyroid Center, [4] Public Health, [7] Clinical Chemistry, [8] Epidemiology, [9] Pediatrics, [10] Child and Adolescent Psychiatry; Erasmus University Medical Center – Sophia Children's Hospital, Rotterdam, The Netherlands

Departments of [5] Endocrinology, [6] Reproductive Medicine, [11] Public Health; Academic Medical Center, Amsterdam, the Netherlands

Short title: Clinical prediction model for maternal thyroid dysfunction during pregnancy.

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/cen.13153

Key words: prediction - thyroid – TSH – FT4 – TPO – pregnancy

Disclosure Summary: The authors have nothing to disclose.

Correspondence to: Tim Korevaar, M.D. , MSc

Room NA-2913, Department of Endocrinology/Epidemiology, Erasmus MC

Dr Molewaterplein 50, 3015 GE

Rotterdam, The Netherlands

E-mail: t.korevaar@erasmusmc.nl

BACKGROUND: Low maternal thyroid function during early pregnancy is associated with various adverse outcomes including impaired neurocognitive development of the offspring, premature delivery and abnormal birth weight.

AIM: To aid doctors in the risk assessment of thyroid dysfunction during pregnancy, we set out to investigate clinical risk factors and derive a prediction model based on easily obtainable clinical variables.

METHODS: 9767 women during early pregnancy (≤ 18 week) were selected from two population-based prospective cohorts: the Generation R Study (N=5985) and the ABCD study (N=3782). We aimed to investigate the association of easily obtainable clinical subject characteristics such as maternal age, BMI, smoking status, ethnicity, parity and gestational age at blood sampling with the risk of low free thyroxine (FT4) and elevated thyroid stimulating hormone (TSH), determined according to the 2.5^{th}-97.5^{th} reference range in TPOAb negative women.

RESULTS: BMI, non-smoking and ethnicity were risk factors for elevated TSH levels, however, the discriminative ability was poor (range c-statistic of 0.57 to 0.60). Sensitivity analysis showed that addition of TPOAbs to the model yielded a c-statistic of 0.73-0.75.
Maternal age, BMI, smoking, parity and gestational age at blood sampling were risk factors for low FT4, which taken together provided adequate discrimination (range c-statistic of 0.72 to 0.76).
CONCLUSIONS: Elevated TSH levels depend predominantly on TPOAb levels and prediction of elevated TSH levels is not possible with clinical characteristics only. In contrast, the validated clinical prediction model for FT4 had high discriminative value to assess the likelihood of low FT4 levels.

INTRODUCTION

Adequate thyroid hormone (TH) availability during pregnancy is crucial for the regulation of metabolic demand, energy homeostasis and adequate supply of THs to the developing fetus.[1] Maternal thyroid hormone deficiency occurs in approximately 4.8-18% of all pregnant women depending on the definition used.[2-5] Overt and subclinical hypothyroidism during early pregnancy are associated with pregnancy loss, premature birth, preeclampsia and impaired child neurocognitive development.[2-4, 6, 7] Recently, hypothyroxinemia (or isolated low FT4 levels) has also been associated with adverse outcomes including premature birth, placental abruption and impaired child neurocognitive development.[8-11]

The fetal thyroid is not fully functional until the 18-20th week of pregnancy. Critical early stages of pregnancy and fetal development therefore predominantly depend on the maternal supply of THs. This specific time window illustrates that it is important to identify women with thyroid dysfunction as early as possible. However, the identification of women at high-risk for gestational thyroid dysfunction in clinical practice is difficult because the majority of women do not present with traditional symptoms. In addition, the status of major known risk factors such as TPO-antibodies and iodine status is usually unknown. This results in under-diagnosis in a large proportion of women with gestational thyroid dysfunction.[12-16]

Because of these reasons, a screening approach is needed to identify women with thyroid dysfunction. However, whether or not all women should be screened for gestational thyroid disease is a matter of controversy and debate. The American Thyroid Association (ATA) and Endocrine Society guidelines advocate aggressive case finding based on risk factors instead of a universal screening approach. Nevertheless, a substantial number of clinicians prefer to perform universal screening over case finding (~50/50 in Europe and ~75/25 in USA and Asia).[17-21] This state of mind is a reflection of studies showing that 30-89% of cases are missed with a case finding approach using risk factors recommended by international guidelines.[12-16] The risk factors that have been used to provide screening estimates in these studies are based on only few, small studies mainly performed in non-pregnant populations.[2, 4]

The combination of non-discriminative clinical symptomatology, the lack of evidence-based risk factors, and the ongoing debate regarding screening underlie the need for clinical prediction tools to identify women with thyroid dysfunction during pregnancy. In order to aid physicians during clinical practice, we aimed to develop and validate clinical risk factors and a prediction model for the identification of women with an impaired thyroid function during early pregnancy based only on subject characteristics that are easily obtainable in clinical practice.

MATERIALS AND METHODS

Design

This study was embedded in two population-based prospective cohorts in the Netherlands, the Generation R Study (Rotterdam) and the Amsterdam Born Children and their Development (ABCD) study.[22, 23] Details on data ascertainment are discussed in the supplemental appendix.

Candidate predictors

For the development of this clinical prediction model we selected variables that are readily available or that can be conveniently obtained in clinical practice. This allow for easy implementation into clinical practice. Variables that were considered possible risk factors for elevated TSH and low FT4 (and as a sensitivity analyses also for TPOAb positivity) were selected based on the literature[24-27] and biological plausibility, and availability of robust ascertainment in both cohorts and included maternal age, BMI, smoking status, parity, ethnicity, gestational age at blood sampling, a medical history (coded as yes/no) of miscarriage or stillborn. Since ethnicity may differ in other populations, we repeated the model with ethnicity recoded as Western/Non-Western in order to allow generalizability to non-Dutch populations.

Outcomes

Reference ranges were determined by population-based calculations according to the 2.5th - 97.5th percentiles after exclusion of women with twin pregnancies, pre-existing thyroid disease or thyroid (interfering) medication usage, fertility treatment and TPOAb positivity, as recommended by international guidelines.[2-4] In Generation R, the intra- and interassay coefficients of variation were <4.1% for TSH at a range of 3.97-22.7 mU/L and <5.4% for FT4 at a range of 14.3-25.0 pmol/L (for Vitros ECI; Ortho Clinical Diagnostics, Rochester, NY); total T4 was not assessed.[28] In the ABCD study, the sensitivity of the TSH assay was 0.1 mU/L, and the interassay coefficient of variation was 5.0% while the detection limit of the free T4 assay was 1.9 pmol/L, and the interassay coefficient of variation ranged between 3.1 and 5.0% (access immunoanalyser of Beckman Coultier Inc. (Fullerton, California)). Subsequently, elevated TSH was defined as a TSH level >97.5th percentile (4.04 mU/L in Generation R, 3.09 mU/l in ABCD study). Low FT4 was defined as an FT4 level <2.5th percentile (10.4 pmol/L in Generation R, 7.39 pmol/L in ABCD study). We also investigated a more liberal cut-off value (5th – 95th percentiles). In Generation R, maternal TPOAbs were measured using the Phadia 250

immunoassay (Phadia AB, Uppsala, Sweden) and considered positive when >60 IU/ml. In the ABCD study, TPOAbs were determined by an enzyme-linked immunosorbent assay (ELISA) [ELIZEN TG Ab (E-CK-96), Zentech, Luik, Belgium and considered positive when >80 kU/l.

Statistical analysis

Logistic regression was used to estimate univariable and multivariable odds ratios with 95% confidence intervals (CIs) for each predictor. A full multivariable model was fitted with all candidate predictors with chosen transformation based on linearity assessment with restricted cubic splines. We selected the number of predictors using backward selection based on the Wald statistic of the pooled regression coefficients, with a *P*-value <0.20 as to keep predictors liberally in the model. The regression coefficients in the final model were multiplied with a shrinkage factor which was estimated with bootstrapping.[29] The final model was also presented as a score chart for easy implementation of the model in clinical practice.

The performance of the resulting model was assessed using cross-validation, fitting the model in the Generation R cohort and subsequently validated in the ABCD cohort and vice versa.[30] Prediction models were also internally validated using bootstrapping to calculate optimism corrected estimates of performance. We subsequently assessed calibration and discrimination of the prediction models. Discrimination refers to the ability of a prediction model to distinguish between patients with and without the outcome of interest. Discriminative ability was quantified using the *c*-statistic, which is equivalent to the area under the receiver operating curve for models predicting a binary outcome. For a model with perfect discrimination the *c*-statistic is equal to 1.0 and a *c*-statistic of 0.5 means that the prediction model is equivalent to a coin toss.

Calibration refers to the agreement between predicted probabilities of the prediction model and the observed outcomes. Calibration was assessed graphically using calibration plots and quantified using calibration-in-the-large (calibration-i.t.l.) and the calibration slope. Calibration-i.t.l. measures whether predicted probabilities are on average too high or too low and should ideally be equal to 0.

The calibration slope measures the average predictor strength and should ideally be equal to 1. We assessed the calibration graphically using calibration plots (see Supplementary Appendix). Missing values of the candidate predictors in the Generation R Study and ABCD study data were multiple imputed (five times). The imputation model included all candidate predictor variables, the outcome variable and several relevant variables descriptive of study subjects. All analyses were performed in each of the completed datasets and final results were pooled using Rubin's rules.

All statistical analyses were performed in R 3.1.2, multiple imputation was done using the *mice* package and model fitting was done using the *rms* package.

RESULTS

The final populations comprised 9767 women, N=5985 from the Generation R cohort and N=3782 from the ABCD cohort (Figure S1). Elevated TSH was observed in 217 (3.6%) and 146 (3.9%) women, low FT4 was present in 166 (2.8%) and 108 (2.9%) women and TPOAb positivity was present in 313 (6%) and 227 (6%) women, in Generation R and ABCD respectively. Descriptive statistics of both populations are shown in Table S1, outcomes of the imputation process are show in Table S2 and S3.

Risk factors and prediction model for elevated maternal TSH

Higher levels of maternal BMI and Asian ethnicity were associated with a higher risk of elevated maternal TSH whereas smoking and non-Western ethnicity were associated with a lower risk of elevated maternal TSH (Table 1). The combination of relevant risk factors for elevated maternal TSH levels yielded a c-statistic of 0.57-0.60 (Table 2). This model allowed for the calculation of a

predictive risk score that can estimate a subject's risk of elevated TSH between 2% and 7% (Table S4). Sensitivity analyses showed that recoding of ethnicity (to Western vs non-Western) did not change the c-statistic while the use of a more liberal TSH cut-off (>95th percentile) did not yield higher c-statistic (data not shown).

TPOAbs play an important role in the pathophysiology of elevated TSH. In order to investigate whether the poor discriminative ability of the prediction model for elevated TSH was due to the strong association of TPOAbs with elevated TSH we added TPOAbs to the prediction model for elevated TSH. After addition of TPOAbs to the model for elevated TSH, this model yielded a c-statistic of 0.73-0.75 (data not shown). A prediction model for TPOAb positivity (Table S5) itself yielded a c-statistic of 0.50-0.57 (data not shown).

Risk factors and prediction model for low maternal FT4

Higher gestational age, maternal age (from 30 years onwards), BMI, parity and smoking were all associated with a higher risk of low maternal FT4 levels (Table 3). A combination of relevant risk factors predicted low FT4 with a c-statistic of 0.72-0.76 (Table 2; addition of ethnicity to the model did not increase model performance). Stepwise addition of covariates to the model show a built up to the final c-statistic as follows: BMI alone (c-statistic 0.69); former model + gestational age (0.747); former model + parity (0.749); former model + smoking (0.75); former model + maternal age (0.76). Sensitivity analysis showed similar results after recoding ethnicity (Western vs non-Western), for a more liberal FT4 cut-off value (<5th percentile), a wider range for gestational age at blood sampling, and after addition of TPOAbs to the model (Table S6).

A clinical scoring model that can be used for the risk assessment of low maternal FT4 is presented in Table 4. This model allowed for the calculation of a predictive risk score that can estimate a subject's risk of low FT4 that will vary between <0.5% and 27% (Figure 1). The more detailed model can be accessed through an online calculator (per journal request, will be made available upon acceptance). Regression formulas for all models are shown in the supplementary appendix.

DISCUSSION

This study reports several easily obtainable clinical risk factors, and the development of the first clinical prediction models for low maternal thyroid function during early pregnancy. The prediction model for low FT4 levels demonstrated a good overall discriminative ability and external generalizability and we provide clinical tools (i.e. a score chart and online calculator) that can be used to assess the risk of low maternal FT4 in clinical practice. We identified several risk factors for elevated TSH levels, however, even a combination of these risk factors lacked proper overall discriminative ability and we show that this is most likely due to the strong association of TPOAbs with elevated TSH.

The American Thyroid Association and Endocrine Society guidelines recommend aggressive case-finding screening utilizing risk factors such as a medical history of head/neck radiation, family history of thyroid dysfunction, obesity (BMI >40kg/m^2), age (>30 years) or symptoms of thyroid dysfunction.[2, 4] These recommendations are based on twelve studies of which the majority (11 out of 12) were performed in non-pregnant populations and lacked (cross) replication.[2, 4] However, the majority of women with gestational thyroid dysfunction do not have classical or distinct symptoms. Moreover, current risk factors recommended by international guidelines are poorly associated with the risk of abnormal thyroid function and are only present in a very small number of women.[12-16] Because currently used clinical risk factors do not enable to distinguish the low from high-risk

groups, physicians are forced to choose between universal screening or screening of a very selected group with distinct symptomatology. Both approaches are suboptimal with regards to efficiency and the ratio of benefits over harms. This study aimed to optimally study risk factors for a risk assessment based on a high-case finding approach, this should allow for better comparison of screening approaches in the future.

Low FT4 levels are associated with impaired child neurocognitive development, abnormal birth weight and a higher risk of premature birth.[8, 31-33] Assessment of the pre-test risk is a prerequisite for evidence –based medical decision making, yet in the majority of cases the risk of thyroid dysfunction can only be assessed by expert opinion. Our prediction model provides a clinical tool for assessing the pre-test risk of low FT4 which, together with clinical expertise, will allow clinicians to make an informed decision on whether or not to test the patient. The prediction model for low FT4 has a good discriminative ability and is particularly able to identify a large number of women that can be considered as a low-risk group. This suggests that the use of the prediction model in clinical practice can optimize both universal and case finding screening strategies. Currently, the effects of levothyroxine treatment in women with isolated low FT4 are unknown and most international guidelines do not advocate treatment.[2-4] It is important to note that the optimal cut-off for any prediction model is based on the benefits (identification of at risk women and providing treatment) and harms (missing at risk women, overtreatment) for each specific cut-off. In order to determine such a cut-off, knowledge on the downstream effects of treatment are required, and such information is currently not available in the field. Therefore, our model can currently be used to assess risk, replacing or adjacent to current known risk factors. However, until further data is available on the optimal clinical decision cut-offs for maternal TSH and FT4 during pregnancy and the harms and benefits of treatment, it is not possible to identify single cut-offs within our model that will optimally distinguish women who will benefit from thyroid function testing.

The small number of women with both elevated TSH and low FT4 levels in the two cohorts (0.2% and 0.7%) and the difference in risk factors for elevated TSH and low FT4 suggest that there is a different pathophysiological mechanism behind these abnormal thyroid function test outcomes. TPOAbs are considered to be a major risk factor for gestational thyroid dysfunction, yet in this study TPOAbs were only associated with a higher risk of elevated TSH. The poor discriminative ability of the prediction model for elevated TSH positivity is explained by the strong association of TPOAbs with elevated TSH as is supported by the results of our sensitivity analysis showing that addition of TPOAbs to the prediction model for elevated TSH substantially improved the discriminative ability. To further substantiate this, we performed a further sensitivity analysis which showed that a prediction model for TPOAb positivity, utilizing the same risk factors as the prediction model for elevated TSH, yielded a poor discriminative ability. Notably, the prediction model for high TSH that incorporates TPOAbs would defeat the purpose of our study and was therefore only devised to investigate the reason for the poor predictive ability of the model. The overlap between TPOAb positivity and elevated TSH also plays an important role in association studies on adverse clinical outcomes as it has been shown that the effect estimates for elevated TSH are much higher with increasing TPOAb levels.[31, 34, 35] Interestingly, addition of TPOAbs to the prediction model for low FT4 did not improve prediction, although it is likely that TPOAbs are a risk factor for low FT4.

The association of a potential risk factor with TPOAb positivity may also underlie the associations of risk factors with high TSH or low FT4. For example, higher parity was associated with a lower risk of TPOAb positivity while it showed a lower risk estimate for high TSH and a higher risk estimate for low FT4. Interestingly, smoking was associated with a lower risk of TPOAb positivity which fits with data showing that smoking protects against the development of TPOAbs and Hashimoto's disease.[36] In line with this results, smoking was associated with a lower risk of high TSH, yet it was associated with a higher risk low FT4. This discrepancy is likely to represent two different pathways via which

smoking can affect thyroid function. Apart from the differences in TPOAb positivity, smoking has also been shown to decrease hCG levels and via this pathway it may affect FT4 levels.[37]

This study was designed for easy implementation into clinical practice, and included a large number of participants with detailed information on clinical characteristics. We used state of the art prediction modelling techniques that aim to accurately predict future patients rather than predictions that are merely correct for patients of the development dataset. The main limitation of this study is that some of the variables were self-reported and derived in a cross-sectional manner. However, this mimics clinical practice where the patient interview is a crucial part of medical decision making.

Another potential limitation is that data on certain variables were derived differently in the Generation R study as compared to the ABCD study. Therefore the external replication analyses may have been suboptimal. The main differences between data derived in the Generation R Study compared to the ABCD study were; measured versus self-reported BMI, measured versus registry derived gestational age at blood sampling, self-reported versus registry derived parity, respectively, and differences in assays used to measure TSH, FT4 and TPOAb levels. In general, differences in assay usage affect the generalizability of virtually all research outcomes in this field, however in this paper we show that our results are replicable for two different assays when population-based reference ranges are used. This is despite differences in absolute levels of TSH and FT4 cut-offs, which are most likely due to the different assays used, but could also be due to population differences in subject characteristics.[38, 39] Moreover, it is important to note that although absolute values are likely to differ between populations, relative values are highly correlated all clinically used assays which is illustrated by excellent replication of the prediction model for low FT4.[38]

Although we found similar results after re-categorizing ethnicity groups, further studies are needed to verify if the results of this study are also applicable to other populations. In addition, it should be noted that self-reported variables such as smoking and BMI (ABCD only) may lead to measurement error that may introduce bias in this study. However, self-reported variables mimic clinical practice which allows for better generalizability.

In conclusion, the prediction models presented in this study shows that easily obtainable clinical characteristics are useful for estimating the pre-test probability of low maternal FT4 levels during pregnancy. This model may aid doctors in identification of women at risk for hypothyroxinemia and overt hypothyroidism.

ACKNOWLEDGEMENTS

The contributions of Professor Herbert Hooijkaas and the endocrine laboratory technicians are highly appreciated. The Generation R Study is conducted by the Erasmus Medical Center (Rotterdam) in close collaboration with the School of Law and Faculty of Social Sciences of the Erasmus University Rotterdam; the Municipal Health Service Rotterdam area, Rotterdam; the Rotterdam Homecare Foundation, Rotterdam; and the Stichting Trombosedienst and Artsenlaboratorium Rijnmond, Rotterdam. We gratefully acknowledge the contribution of children and parents, general practitioners, hospitals, midwives, and pharmacies in Rotterdam. The general design of the Generation R Study is made possible by financial support from the Erasmus Medical Center, Rotterdam; the Erasmus University Rotterdam; The Netherlands Organization for Health Research and Development; The Netherlands Organisation for Scientific Research; the Ministry of Health, Welfare, and Sport; and the Ministry of Youth and Families.

We thank the women who participated in the Amsterdam Born Children and their Development study. We also thank all the midwife practices and hospitals in Amsterdam, the Netherlands, for their assistance and data extraction. The ABCD study was financial supported by The Netherlands Organization for Health Research and Development. This work was supported by a clinical fellowship from ZonMw, project number 90700412 (RPP). V.W.V. Jaddoe received an additional grant from the Netherlands Organization for Health Research and Development (VIDI 016.136.361) and a European Research Council Consolidator Grant (ERC-2014-CoG-648916).

SUPPORTING INFORMATION

This manuscript is accompanied by additional materials including:

- Figure S1; A flowchart depicting the selection procedures of study subjects
- Table S1; Showing descriptive statistics of both cohorts
- Table S2 and S3; Showing the outcome of multiple imputation procedures
- Table S4; A prediction score for high TSH
- Table S5; Risk factor analyses for TPOAb positivity
- Table S6; C-statistics for various sensitivity analyses

CONTRIBUTION STATEMENTS: TIMK and DN performed analyses and were involved in writing the manuscript. PHLTB, MG, MM, LC, VVWJ, HT, TW, TJV, EWS, and TGV contributed to analyses and writing of the manuscript. YBR was involved in data collection and contributed to manuscript. RPP supervised analyses, was involved in writing of the manuscript and directed the project.

ETHICS APPROVAL: The general design, all research aims and the specific measurements in the Generation R Study have been approved by the Medical Ethical Committee of the Erasmus Medical Center, Rotterdam. Written informed consent was obtained from all participants.

TRANSPARENCY DECLARATION: RPP affirms that the manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned have been explained.

REFERENCES

1 Krassas, G.E., Poppe, K. & Glinoer, D. (2010) Thyroid function and human reproductive health. *ENDOCR REV* **31**, 702-755.

2 De Groot, L., Abalovich, M., Alexander, E.K., Amino, N., Barbour, L., Cobin, R.H., Eastman, C.J., Lazarus, J.H., Luton, D., Mandel, S.J., Mestman, J., Rovet, J. & Sullivan, S. (2012) Management of thyroid dysfunction during pregnancy and postpartum: an Endocrine Society clinical practice guideline. *J Clin Endocrinol Metab* **97**, 2543-2565.

3 Lazarus, J., Brown, R.S., Daumerie, C., Hubalewska-Dydejczyk, A., Negro, R. & Vaidya, B. (2014) 2014 European thyroid association guidelines for the management of subclinical hypothyroidism in pregnancy and in children. *Eur Thyroid J* **3**, 76-94.

4 Stagnaro-Green, A., Abalovich, M., Alexander, E., Azizi, F., Mestman, J., Negro, R., Nixon, A., Pearce, E.N., Soldin, O.P., Sullivan, S., Wiersinga, W., American Thyroid Association Taskforce on Thyroid Disease During, P. & Postpartum (2011) Guidelines of the American Thyroid Association for the diagnosis and management of thyroid disease during pregnancy and postpartum. *Thyroid* **21**, 1081-1125.

5 Lazarus, J.H., Bestwick, J.P., Channon, S., Paradice, R., Maina, A., Rees, R., Chiusano, E., John, R., Guaraldo, V., George, L.M., Perona, M., Dall'Amico, D., Parkes, A.B., Joomun, M. & Wald, N.J. (2012) Antenatal thyroid screening and childhood cognitive function. *N Engl J Med* **366**, 493-501.

6 Negro, R. & Stagnaro-Green, A. (2014) Diagnosis and management of subclinical hypothyroidism in pregnancy. *BMJ* **349**, g4929.

7 Chan, S. & Boelaert, K. (2015) Optimal management of hypothyroidism, hypothyroxinaemia and euthyroid TPO antibody positivity preconception and in pregnancy. *Clin Endocrinol (Oxf)* **82**, 313-326.

8 Henrichs, J., Ghassabian, A., Peeters, R.P. & Tiemeier, H. (2013) Maternal hypothyroxinemia and effects on cognitive functioning in childhood: how and why? *Clin Endocrinol (Oxf)* **79**, 152-162.

9 Roman, G.C., Ghassabian, A., Bongers-Schokking, J.J., Jaddoe, V.W., Hofman, A., de Rijke, Y.B., Verhulst, F.C. & Tiemeier, H. (2013) Association of gestational maternal hypothyroxinemia and increased autism risk. *Ann Neurol* **74**, 733-742.

10 Gyllenberg, D., Sourander, A., Surcel, H.M., Hinkka-Yli-Salomaki, S., McKeague, I.W. & Brown, A.S. (2015) Hypothyroxinemia During Gestation and Offspring Schizophrenia in a National Birth Cohort. *Biol Psychiatry*.

11 Korevaar, T.I., Muetzel, R., Medici, M., Chaker, L., Jaddoe, V.W., de Rijke, Y.B., Steegers, E.A., Visser, T.J., White, T., Tiemeier, H. & Peeters, R.P. (2015) Association of maternal thyroid function during early pregnancy with offspring IQ and brain morphology in childhood: a population-based prospective cohort study. *Lancet Diabetes Endocrinol*.

12 Yang, H., Shao, M., Chen, L., Chen, Q., Yu, L., Cai, L., Lin, Z., Zhang, C. & Lu, X. (2014) Screening strategies for thyroid disorders in the first and second trimester of pregnancy in China. *PLoS One* **9**, e99611.

13 Qian, W., Zhang, L., Han, M., Khor, S., Tao, J., Song, M. & Fan, J. (2013) Screening for thyroid dysfunction during the second trimester of pregnancy. *Gynecol Endocrinol* **29**, 1059-1062.

14 Horacek, J., Spitalnikova, S., Dlabalova, B., Malirova, E., Vizda, J., Svilias, I., Cepkova, J., Mc Grath, C. & Maly, J. (2010) Universal screening detects two-times more thyroid disorders in early pregnancy than targeted high-risk case finding. *Eur J Endocrinol* **163**, 645-650.

15 Vaidya, B., Anthony, S., Bilous, M., Shields, B., Drury, J., Hutchison, S. & Bilous, R. (2007) Detection of thyroid dysfunction in early pregnancy: Universal screening or targeted high-risk case finding? *J Clin Endocrinol Metab* **92**, 203-207.

16 Negro, R., Schwartz, A., Gismondi, R., Tinelli, A., Mangieri, T. & Stagnaro-Green, A. (2010) Universal screening versus case finding for detection and treatment of thyroid hormonal dysfunction during pregnancy. *J Clin Endocrinol Metab* **95**, 1699-1707.

17 Vaidya, B., Hubalewska-Dydejczyk, A., Laurberg, P., Negro, R., Vermiglio, F. & Poppe, K. (2012) Treatment and screening of hypothyroidism in pregnancy: results of a European survey. *Eur J Endocrinol* **166**, 49-54.

18 Srimatkandada, P., Stagnaro-Green, A. & Pearce, E.N. (2014) Attitudes of ATA Survey Respondents Toward Screening and Treatment of Hypothyroidism in Pregnancy. *Thyroid*.

19 Chang, D.L., Leung, A.M., Braverman, L.E. & Pearce, E.N. (2011) Thyroid testing during pregnancy at an academic Boston Area Medical Center. *J Clin Endocrinol Metab* **96**, E1452-1456.

20 Azizi, F., Amouzegar, A., Mehran, L., Alamdari, S., Subekti, I., Vaidya, B., Poppe, K., San Luis, T., Jr. & Akamizu, T. (2014) Screening and management of hypothyroidism in pregnancy: results of an Asian survey. *Endocrine Journal* **61**, 697-704.

21 Haddow, J.E., McClain, M.R., Palomaki, G.E., Kloza, E.M. & Williams, J. (2006) Screening for thyroid disorders during pregnancy: results of a survey in Maine. *Am J Obstet Gynecol* **194**, 471-474.

22 Jaddoe, V.W., van Duijn, C.M., Franco, O.H., van der Heijden, A.J., van Iizendoorn, M.H., de Jongste, J.C., van der Lugt, A., Mackenbach, J.P., Moll, H.A., Raat, H., Rivadeneira, F., Steegers, E.A., Tiemeier, H., Uitterlinden, A.G., Verhulst, F.C. & Hofman, A. (2012) The Generation R Study: design and cohort update 2012. *Eur J Epidemiol* **27**, 739-756.

23 van Eijsden, M., Vrijkotte, T.G.M., Gemke, R.J.B.J. & van der Wal, M.F. (2011) Cohort Profile: The Amsterdam Born Children and their Development (ABCD) Study. *International Journal of Epidemiology* **40**, 1176-1186.

24 Korevaar, T.I., Medici, M., de Rijke, Y.B., Visser, W., de Muinck Keizer-Schrama, S.M., Jaddoe, V.W., Hofman, A., Ross, H.A., Visser, W.E., Hooijkaas, H., Steegers, E.A., Tiemeier, H., Bongers-Schokking, J.J., Visser, T.J. & Peeters, R.P. (2013) Ethnic differences in maternal thyroid parameters during pregnancy: the Generation R study. *J Clin Endocrinol Metab* **98**, 3678-3686.

25 Dieguez, M., Herrero, A., Avello, N., Suarez, P., Delgado, E. & Menendez, E. (2014) Prevalence of thyroid dysfunction in women in early pregnancy: does it increase with maternal age? *Clin Endocrinol (Oxf)*.

26 Mannisto, T., Hartikainen, A.L., Vaarasmaki, M., Bloigu, A., Surcel, H.M., Pouta, A., Jarvelin, M.R., Ruokonen, A. & Suvanto, E. (2012) Smoking and early pregnancy thyroid hormone and anti-thyroid antibody levels in euthyroid mothers of the Northern Finland Birth Cohort 1986. *Thyroid* **22**, 944-950.

27 Mannisto, T., Surcel, H.M., Ruokonen, A., Vaarasmaki, M., Pouta, A., Bloigu, A., Jarvelin, M.R., Hartikainen, A.L. & Suvanto, E. (2011) Early pregnancy reference intervals of thyroid hormone concentrations in a thyroid antibody-negative pregnant population. *Thyroid* **21**, 291-298.

28 Korevaar, T.I., Chaker, L., Medici, M., de Rijke, Y.B., Jaddoe, V.W., Steegers, E.A., Tiemeier, H., Visser, T.J. & Peeters, R.P. (2016) Maternal Total T4 During the First Half of Pregnancy: Physiologic Aspects and the Risk of Adverse Outcomes in comparison to Free T4. *Clin Endocrinol (Oxf)*.

29 Steyerberg, E. (2008) *Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating*. Springer New York.

30 Steyerberg, E.W. & Harrell, F.E., Jr. (2015) Prediction models need appropriate internal, internal-external, and external validation. *J Clin Epidemiol*.

31 Korevaar, T.I., Schalekamp-Timmermans, S., de Rijke, Y.B., Visser, W.E., Visser, W., de Muinck Keizer-Schrama, S.M., Hofman, A., Ross, H.A., Hooijkaas, H., Tiemeier, H., Bongers-Schokking, J.J., Jaddoe, V.W., Visser, T.J., Steegers, E.A., Medici, M. & Peeters, R.P. (2013) Hypothyroxinemia and TPO-antibody positivity are risk factors for premature delivery: the generation R study. *J Clin Endocrinol Metab* **98**, 4382-4390.

32 Medici, M., Timmermans, S., Visser, W., de Muinck Keizer-Schrama, S.M., Jaddoe, V.W., Hofman, A., Hooijkaas, H., de Rijke, Y.B., Tiemeier, H., Bongers-Schokking, J.J., Visser, T.J., Peeters, R.P. & Steegers, E.A. (2013) Maternal thyroid hormone parameters during early pregnancy and birth weight: the Generation R Study. *J Clin Endocrinol Metab* **98**, 59-66.

33 Julvez, J., Alvarez-Pedrerol, M., Rebagliato, M., Murcia, M., Forns, J., Garcia-Esteban, R., Lertxundi, N., Espada, M., Tardon, A., Riano Galan, I. & Sunyer, J. (2013) Thyroxine levels during pregnancy in healthy women and early child neurodevelopment. *Epidemiology* **24**, 150-157.

34 Kumru, P., Erdogdu, E., Arisoy, R., Demirci, O., Ozkoral, A., Ardic, C., Ertekin, A.A., Erdogan, S. & Ozdemir, N.N. (2015) Effect of thyroid dysfunction and autoimmunity on pregnancy outcomes in low risk population. *ARCH GYNECOL OBSTET* **291**, 1047-1054.

35 Liu, H., Shan, Z., Li, C., Mao, J., Xie, X., Wang, W., Fan, C., Wang, H., Zhang, H., Han, C., Wang, X., Liu, X., Fan, Y., Bao, S. & Teng, W. (2014) Maternal subclinical hypothyroidism, thyroid autoimmunity, and the risk of miscarriage: a prospective cohort study. *Thyroid* **24**, 1642-1649.

36 Krassas, G.E. & Wiersinga, W. (2006) Smoking and autoimmune thyroid disease: the plot thickens. *Eur J Endocrinol* **154**, 777-780.

37 Korevaar, T.I., Steegers, E.A., de Rijke, Y.B., Schalekamp-Timmermans, S., Visser, W.E., Hofman, A., Jaddoe, V.W., Tiemeier, H., Visser, T.J., Medici, M. & Peeters, R.P. (2015) Reference ranges and determinants of total hCG levels during pregnancy: the Generation R Study. *Eur J Epidemiol* **30**, 1057-1066.

38 Berta, E., Samson, L., Lenkey, A., Erdei, A., Cseke, B., Jenei, K., Major, T., Jakab, A., Jenei, Z., Paragh, G., Nagy, E.V. & Bodor, M. (2010) Evaluation of the thyroid function of healthy pregnant women by five different hormone assays. *Pharmazie* **65**, 436-439.

39 Medici, M., Korevaar, T.I., Visser, W.E., Visser, T.J. & Peeters, R.P. (2015) Thyroid function in pregnancy: what is normal? *Clin Chem* **61**, 704-713.

FIGURE CAPTIONS

Figure 1. Figure 1: Figure shows the association of the outcome of the risk score with the prevalence of low FT4. The dotted horizontal black line depicts the baseline risk in the whole population.

Table 1. Risk factors for elevated maternal TSH during pregnancy.

		Odds ratios for potential risk factors amongst both populations combined		
Predictor	**Category/Measure**	**Univariable**	**Multivariable**	**Prediction model**
Age	*Per year**	1.05 (1.01, 1.09)	1.03 (0.98, 1.07)	
BMI	*Per point*	1.02 (0.99, 1.04)	1.03 (1.01, 1.06)	1.03 (1.00, 1.05)
Smoking	*No*	ref	ref	ref
	Stopped	0.79 (0.56, 1.12)	0.76 (0.53, 1.08)	0.77 (0.54, 1.09)
	Yes	0.70 (0.50, 0.99)	0.71 (0.50, 1.02)	0.70 (0.49, 0.99)
Parity	*0*	ref	ref	
	1	0.87 (0.68, 1.10)	0.82 (0.64, 1.05)	
	2	0.87 (0.59, 1.29)	0.83 (0.55, 1.24)	
	≥3	0.47 (0.21, 1.06)	0.47 (0.20, 1.08)	

Ethnicity	*Dutch*	ref	ref	ref
	Other western	0.87 (0.60, 1.28)	0.88 (0.60, 1.28)	0.88 (0.60, 1.28)
	Moroccan	0.40 (0.22, 0.73)	0.40 (0.22, 0.76)	0.37 (0.20, 0.68)
	Surinamese	0.37 (0.20, 0.66)	0.38 (0.21, 0.69)	0.35 (0.19, 0.63)
	Turkish	0.76 (0.48, 1.19)	0.82 (0.50, 1.33)	0.75 (0.47, 1.19)
	Asian	1.49 (0.92, 2.42)	1.49 (0.91, 2.43)	1.42 (0.87, 2.31)
	Other non-western	0.56 (0.39, 0.81)	0.57 (0.39, 0.83)	0.55 (0.38, 0.79)
Previous miscarrriage or stilborn	*Yes*	1.03 (0.77, 1.37)	1.02 (0.76, 1.37)	
Gestational age	*Per week*	1.02 (0.97, 1.07)	1.04 (0.99, 1.10)	

Table 2. Discriminative ability of prediction models for elevated TSH, TPOAb positivity and decreased FT4.

	Populations combined	*Generation R study*	*ABCD study*
Outcome	**C-statistic** [a]	**C-statistic** [b]	**C-statistic** [c]
Elevated TSH	0.60 (0.57*)	0.58	0.58
Decreased FT4	0.76 (0.75*)	0.75	0.72

Table 3. Risk factors for decreased maternal FT4 during pregnancy.

		Odds ratios for potential risk factors amongst both populations combined		
Predictor	**Category/Measure**	**Univariable**	**Multivariable**	**Prediction model**
Age	*Per year**	1.08 (1.03, 1.13)	1.06 (1.01, 1.11)	1.06 (1.01, 1.10)
BMI**	*25 vs. 20*	2.20 (1.89, 2.58)	2.02 (1.70, 2.40)	2.05 (1.73, 2.42)
	30 vs. 20	4.21 (3.17, 5.59)	3.59 (2.63, 4.90)	3.67 (2.70, 4.99)
Smoking	*No*	ref	ref	ref
	Stopped	1.10 (0.71, 1.71)	1.44 (0.91, 2.27)	1.41 (0.89, 2.22)
	Yes	1.72 (1.21, 2.45)	1.65 (1.15, 2.36)	1.60 (1.12, 2.29)
Parity	*0*	ref	ref	ref
	1	1.31 (0.96, 1.78)	1.11 (0.80, 1.53)	1.12 (0.82, 1.55)
	2	2.16 (1.44, 3.23)	1.35 (0.88, 2.08)	1.42 (0.92, 2.17)
	≥3	4.92 (3.15, 7.70)	2.02 (1.21, 3.37)	2.21 (1.34, 3.63)

Ethnicity	*Dutch*	ref	ref	
	Other western	1.39 (0.84, 2.29)	1.34 (0.81, 2.23)	
	Moroccan	2.30 (1.45, 3.65)	1.33 (0.82, 2.18)	
	Surinamese	1.93 (1.19, 3.14)	1.33 (0.80, 2.22)	
	Turkish	1.78 (1.06, 2.98)	1.10 (0.64, 1.90)	
	Asian	1.48 (0.68, 3.21)	1.33 (0.60, 2.93)	
	Other non-western	2.14 (1.48, 3.08)	1.65 (1.13, 2.42)	
Previous miscarriage or stillborn	*Yes*	1.28 (0.88, 1.86)	1.14 (0.79, 1.66)	
Gestational age	*Per week*	1.41 (1.32, 1.50)	1.35 (1.26, 1.45)	1.37 (1.28, 1.46)

.

Table 4. Clinical prediction score for decreased maternal FT4 levels during pregnancy.

Age		BMI		Week of pregnancy		Parity		Smoking	
≤30	0	**≤20**	0	**≤10**	0	**0**	0	**None**	0
31-33	1	**21-25**	3	**11-14**	10	**1**	1	**Stopped**	3
34-35	2	**26-29**	9	**15-18**	20	**2**	3	**Yes**	5
36-37	3	**≥30**	15			**≥3**	8		
38-39	4								
≥40	5								

Total score	Risk of decreased FT4 (%)	N(%) per group
Whole population	**2.8**	**9415 (100)**
≤10	**<0.5**	**818 (8.7)**
11-17	**0.5 - 1**	**2173 (23.1)**
18-24	**1 - 2**	**2926 (31.1)**
25-31	**2 - 4**	**2034 (21.6)**
32-37	**4 - 7**	**915 (9.7)**
38-45	**7 - 15**	**458 (4.9)**
≥46	**>15**	**91 (1.0)**

Figure 1 The risk of low FT4 according to the risk score chart