Controlled Antenatal Thyroid Screening II: effect of treating maternal sub-optimal thyroid function on child behaviour

Charlotte Hales, Peter N Taylor, Sue Channon, Kirsten McEwan, Anita Thapar, Kate Langley, Ilaria Muller, Mohd S Draman, Colin Dayan, John W Gregory, Onyebuchi Okosieme, John H Lazarus, D Aled Rees and Marian Ludgate

Centre for Endocrine and Diabetes Sciences, School of Medicine, Cardiff University, UK (C Hales PhD, P N Taylor PhD, O Okosieme MD, I Muller PhD, M S Draman PhD, JW Gregory MD, C Dayan PhD, JH Lazarus MD, M Ludgate PhD);

Centre for Trials Research, Cardiff University, UK (S Channon DClinPsych, K McEwan PhD); School of Psychology, Cardiff University, UK (Kate Langley)Division of Psychological Medicine and Clinical Neurosciences, School of Medicine, Cardiff University, UK (Anita Thapar); Neuroscience and Mental Health Research Institute, Cardiff University, UK (D Aled Rees PhD).

SHORT TITLE; Maternal thyroid function effects on child behaviour.

KEY WORDS; Thyroid, Pregnancy, Thyroxine, ADHD, Autism, Childhood

CORRESPONDENCE TO: Prof Marian Ludgate, Division of Infection and Immunity, School of Medicine, Cardiff University, Cardiff, Wales UK, ludgate@cardiff.ac.uk

FUNDED BY: The Charles Wolfson Trust, Action Medical Research (project code GN2033)/The Henry Smith Charity (20122759 GN 2033, grants to ML) and The American Thyroid Association (project code ATA-2014-033.R1, grant to PNT)

DISCLOSURE STATEMENT: None of the authors have any conflicting interests to declare.

Word Count: 4036

ABSTRACT

Context & Objectives The Controlled Antenatal Thyroid Screening (CATS) study was the first randomised controlled trial to investigate effects of treating suboptimal-gestational-thyroid-function (SGTF) on child cognition. Since observational studies indicated that SGTF may also increase symptoms of autism and attention deficit hyperactivity disorder (ADHD), the CATS cohort was used to investigate whether treatment of mothers affected their children's behavior.

Design & Participants Mothers (N=475) completed 3 questionnaires, strengths and difficulties (SDQ), attention deficit hyperactivity disorder (ADHD) and social communication (SCQ, used as a screen for autism spectrum disorder (ASD)), about their children (aged 9.5 years). Group comparisons of total scores, numbers of children above clinical thresholds and association between high maternal FT4 (>97.5th percentile of the UK cohort, 'over-treated') and child neurodevelopment were reported.

Results There were no differences in total scores between normal-GTF (n=246), treated (n=125) and untreated (n=104) SGTF groups. More children of treated mothers scored above clinical thresholds, particularly the over-treated. Scores were above thresholds in SDQ conduct 22% vs 7%, SCQ total scores 7% vs 1%, and ADHD hyperactivity 17% vs 5% when comparing over-treated (n=40) and untreated (N=100) respectively. We identified significantly higher mean scores for SDQ conduct (adjusted mean difference (AMD) 0.74, 95% confidence interval (CI) 0.021 to 1.431, P=0.040, effect size 0.018) and ADHD hyperactivity (AMD 1.60, CI 0.361 to 2.633, P=0.003, effect size 0.028) comparing over-treated with normal-GTF children.

Conclusions There was no overall association between SGTF and offspring ADHD, ASD or behavior questionnaire scores. However, children of 'over-treated' mothers displayed significantly more ADHD symptoms and behavioral difficulties than normal-GTF. Thyroxine supplementation during pregnancy requires monitoring to avoid over-treatment.

PRÉCIS

We investigated effects of treating sub-optimal gestational thyroid function (SGTF) on child behaviour. No associations found between SGTF & offspring behaviour; children of 'over-treated' mothers had significantly more difficulties than untreated.

INTRODUCTION

The cognitive impairment displayed in children with congenital hypothyroidism or born to mothers with iodine deficiency illustrates the essential role of thyroid hormones, triiodothyronine (T3) and thyroxine (T4), in brain development (1). Although fetal T3 receptors in the human brain can be detected from about 10 weeks gestation and increase 50 fold by 16 weeks, the thyroid develops relatively late with full function being acquired at around 36 weeks gestation (1); thus thyroid hormones must be supplied via placental transfer (2). Sub-optimal gestational thyroid function (SGTF) (classified as low T4 and/or high thyroid stimulating hormone (TSH) values, or both) in pregnancy can occur in women with no previous thyroid dysfunction. Children born to mothers with SGTF have been reported to have lower IQ compared with children born to euthyroid mothers (3-11). However, two large randomized controlled trials (RCTs) found no evidence of thyroid hormone treatment improving the IQ of children born to mothers with SGTF (12,13) tested at ages 3 and 5 respectively. These results were confirmed recently in a follow-up to the first study, in children aged 9 (14).

In addition to intelligence, SGTF may also affect other aspects of child neurodevelopment; specifically attention deficit hyperactivity disorder (ADHD, characterised by hyperactivity, inattention and impulsiveness) and autism spectrum disorder (ASD, characterised by social and communication problems, and repetitive behaviors). Multiple observational studies have reported higher ADHD scores in children born to mothers with SGTF, particularly hypothyroxinaemia, compared to controls (3,15-19). Childhood ASD has been reported to be increased in those born to mothers with SGTF (20,21). Maternal thyroid peroxidase autoantibodies, markers of thyroid autoimmunity, are also associated with increased ADHD (22) and ASD risk

(23) in offspring. In contrast, others report no link between SGTF and ADHD or ASD symptoms in the child (20,24,25).

The two RCTs described above (12,13) included additional psychological assessments of behavioral and social competency plus attention. Neither reported any differences in these parameters in children of treated and untreated mothers; although they were not powered specifically for these aspects. Furthermore, data from children of mothers with normal-GTF were not collected. The current study is an extension of the first RCT of treatment of SGTF during pregnancy (12) in which the offspring were examined for behavior, social competency and attention after long-term follow-up. The inclusion of children of normal-GTF mothers enabled us to assess whether there is a deficit in neurodevelopment requiring treatment.

Furthermore, since a number of mothers demonstrated high FT4 concentrations suggesting a degree of over-treatment, we were also interested in the effects of high-FT4 on child neurodevelopment outcomes. The rationale for this derives from studies by Korevaar and colleagues, who reported adverse effects on brain morphology and cognition of both too little and too much thyroid hormone (26). In addition, a recent study from Maraka et al, demonstrated worse obstetric outcomes such as pre-term delivery and pre-eclampsia in treated women with SGTF (27).

MATERIALS AND METHODS

The Controlled Antenatal Thyroid Screening study

We report a follow-up to a treatment trial, the Controlled Antenatal Thyroid Screening Study (CATS) (12), in which women were screened at a median of 12 weeks 3 days' gestation between June 13, 2002 and May 31, 2006. Briefly, in CATS-I (12) midwives and obstetric care-givers in the UK and Italy recruited ~22,000 women (excluding

history of thyroid disease, twin pregnancies, maternal age <18 years or gestational age >15 weeks and 6 days), at their first antenatal hospital appointment. Participants were randomized, by computer generated block design, either to screen or control groups; both provided serum samples at recruitment with the screen group having their thyroid function tested immediately and the controls after their child was born (sera stored at -40°C). If the mother's FT4 was <2.5th percentile and/or TSH >97.5th percentile, they were classified as having SGTF; percentiles being calculated from the CATS cohort. Women in the screen group with SGTF were treated with levothyroxine (starting dosage 150μg) for the remainder of their pregnancies. Dosages were adjusted where necessary, to maintain a serum TSH of 0.1-1.0mIU/L, following measurement of TSH and FT4 6 weeks after treatment was initiated and at 30 weeks gestation. SGTF women diagnosed after delivery were advised to visit their general practitioner for further management. The primary outcome was children's IQ at age 3 from 390 (303 in the UK) treated (screen) and 404 (306 in the UK) untreated (control) mothers (assessors blinded to treatment group). A further 20,789 (15,593 in the UK) women made up the normal-GTF group and recruitment ended when sufficient numbers, determined by prior power calculation, had been reached. Figure 1 shows the study flow chart of the UK participants.

Follow-up of children at ages 7-10 years; current data

Inclusion criteria were mothers living in the UK and from the CATS-I cohort whose children were aged 7 to 10 years. Offspring follow-up data were collected between August 8, 2011 and August 7, 2015. The target sample size of 480 participants was pre-determined and informed by prior power calculations to assess cognition. Samples of 120 from each of the treated and untreated SGTF groups provided 80% power with

a 5% two-sided significance level to detect a 1.97 increase in odds of IQ <85 in untreated SGTF offspring assuming treated SGTF offspring have a mean IQ of 100 with a SD of 15. An additional 240 participants from the normal GTF group, randomly selected from the 15,593 UK cohort, were used to assess interaction with maternal thyroid status and levothyroxine treatment on offspring IQ.

For full details see Hales et al 2014 (28).

CATS mothers from the UK SGTF treated and untreated groups (n=609) were invited to participate by letter. The Welsh Demographics Service and Patient Data Registrar provided current addresses. Those without SGTF in the control and screen branches of the RCT, were pooled (UK n=15,593), and named 'normal-GTF'; a random sample of 4,000 from this group was also invited to participate. Data were collected at a research centre, via a home visit, or by post; this approach was used to maximise the follow-up rate. Informed consent was provided by all mothers who completed the questionnaires.

Results of cognitive assessments in the children have been reported (14). Cardiovascular, metabolic, bone data and DNA samples were also collected in those participants attending the visit at the research centre but are not described here (ms in preparation).

Mothers (mean age 31 (SD 5.12) years) were asked to complete three questionnaire measures on their children (mean age 9.5 (SD 0.77) years): The Strengths and Difficulties Questionnaire (SDQ) (29), The Child ADHD Questionnaire (ADHDq) (30), and the Social Communication Questionnaire (SCQ) (31). The SDQ consisted of 25 items, scored by a Likert scale (0-2), grouped into five subscales; Hyperactivity (5 ADHD items), Emotional symptoms, Conduct problems, Peer problems, and Pro-

social behavior as well as a total score. The SDQ provides a measure of child mental health difficulties, has been internationally validated and has good psychometric properties (32-35). The ADHDq consists of 18 items addressing ICD-10 symptoms by a Likert scale (0-3, Thapar et al.(30)); subscales include inattention, hyperactivity, and impulsivity. The Social Communication Questionnaire has good external validity (36) and is one of the most widely adopted screening tools for ASD (37). The SCQ generates a total score from 'yes'/'no' answers; a score of ≥15, indicates possible ASD (38). Data were collected, inputted, and cleaned by one researcher who was blinded to participant study group.

Statistical analysis

All analyses (unless otherwise stated) were adjusted for child gender, maternal age at time of consent into CATS, whether the mother breastfed for over one month, and quintile of social deprivation (the lower the quintile the more deprived, quintiles enabled comparisons between Welsh and English postcodes); these variables have been reported previously to be associated with questionnaire scores (39-47). Child age was not included as a covariate given the narrow age range of the children (3 years).

Descriptive statistics were calculated to identify any potential recruitment bias in the three groups; continuous variables are presented as means (standard deviations) and categorical data as percentages (unless otherwise stated). Thyroid function test results at time of consent into CATS are presented per group, as well as the treated SGTF group's additional test results. As a proportion of CATS mothers had high FT4 concentrations (26), the study was able to analyse an 'over-treatment' effect; a FT4 measurement >17.7pmol/L at either 20 weeks or 30 weeks' gestation as defined by the top 2.5th percentile of FT4 in the entire CATS UK cohort at consent into the RCT.

The over-treatment value was calculated using a large sample-size (n=16,346) and was higher than the upper limit of the regional reference range. Mean FT4 and median TSH values of those who were over-treated were compared to the optimally treated mothers (FT4 <17.6pmol/L). Effects of the treated SGTF group's thyroid function tests throughout pregnancy were related to the questionnaire subdomains and total scores.

List-wise deletion was applied for questionnaires having >15% missing data (previously reported 10-20% acceptable (48,49), with >15% requiring consideration (50)), leading to exclusion of 21 participants. Of the remaining 454 participants (normal-GTF=237, treated SGTF=117, untreated SGTF=100), 54 (11.9%) had missing data; total values of missing data were n=85 (0.2%). The remaining participants' data were analysed by Little's missing completely at random test (non-significant), with missing values replaced by the 'series mean' in SPSS (51).

All Kolmogorov-Smirnov (and Shapiro-Wilk) normality tests were returned as $p < 0.001$ for all questionnaires on every totalled domain. However, means and medians appeared close and skewness and kurtosis ranges were largely just above ± 1 and thus were returned as normal with caution.

In the primary analysis, we compared the mean total scores from the three questionnaires administered to the normal-GTF, treated and untreated SGTF groups. This was by a multivariate analysis of covariance (MANCOVA), which was followed by univariate tests and pairwise comparisons that were Bonferroni corrected, where relevant.

In the secondary analysis, the frequency of children scoring above the questionnaire thresholds that indicate likely disorder was presented to enable comparisons between the groups. For SDQ, the clinically significant cut-offs were adopted, those classified

as ‘high’ were in the 90-95th percentile range, and ‘very high’ were above the 95th percentile (29). For the ADHDq, the authors calculated the SDs (30); scores >2SDs were reported. For the SCQ, scores >15 were identified as the child potentially having an ASD (31); a UK sample found a prevalence of 4-5% for scores >15 (52).

The exploratory analysis investigated the effect of treatment. Mean scores from the questionnaires (total and subdomains) were compared between the normal-GTF, untreated SGTF, optimally treated SGTF (FT4 <17.6pmol/L), and over-treated SGTF (FT4 >17.7pmol/L). Further exploratory analyses are in the Supplemental Material, including effects of hypothyroxinemia or subclinical hypothyroidism and varying levels of FT4 following treatment.

Standard regression analysis was performed to compare behavior of participating children at age 3 (Child behavior check-list) and age 9 (strengths and difficulties questionnaire).

All statistical analyses were performed with SPSS version 20 software. P<0.05 was considered significant.

RESULTS

Characteristics of the cohort

A total of 485 women/child pairs participated but 10 were excluded (9 ineligible, one whose results were deemed unreliable due to their approach to the assessment) leaving 475 who completed the questionnaires (normal-GTF=246, treated SGTF=125, and untreated SGTF=104); table 1 reports their demographic data. The SGTF groups had younger children participating (by four months) than the normal-GTF group, and the treated SGTF also had significantly younger mothers by ten months. The treated SGTF had a higher median baseline TSH compared to the untreated SGTF group; this

suggested a potential recruitment bias and was reflected in SGTF group age 3 to age 9 comparisons; supplemental data table 1 (53).

Pregnancy thyroid function tests of the treated SGTF group at 20 weeks' gestation identified a mean (SD) FT4 of 16.2pmol/L (2.89), reducing to 15.5pmol/L (2.47) at 30 weeks; median TSH values at 20 and 30 weeks gestation were 0.3mIU/L at both time points. Of these women, 41 were over-treated, with FT4 concentration >17.7pmol/L at 20 or 30 weeks' gestation (see supplemental material for figures); with 13 of these sustaining over-treatment throughout their pregnancies. The mean FT4 concentrations at 20 weeks and 30 weeks for the over-treated group were 19.2pmol/L (2.53) and 17.8pmol/L (2.46), respectively, whilst median TSH values at these times were below target at 0.08mIU/L and 0.05mIU/L, respectively. In the remaining optimally treated SGTF group, the mean FT4 concentrations at 20 weeks' and 30 weeks' gestation were 14.8pmol/L (1.78) and 14.5pmol/L (1.73), respectively, with median TSH values of 0.52mIU/L and 0.35mIU/L.

Main analyses

After removing participants with >15% missing data, 454 were analysed, revealing no significant differences between the means of the 3 administered questionnaires (table 2) in the normal-GTF, treated SGTF and untreated SGTF groups (P=0.261). Total questionnaire scores comparing the three participant groups were significantly affected by child gender (/sex), mother age at time of consent and social deprivation (P=0.001, P=0.030, and P=0.001, respectively). Higher questionnaire scores were more likely for male children and children born to younger mothers during pregnancy or to families with greater social deprivation.

Secondary and Exploratory Analyses

At the point of recruitment into CATS-I, there was a negative association between maternal FT4 level and aspects of child behaviour assessed at age 9; i.e. children of mothers with low FT4 before any treatment tended to have higher questionnaire scores indicating more behaviour problems. However, the association between child behaviour problems and FT4 in treated mothers was positive at 20 weeks' gestation (figure 2). Thus we investigated the possible effects of over-treatment as defined by the top 2.5th percentile of FT4 in the CATS UK cohort at recruitment (N=16,346, FT4>17.7 pmol/L). We observed significantly higher mean scores for some domains: SDQ conduct (adjusted mean difference (AMD) 0.74, 95% confidence interval (CI) 0.021 to 1.431, P=0.040, effect size 0.018) and ADHD hyperactivity (AMD 1.60, CI 0.361 to 2.633, P=0.003, effect size 0.028) in over-treated compared with normal-GTF.

Furthermore, individuals within each group had 'high scores' on domains reported in the analyses (table 3); these were defined as clinically significant cut-offs (SDQ), standard deviations calculated by Thapar et al.(30) (ADHDq), and a score >15 indicative of a child potentially having an ASD (31) (SCQ). The largest and significant percentage differences were noted on SDQ Conduct, SDQ Hyperactivity, ADHDq Hyperactivity, and SCQ Total; the treated SGTF had proportionately more than twice as many individuals in the highest classifications compared to the normal-GTF and untreated SGTF, which appeared to be driven by those children born to over-treated mothers. When comparing the normal-GTF and untreated SGTF children, we did not observe any difference between numbers of individuals in the highest classifications.

Effect of subclinical hypothyroidism, hypothyroxinemia and overt hypothyroidism

Similar non-significant differences were found between the normal GTF (n=237), treated (n=67), and untreated (n=49) subclinical hypothyroid groups for total

questionnaire scores (P=0.909). When exploring the effect of treatment to the 17.7pmol/L threshold, there was also no difference in mean scores; supplemental data table 2

Total questionnaire scores were not significantly different between the normal GTF (n=237), treated (n=36) and untreated (n=44) hypothyroxinaemia groups; supplemental data table 3.

There were 8.55% (n=10) of the treated SGTF group, and 9.00% (n=9) of the untreated SGTF group classified as having overt hypothyroidism (OH). Comparing these groups (combined) with the normal GTF for total questionnaire scores identified a significant difference at the multivariate level (V=0.063, F(6,496)=2.669, P=0.015, η_p^2=0.031), with a difference found for ADHD Total scores (F(2,249)=4.680, P=0.010, η_p^2=0.036). The untreated OH group had significantly higher scores compared to the normal GTF for ADHD Total (P=0.022 (0.945, 16.876)).

When exploring the effect of treatment to the 17.7pmol/L threshold for maternal FT4, there was a difference at the multivariate level (V=0.203, F(27,726)=1.948, P=0.003 η_p^2=0.068). At the univariate level SDQ Conduct (F(3,248)=4.402, P=0.005, η_p^2=0.051), ADHD Hyperactivity (F(3,248)=5.665, P=0.001, η_p^2=0.064), and ADHD Total (F(3,248)=3.487, P=0.016, η_p^2=0.040) were significantly different. The over-treated OH group had significantly higher scores for SDQ Conduct (P=0.038 (0.072, 4.273) and ADHD Hyperactivity (P=0.020 (0.357, 6.597) compared to the normal GTF. Again, for ADHD Total, the untreated OH had significantly higher scores than the normal GTF (P=0.045 (0.117, 7.693).

Children born to mothers who were untreated for OH in pregnancy, had higher scores for the Child ADHD Questionnaire, optimal treatment improved this with some means

moving closer to those of the normal GTF; supplemental data table 4. Over-treatment for OH had a significant effect on the same domains identified with the entire SGTF sample; SDQ Conduct and ADHD Hyperactivity. Caution is advised, however, as the sample sizes are small for OH.

Finally, we applied standardised regression analysis to compare behavior assessed at age 3, using the CBCL, and at age 9, using the SDQ. We observed that behavior at ages 3 and 9 was highly correlated B (Std) = 0.54 (95%CI 0.48 – 0.66) p< 0.001, with variation at age 3 predicting 30% of the variation at age 9. Adjusting for other key factors such as sex, social class, maternal age and breastfeeding had no substantial impact on relationship or effect estimates.

DISCUSSION

We have applied well-validated questionnaires to assess neurodevelopment and behavior in a large cohort of children born to mothers recruited to CATS-I, the first RCT to explore the effect of levothyroxine treatment on childhood outcomes. In addition, our inclusion of children from normal-GTF mothers allowed us to investigate whether SGTF has any effect on conduct and hyperactivity.

We found no difference in the mean total scores of questionnaires to assess ADHD and ASD in children of mothers with normal-GTF compared with SGTF, whether treated or not. Results were the same when considering treated and untreated hypothyroxinemia and subclinical hypothyroidism separately, although numbers were inevitably smaller in the sub-sets. Boys, children born to younger mothers or to socially deprived families were more likely to have higher scores. FT4 was negatively associated with measures of hyperactivity in children of treated mothers at recruitment but positively associated after thyroxine supplementation. Reflecting this switch, there were significantly more children of treated mothers with scores above clinical

thresholds for conduct and hyperactivity than normal-GTF mothers. In particular, there were either significantly more children of over-treated mothers above defined cut-offs for these parameters or with mean scores significantly higher than the untreated SGTF or normal-GTF groups respectively.

We are aware that our study sample may be biased but group means were within ±1SD of the ADHDq mean (24) or similar to the means reported for the SDQ and SCQ from other large UK cohorts (52) suggesting that participants are representative. However, numbers in the SGTF groups, especially untreated, were below our target and recruitment was a challenge from the outset, mainly because participants had re-located and did not respond to invitation.

Incomplete questionnaires were minimised by spot-checks and thus less than 1% of participants were excluded. However, questionnaires were completed only by the mothers and the study would have benefited from the child's mental health and neurodevelopment being assessed by another independent source, such as their teacher. Furthermore, significant mean effects are mainly seen on individual domains of the questionnaires, although these sub-domains mostly relate to hyperactivity.

In the two RCTs (12, 13) of the effects of SGTF treatment on child outcomes, neither reported any differences (mean scores) in behavior, social competency or attention in children of treated and untreated mothers at age 3 and 5 respectively. Our results in children aged 9 confirm both studies and demonstrate that behavior scores at age 3 predict 30% of the variability in behavior scores at age 9.

In contrast to other studies, we found no differences in ADHD (3,15-18) and ASD (20,21) questionnaire domains between the SGTF and normal-GTF groups. However, we recognise that our findings are based on a smaller sample size, particularly of normal-GTF, than the large observational cohorts.

A small number of mothers in our study had overt hypothyroidism (n=19) and their children, whether in the untreated or over-treated groups, had more conduct, total ADHDq and hyperactivity symptoms then those in the normal-GTF group (supplemental material). In studies of congenital hypothyroidism, similar findings have also been reported: increased hyperactivity with higher levothyroxine dosages (49), and if the child had a high T4, they were significantly more distractible (50). Haddow et al.(3) identified more children with attention difficulties born to mothers who were treated in pregnancy, compared to untreated; however this was not an RCT and there was limited information on thyroid hormone levels of those who were treated. Recent research confirms that both high and low maternal FT4 may be detrimental to offspring in respect of intelligence and brain morphology (26). Our results also hint at a biphasic effect of maternal FT4 on neurodevelopment as illustrated by the switch from negative to positive association between FT4 and measures of hyperactivity before and after initiation of treatment. Mothers received a starting Levothyroxine dose of 150 µg daily after which the dose was adjusted to maintain TSH in the lower end of the reference range according to recommended treatment targets at the time of the study (54). This resulted in one third having an FT4 concentration above 17.7pmol/L, a threshold derived from the top 2.5th percentile of the UK CATS-I cohort (n=16 346), a cohort larger than those used to set reference ranges, and which excluded mothers with a history of thyroid disease. Thus subclinical hyperthyroidism may have inadvertently been induced in the over-treated group (based on current guidelines of a TSH value

<0.1mIL/l (55). A nationwide study in Europe (n=857, 014) reported that if the mother was hyperthyroid throughout pregnancy, the child was at an increased risk (OR 1.23, 1.05 to 1.44) of being diagnosed with ADHD (20). Yau et al.(25) reported that lower maternal TSH values were associated with more ASD symptoms in the offspring; whilst individuals with resistance to thyroid hormone (n=75) displayed a significant positive correlation between T4 levels and both ADHD hyperactivity and impulsivity symptoms (56). The same study also found a significant positive correlation between T3 and both ADHD hyperactivity and impulsivity in individuals without resistance to thyroid hormone (case family members, n=77). Thus, our results confirm these earlier studies indicating that higher levels of thyroid hormones can affect ADHD symptoms.

Our results are relevant to clinicians managing pregnant women with hypothyroidism. If a woman is already prescribed levothyroxine before pregnancy, current guidelines (55) advise an increase in the levothyroxine dose of 20-30% during pregnancy. Recommended starting doses for women diagnosed during pregnancy range from 100 µg for OH but 25-50 µg daily for subclinical hypothyroidism and depend on 'normal reference ranges' for T4 and TSH according to geographic region and ethnic origin (57,58). In all cases treatment focuses on reducing TSH levels, and FT4 levels are seldom monitored, hence over-treatment in pregnancy may be more commonplace than anticipated (57,59,60). Of note, 17% of the women found to be over-treated in the present study had TSH values within the normal reference range. We hypothesise that there could be a beneficial mid-range of maternal FT4 during pregnancy for the offspring; too little or too much may affect the child. To investigate a possible 'tipping point', when differences in mental health and neurodevelopment between groups were no longer significant, FT4 thresholds of 17, 16, 15 and 14pmol/L at 20 weeks' gestation were applied (supplemental data table 5). It was identified that most differences

399 diminished below the 17.7pmol/L category, with lowering FT4 also increasing
400 questionnaire scores; this also supports the biphasic effect of maternal FT4.

401 While screening and treatment of thyroid dysfunction in pregnancy will clearly benefit
402 patients with overt hypothyroidism, the benefits for SGTF remain to be proven. In the
403 light of our findings screening programmes should be carefully considered in order to
404 avoid the unintended consequences of over-treatment. To fully explore the impact of
405 SGTF and its treatment on behavioral problems requires large-scale studies with early
406 intervention. Before embarking on such studies, consensus needs to be reached on
407 trimester-specific reference ranges for TSH, FT4 and FT3, all derived from large
408 groups of healthy women, free of thyroid autoantibodies, in an iodine sufficient area
409 and with no history of thyroid disease.

410 We conclude that SGTF and optimal management of levothyroxine treatment during
411 pregnancy have no effect on ADHD and ASD-symptoms of offspring at age 9. The
412 children of women with FT4>97.5th percentile had more conduct, ADHDq, and ASD-
413 symptoms and suggests that levothyroxine dose in pregnancy needs to be carefully
414 monitored.

415

ACKNOWLEDGMENTS

CATS-II was funded by The Charles Wolfson Trust, Action Medical Research (project code GN2033)/The Henry Smith Charity (20122759 GN 2033, grants to ML) and The American Thyroid Association (project code ATA-2014-033.R1, grant to PNT) for genotyping data. We are extremely grateful to the children, parents and families who participated in the study. Special thanks are extended to Dionne Shillabeer, Julie Pell, Julie Evans, Sophie Fuller, and Beverley Carey for their dedicated support of the CATS project.

DATA AVAILABILITY

Data sharing is not applicable to this article as no datasets were generated or analysed during the current study.

REFERENCES

1. J.G.Thorpe-Beeston KHN, McGregor AM Fetal Thyroid Function. Thyroid 1992; 2:207-217

2. Chan SY, Vasilopoulou E, Kilby MD. The role of the placenta in thyroid hormone delivery to the fetus. Nature Clinical Practice Endocrinology and Metabolism 2009; 5:45-54

3. Haddow JE, Palomaki GE, Allan WC, Williams JR, Knight GJ, Gagnon J, O'Heir CE, Mitchell ML, Hermos RJ, Waisbren SE, Faix JD, Klein RZ. Maternal thyroid deficiency during pregnancy and subsequent neuropsychological development of the child. New England Journal of Medicine 1999; 341:549-555

4. Li Y, Shan Z, Teng W, Yu X, Li Y, Fan C, Teng X, Guo R, Wang H, Li J, Chen Y, Wang W, Chawinga M, Zhang L, Yang L, Zhao Y, Hua T. Abnormalities of maternal thyroid function during pregnancy affect neuropsychological development of their children at 25-30 months. Clinical Endocrinology (Oxf) 2010; 72:825-829

5. Su PY, Huang K, Hao JH, Xu YQ, Yan SQ, Li T, Xu YH, Tao FB. Maternal thyroid function in the first twenty weeks of pregnancy and subsequent fetal and infant development: A prospective population- based cohort study in China. Journal of Clinical Endocrinology and Metabolism 2011; 96:3234-3241

6. Klein RZ, Sargent JD, Larsen PR, Waisbren SE, Haddow JE, Mitchell ML. Relation of severity of maternal hypothyroidism to cognitive development of offspring. Journal of Medical Screening 2001; 8:18-20

7. Ghassabian A, El Marroun H, Peeters RP, Jaddoe VW, Hofman A, Verhulst FC, Tiemeier H, White T. Downstream effects of maternal hypothyroxinemia in early pregnancy: Nonverbal IQ and brain morphology in school-age children. Journal of Clinical Endocrinology and Metabolism 2014; 99:2383-2390

8. Henrichs J, Bongers-Schokking JJ, Schenk JJ, Ghassabian A, Schmidt HG, Visser TJ, Hooijkaas H, De Muinck Keizer-Schrama SMPF, Hofman A, Jaddoe VVW, Visser W, Steegers EAP,

Verhulst FC, De Rijke YB, Tiemeier H. 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

9. Suárez-Rodríguez M, Azcona-San Julián C, Alzina de Aguilar V. Hypothyroxinemia during pregnancy: The effect on neurodevelopment in the child. International Journal of Developmental Neuroscience 2012; 30:435-438

10. Pop VJ, Brouwers EP, Vader HL, Vulsma T, Van Baar AL, De Vijlder JJ. Maternal hypothyroxinaemia during early pregnancy and subsequent child development: A 3-year follow-up study. Clinical Endocrinology 2003; 59:282-288

11. Berbel P, Mestre JL, Santamaría A, Palazón I, Franco A, Graells M, González-Torga A, de Escobar GM. Delayed neurobehavioral development in children born to pregnant women with mild hypothyroxinemia during the first month of gestation: the importance of early iodine supplementation. Thyroid 2009; 19:511-519

12. Lazarus JH, Bestwick JP, Channon S, Paradice R, Maina A, Rees R, Chiusano E, John R, Guaraldo V, George LM, Perona M, Dall'Amico D, Parkes AB, Joomun M, Wald NJ. Antenatal thyroid screening and childhood cognitive function. New England Journal of Medicine 2012; 366:493-501

13. Casey BM, Thom EA, Peaceman AM, Varner MW, Sorokin Y, Hirtz DG, Reddy UM, Wapner RJ, Thorp Jr JM, Saade G. Treatment of Subclinical Hypothyroidism or Hypothyroxinemia in Pregnancy. New England Journal of Medicine 2017; 376:815-825

14. Hales C, Taylor P, Channon S, Paradice R, McEwan K, Zhang L, Gyedu M, Bakhsh A, Okosieme O, I IM, Draman M, Gregory J, Dayan C, Lazarus J, Rees D, Ludgate M. Controlled Antenatal Thyroid Screening II: effect of treating maternal sub-optimal thyroid function on child cognition The Journal of Clinical Endocrinology & Metabolism 2018; 103:1-9

15. Modesto T, Tiemeier H, Peeters RP, Jaddoe VV, Hofman A, Verhulst FC, Ghassabian A. Maternal mild thyroid hormone insufficiency in early pregnancy and attention-deficit/hyperactivity disorder symptoms in children. JAMA Pediatrics 2015; 169:838-845

16. Päkkilä F, Männistö T, Pouta A, Hartikainen AL, Ruokonen A, Surcel HM, Bloigu A, Vääräsmäki M, Järvelin MR, Moilanen I, Suvanto E. The impact of gestational thyroid hormone concentrations on ADHD symptoms of the child. Journal of Clinical Endocrinology and Metabolism 2014; 99:E1-E8

17. Ghassabian A, Bongers-Schokking JJ, Henrichs J, Jaddoe VWV, Visser TJ, Visser W, de Muinck Keizer-Schrama SMPF, Hooijkaas H, Steegers EAP, Hofman A, Verhulst FC, van der Ende J, de Rijke YB, Tiemeier H. Maternal thyroid function during pregnancy and behavioral problems in the offspring: the generation R study. Pediatric Research 2011; 69:454-459

18. Vermiglio F, Lo Presti VP, Moleti M, Sidoti M, Tortorella G, Scaffidi G, Castagna MG, Mattina F, Violi MA, Crisa A. Attention deficit and hyperactivity disorders in the offspring of mothers exposed to mild-moderate iodine deficiency: a possible novel iodine deficiency disorder in developed countries. The Journal of Clinical Endocrinology & Metabolism 2004; 89:6054-6060

19. Oostenbroek, W. MH, Kersten RHJ, Tros B, Kunst AE, Vrijkotte TGM, Finken MJJ. Maternal hypothyroxinaemia in early pregnancy and problem behavior in 5 year-old offspring. PSYCHONEUROENDOCRINOLOGY 2017; 81:29-35

20. Andersen SL, Laurberg P, Wu CS, Olsen J. Attention deficit hyperactivity disorder and autism spectrum disorder in children born to mothers with thyroid dysfunction: A Danish nationwide cohort study. BJOG: An International Journal of Obstetrics and Gynaecology 2014; 121:1365-1374

21. Román GC, Ghassabian A, Bongers-Schokking JJ, Jaddoe VWV, Hofman A, Rijke YB, Verhulst FC, Tiemeier H. Association of gestational maternal hypothyroxinemia and increased autism risk. Annals of neurology 2013; 74:733-742

22. Ghassabian A, Bongers-Schokking JJ, De Rijke YB, Van Mil N, Jaddoe VWV, De Muinck Keizer-Schrama SMPF, Hooijkaas H, Hofman A, Visser W, Roman GC, Visser TJ, Verhulst FC, Tiemeier H. Maternal thyroid autoimmunity during pregnancy and the risk of attention deficit/hyperactivity problems in children: The generation r study. Thyroid 2012; 22:178-186

23. Brown AS, Surcel HM, Hinkka-Yli-Salomäki S, Cheslack-Postava K, Bao Y, Sourander A. Maternal thyroid autoantibody and elevated risk of autism in a national birth cohort. Progress in Neuro-Psychopharmacology and Biological Psychiatry 2015; 57:86-92

24. Croen LA, Grether JK, Yoshida CK, Odouli R, Van de Water J. Maternal autoimmune diseases, asthma and allergies, and childhood autism spectrum disorders: a case-control study. Archives of pediatrics & adolescent medicine 2005; 159:151-157

25. Yau VM, Lutsky M, Yoshida CK, Lasley B, Kharrazi M, Windham G, Gee N, Croen LA. Prenatal and Neonatal Thyroid Stimulating Hormone Levels and Autism Spectrum Disorders. Journal of Autism and Developmental Disorders 2014;

26. Korevaar T, Muetzel R, Medici M, Chaker L, Jaddoe VWV, de Rijke YB, Steegers EAP, Visser TJ, White T, Tiemeier H. Association of maternal thyroid function during early pregnancy with offspring IQ and brain morphology in childhood: a population-based prospective cohort study. The Lancet Diabetes & Endocrinology 2016; 4:35-43

27. Spyridoula Maraka RM, Rozalina G McCoy , Xiaoxi Yao ,, Lindsey R Sangaralingham NMS, Derek T O'Keeffe , Ana, E Espinosa De Ycaza RR-G, Charles C Coddington III, Marius N Stan,, Juan P Brito VMM. Thyroid hormone treatment among pregnant women with subclinical hypothyroidism: US national assessment. British Medical Journal 2017; 356

28. Hales C, Channon S, Taylor PN, Draman MS, Muller I, Lazarus J, Paradice R, Rees A, Shillabeer D, Gregory JW. The second wave of the Controlled Antenatal Thyroid Screening (CATS II) study: the cognitive assessment protocol. BMC Endocrine Disorders 2014; 14:95

29. Goodman R. The Strengths and Difficulties Questionnaire: a research note. Journal of child psychology and psychiatry 1997; 38:581-586

30. Thapar A, Harrington R, Ross K, McGuffin P. Does the definition of ADHD affect heritability? Journal of the American Academy of Child & Adolescent Psychiatry 2000; 39:1528-1536

31. Rutter M, Bailey A, Lord C. The social communication questionnaire: Manual. Los Angeles, USA: Western Psychological Services.

32. Goodman R. Psychometric properties of the strengths and difficulties questionnaire. Journal of the American Academy of Child & Adolescent Psychiatry 2001; 40:1337-1345

33. Muris P, Meesters C, van den Berg F. The strengths and difficulties questionnaire (SDQ). European child & adolescent psychiatry 2003; 12:1-8

34. Palmieri PA, Smith GC. Examining the structural validity of the Strengths and Difficulties Questionnaire (SDQ) in a US sample of custodial grandmothers. Psychological assessment 2007; 19:189

35. Woerner W, Fleitlich-Bilyk B, Martinussen R, Fletcher J, Cucchiaro G, Dalgalarrondo P, Lui M, Tannock R. The Strengths and Difficulties Questionnaire overseas: evaluations and applications of the SDQ beyond Europe. European Child & Adolescent Psychiatry 2004; 13:ii47-ii54

36. Wei T, Chesnut SR, Barnard-Brak L, Richman D. Psychometric Analysis of the Social Communication Questionnaire Using an Item-Response Theory Framework: Implications for the Use of the Lifetime and Current Forms. Journal of Psychopathology and Behavioral Assessment 2014; 37:469-480

37. Wilkinson L. A best practice guide to assessment and intervention for autism and Asperger syndrome in schools. London: Jessica Kingsley Publishers.

38. Eaves LC, Wingert HD, Ho HH, Mickelson ECR. Screening for autism spectrum disorders with the social communication questionnaire. Journal of Developmental & Behavioral Pediatrics 2006; 27:S95-S103

39. Skogli EW, Teicher MH, Andersen PN, Hovik KT, Øie M. ADHD in girls and boys - gender differences in co-existing symptoms and executive function measures. BMC Psychiatry 2013; 13

40. Halladay AK, Bishop S, Constantino JN, Daniels AM, Koenig K, Palmer K, Messinger D, Pelphrey K, Sanders SJ, Singer AT, Taylor JL, Szatmari P. Sex and gender differences in autism spectrum disorder: Summarizing evidence gaps and identifying emerging areas of priority. Molecular Autism 2015; 6

41. Zwirs B, Burger H, Schulpen T, Vermulst AA, HiraSing RA, Buitelaar J. Teacher ratings of children's behavior problems and functional impairment across gender and ethnicity: Construct equivalence of the strengths and difficulties questionnaire. Journal of Cross-Cultural Psychology 2011; 42:466-481

42. Orlebeke JF, Knol DL, Boomsma DI, Verhulst FC. Frequency of parental report of problem behavior in children decreases with increasing maternal age at delivery. Psychological Reports 1998; 82:395-404

43. Durkin MS, Maenner MJ, Newschaffer CJ, Lee L, Cunniff CM, Daniels JL, Kirby RS, Leavitt L, Miller L, Zahorodny W. Advanced parental age and the risk of autism spectrum disorder. American Journal of Epidemiology 2008; 168:1268-1276

44. Heikkilä K, Sacker A, Kelly Y, Renfrew MJ, Quigley MA. Breast feeding and child behaviour in the millennium cohort study. Archives of Disease in Childhood 2011; 96:635-642

45. Stadler DD, Musser ED, Holton KF, Shannon J, Nigg JT. Recalled Initiation and Duration of Maternal Breastfeeding Among Children with and Without ADHD in a Well Characterized Case–Control Sample. Journal of Abnormal Child Psychology 2015; 8:1-9

46. Al-Farsi YM, Al-Sharbati MM, Waly MI, Al-Farsi OA, Al-Shafaee MA, Al-Khaduri MM, Trivedi MS, Deth RC. Effect of suboptimal breast-feeding on occurrence of autism: A case-control study. Nutrition 2012; 28:e27-e32

47. Piotrowska PJ, Stride CB, Croft SE, Rowe R. Socioeconomic status and antisocial behaviour among children and adolescents: A systematic review and meta-analysis. Clinical Psychology Review 2015; 35:47-55

48. Bennett DA. How can I deal with missing data in my study? Australian and New Zealand Journal of Public Health 2001; 25:464-469

49. Downey RG, King CV. Missing data in Likert ratings: A comparison of replacement methods. *The Journal of general psychology* 1998; 125:175-191

50. Roth PL. Missing data: A conceptual review for applied psychologists. *Personnel psychology* 1994; 47:537-560

51. Peyre H, Leplège A, Coste J. Missing data methods for dealing with missing items in quality of life questionnaires. A comparison by simulation of personal mean score, full information maximum likelihood, multiple imputation, and hot deck techniques applied to the SF-36 in the French 2003 decennial health survey. Quality of Life Research 2011; 20:287-300

52. Chandler S, Charman T, Baird G, Simonoff E, Loucas T, Meldrum D, Scott M, Pickles A. Validation of the social communication questionnaire in a population cohort of children with autism spectrum disorders. Journal of the American Academy of Child & Adolescent Psychiatry 2007; 46:1324-1332

53. Ludgate M. Supplemental data. 2019; supplemental figures and tables. Available at: https://figshare.com/s/e2a9d93e558716d58733

54. Stagnaro-Green A, Abalovich M, Alexander E, Azizi F, Mestman J, Negro R, Nixon A, Pearce EN, Soldin OP, Sullivan S, W W. Guidelines of the American Thyroid Association for the diagnosis and management of thyroid disease during pregnancy and postpartum. Thyroid 2011; 21:1081-1125

55. Alexander EK, Pearce EN, Brent GA, Brown RS, Chen H, Dosiou C, Grobman W, Laurberg P, LazarusJH, Mandel SJ, Peeters R, S. S. Guidelines of the American Thyroid Association for the

Diagnosis and Management of Thyroid Disease during Pregnancy and the Postpartum Thyroid 2017; 27:315-389

56. Hauser P, Soler R, Brucker-Davis F, Weintraub BD. Thyroid hormones correlate with symptoms of hyperactivity but not inattention in attention deficit hyperactivity disorder. Psychoneuroendocrinology 1997; 22:107-114

57. 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

58. Korevaar TI, Medici M, de Rijke YB, Visser W, de Muinck Keizer-Schrama SM, Jaddoe VW, Hofman A, Ross HA, Visser WE, Hooijkaas H. Ethnic differences in maternal thyroid parameters during pregnancy: the Generation R study. The Journal of Clinical Endocrinology & Metabolism 2013; 98:3678-3686

59. De Groot L, Abalovich M, Alexander EK, Amino N, Barbour L, Cobin RH, Eastman CJ, Lazarus JH, Luton D, Mandel SJ. Management of thyroid dysfunction during pregnancy and postpartum: an Endocrine Society clinical practice guideline. The Journal of Clinical Endocrinology & Metabolism 2012; 97:2543-2565

60. National, Institute, Health, Care, Excellence. Hypothyroidism; Scenario: Preconception or pregnant. 2016; http://cks.nice.org.uk/hypothyroidism#!scenario:3. Accessed 12.07, 2016.

LEGENDS FOR FIGURES AND TABLES

Figure 1: Flow diagram of the CATS studies showing UK participants only.

Figure 2: Correlations between elements of child behaviour and FT4 levels at recruitment and following several weeks of thyroxine treatment.

Table 1: Characteristics of the cohort.

Table 2: Group total scores for the questionnaires

Table 3: High questionnaire scores from the normal gestational thyroid function (GTF), treated, and untreated suboptimal GTF groups

SUPPLEMENTARY MATERIALS

Characteristics of the follow-on cohort

The SGTF groups were compared between first participation of offspring, and at follow-up (current). The sample appeared mostly similar between the two time points (table 1). Those in the treated SGTF group who continued participation into the follow-up of CATS had significantly raised TSH values. A potential reason for this could have been that those who felt most benefit from treatment, continued to aid the study. The mothers in the untreated SGTF group who participated in the follow-up were significantly older at time of pregnancy consent compared to the original CATS sample, this was by an age increase of 1.1 years.

Table 1 | Comparison of study population to original cohort. Figures are mean values (standard deviation). P for difference by t-test unless stated otherwise.

Characteristics at time of study entry	Treated SGTF		P for difference	Untreated SGTF		P for difference
	CATS N=303	Follow-up N= 125		CATS N=306	Follow-up N=104	
Male children (%)	54.1	52.0	0.921*	51.3	48.1	0.569
Social deprivation**	3 IQR=2-5 (mean=3.3)	4 IQR=3-5 (mean=3.8)	0.358*	3 IQR=2-5 (mean=3.3)	4 IQR=2-5 (mean=3.4)	0.981
Mother age	29.8 (5.6)	30.2 (5.1)	0.447	29.8 (5.4)	30.9 (4.8)	0.050 (95% CI -0.002, 2.364)
Maternal TSH	3.6 IQR=1.5-4.7	4.1 IQR=1.9-5.1	0.026†	3.4 IQR=1.2-4.1	3.6 IQR=1.1-4.5	0.372†
Maternal FT4	11.8 (1.8)	12.0 (1.9)	0.262	12.0 (1.9)	11.7 (1.9)	0.310

*Significance tested by chi square.
**The social deprivation scores used were calculated from postcodes at time of pregnancy, rather than the participants' current deprivation scores as used in the main analysis.
†Significance for TSH at consent test by Mann Whitney.
SGTF=suboptimal gestational thyroid function, CATS=controlled antenatal thyroid screening, TSH=thyroid stimulating hormone, FT4= free thyroxine

Thyroid function tests during pregnancy of the treated suboptimal gestational thyroid function group

Thyroid function tests in pregnancy in the treated SGTF group were analysed by repeated measures ANOVAs (FT4) and Friedman's test (TSH).

Mean (SD) FT4 levels at time of consent, at 20 weeks' gestation, and 30 weeks' gestation were 12.0pmol/L (1..91), 16.2pmol/L (2.89), and 15.5pmol/L (2.47), respectively). Maternal FT4 was different at the multivariate level throughout pregnancy for the treated SGTF group (Pillai's Trace=0.687, F(2,116)=127.93, P=0.001, η_p^2=0.687). Pairwise comparisons that were Bonferroni corrected, identified that at time of consent FT4 was significantly lower than at 20 weeks gestation (P=0.001, -5.038 to -3..627), and 30 weeks gestation (P=0.001, -4.151 to -2.929). FT4 at 20 weeks gestation was significantly higher than at 30 weeks gestation (P=0.003, 0.231 to 1.353).

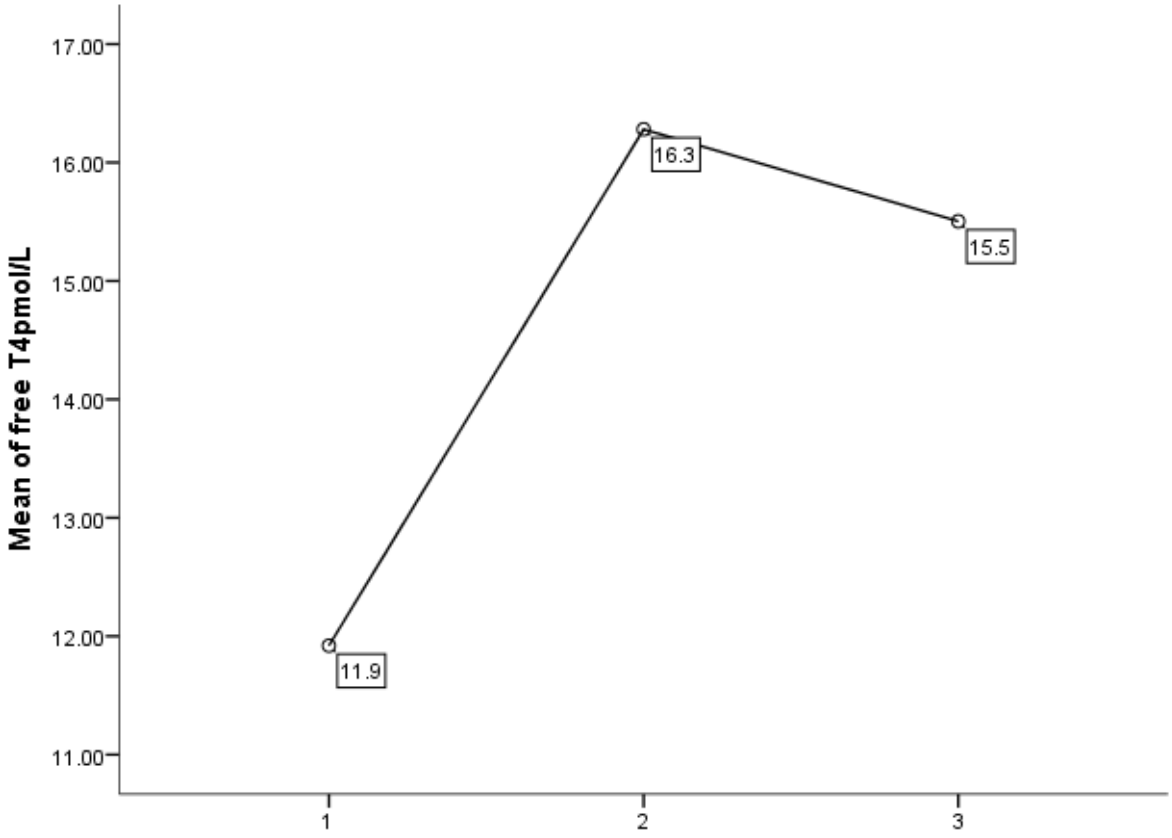


Fig 1|Mean free thyroxine (T4) values of the treated suboptimal gestational thyroid function group taken at: (1) consent into the Controlled Antenatal Thyroid Screening study (median 12 weeks 3 days gestation), (2) 20 weeks (0 days) gestation, and (3) 30 weeks (1 day) gestation. T4 was measured in pmol/L.

There was a statistically significant difference in TSH levels throughout the pregnancies of those from the treated SGTF group, $\chi^2(2)=173.207$, $P=0.001$. Post hoc analysis with Wilcoxon signed-rank tests was conducted with a Bonferroni correction applied, resulting in a significance level at P<0.017. Median (IQR) TSH levels at time of consent, 20 weeks gestation, and 30 weeks gestation were 4.1mIU/L (1.9 to 5.1), 0.3mIU/L (0.1 to 1.0), and 0.3mIU/L (0.03 to 0.9), respectively. There was a significant difference between TSH at time of consent and at 20 weeks gestation (Z=-9.543, P=0.001), as well as at 30 weeks gestation (Z=-9.478, P=0.001). There was also a significant difference between TSH at 20 weeks gestation and 30 weeks gestation (Z=-3.109, P=0.002), even though there was only a slight drop in the median value.

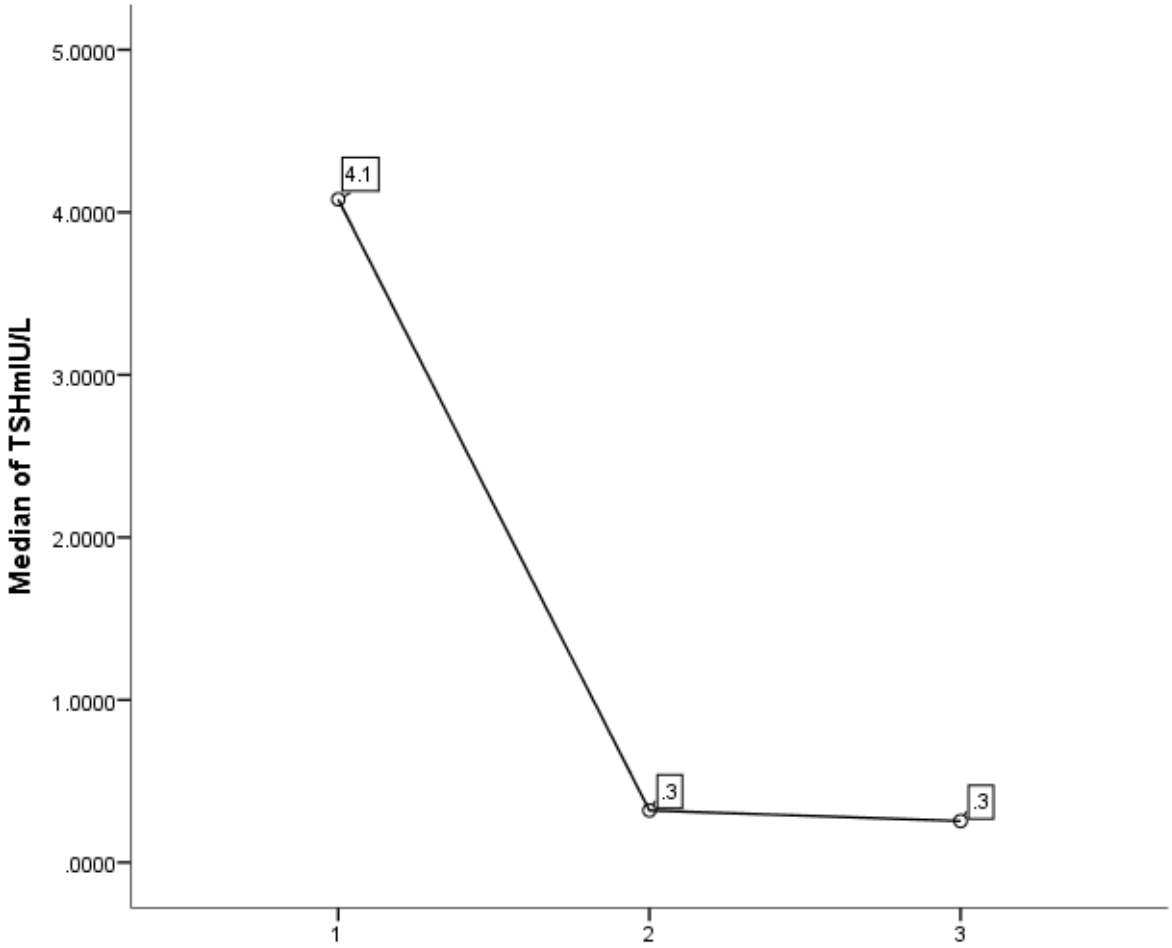


Figure 2 |Median thyroid stimulating hormone (TSH) values of the treated suboptimal gestational thyroid function group taken at: (1) consent into the Controlled Antenatal Thyroid Screening study (median 12 weeks 3 days gestation), (2) 20 weeks (0 days) gestation, and (3) 30 weeks (1 day) gestation. TSH was measured in mIU/L.

Supplementary Statistics

All analyses in this section deal with continuous data; therefore are MANCOVAs and include Bonferroni corrections where applicable. The reported statistical test used in the MANCOVAs was Pillai's trace, V. Every analysis was adjusted for the same covariates as in the main article (child gender, mother age at time of consent, whether the mother breastfed for over one month, and social deprivation). Analyses followed the same pattern, three CATS groups compared for the three total questionnaire scores, then a secondary MANCOVA exploring all questionnaire domains with the treated SGTF split into optimal and over-treatment. Exploration of specific maternal FT4 ranges only included the latter MANCOVAs.

Subclinical hypothyroidism

Women identified as having subclinical hypothyroidism during their pregnancies, had a TSH >97.5th and FT4 >2.5th percentiles at time of consent. In the treated SGTF group, 57.26% (n=67) were in this category, and 49.00% (n=49) of the untreated SGTF group. There was no difference between the normal GTF, treated, and untreated subclinical hypothyroid groups for total questionnaire scores (P=0.909). When exploring the effect of treatment to the 17.7pmol/L threshold, there was also no difference in mean scores (P=0.128). See table 2 for descriptive data.

Table 2 | Mean scores for the subclinical hypothyroid and normal GTF groups. Figures are mean values (standard deviation).

Questionnaire domain		Subclinical hypothyroidism			
	Normal GTF N=237	Treated N=67	Untreated N=49	Optimally treated N=41	Over-treated N=26
Strengths and difficulties questionnaire					
Emotion	2.32 (2.24)		2.32 (2.18)	2.22 (1.98)	2.50 (1.73)
Conduct	1.21 (1.50)		1.33 (1.39)	1.44 (1.83)	1.92 (1.83)
Hyperactivity	3.09 (2.52)		3.47 (2.61)	3.10 (2.91)	4.23 (2.55)
Peer problems	1.58 (1.88)		1.10 (1.29)	1.51 (1.90)	1.00 (1.02)
Total difficulties	8.20 (6.01)	8.81 (5.71)	8.22 (5.41)	8.27 (6.13)	9.65 (4.97)
Prosocial	8.78 (1.78)		8.47 (2.09)	8.90 (1.71)	8.69 (1.93)
Child ADHD questionnaire					
Inattention	6.07 (5.67)		5.75 (4.94)	5.88 (5.84)	7.62 (5.73)
Overactivity	2.05 (2.29)		2.37 (2.42)	2.41 (2.88)	3.58 (3.14)
Impulsivity	3.13 (2.79)		3.34 (2.50)	3.68 (3.55)	4.38 (3.28)
Total	11.25 (9.53)	13.37 (11.20)	11.46 (8.52)	11.98 (11.45)	15.58 (10.63)
Social communication questionnaire	4.28 (3.70)	4.59 (4.14)	3.95 (2.96)	4.05 (3.73)	5.43 (4.65)

SGTF=suboptimal gestational thyroid function, ADHD=attention deficit hyperactivity disorder

Maternal hypothyroxinemia

Women were classified as the having hypothyroxinemia during their pregnancies if their FT4 at consent was <2.5th and their TSH was <97.5th percentiles. There were 30.77% (n=36) of the treated SGTF, and 44.00% (n=44) of the untreated SGTF group in this category. Total questionnaire scores were not significantly different between the normal GTF, treated and untreated hypothyroxinaemia groups P=0.207, see table 3 for scores). There was also no difference when treatment was explored between the groups (P=0.105).

Table 3 | Mean questionnaire scores for the hypothyroxinemic and normal GTF groups. Figures are mean values (standard deviation).

Questionnaire domain		Hypothyroxinemic			
	Normal GTF N=237	Treated N=36	Untreated N=44	Optimally treated N=26	Over-treated N=10
Strengths and difficulties questionnaire					
Emotion	2.32 (2.24)		2.11 (1.94)	2.41 (2.10)	2.90 (2.33)
Conduct	1.21 (1.50)		1.05 (1.40)	1.50 (1.73)	1.70 (1.83)
Hyperactivity	3.09 (2.52)		3.30 (2.79)	3.69 (2.81)	4.30 (4.06)
Peer problems	1.58 (1.88)		1.20 (1.32)	1.31 (1.87)	1.00 (1.33)
Total difficulties	8.20 (6.01)	9.18 (7.17)	7.66 (5.40)	8.91 (6.83)	9.90 (8.35)
Prosocial	8.78 (1.78)		9.02 (1.25)	8.73 (1.93)	9.40 (1.07)
Child ADHD questionnaire					
Inattention	6.07 (5.67)		4.79 (4.63)	6.75 (6.32)	6.90 (7.84)
Overactivity	2.05 (2.29)		2.38 (2.82)	2.31 (2.62)	3.40 (4.14)
Impulsivity	3.13 (2.79)		2.93 (3.04)	3.42 (2.87)	4.10 (4.04)
Total	11.25 (9.53)	13.02 (12.29)	10.10 (9.25)	12.48 (10.99)	14.40 (15.78)
Social communication questionnaire	4.28 (3.70)	5.98 (5.21)	4.26 (2.89)	6.47 (5.44)	4.70 (4.55)

SGTF=suboptimal gestational thyroid function, ADHD=attention deficit hyperactivity disorder

Overt hypothyroid groups

As the CATS study classified women with FT4 <2.5th and/or TSH >97.5th percentiles for a positive result of SGTF, the secondary analysis examined questionnaire scores by women who had overt hypothyroidism (OH, FT4 <2.5th and TSH >97.5th). OH, though uncommon during pregnancy (3 in 1000(1)), is the most severe underactive thyroid function. There were 8.55% (n=10) of the treated SGTF group, and 9.00% (n=9) of the untreated SGTF group classified as having OH. Comparing these groups with the normal GTF for total

questionnaire scores identified a significant difference at the multivariate level (V=0.063, F(6,496)=2.669, P=0.015, η_p^2=0.031), with a difference found for ADHD Total scores (F(2,249)=4.680, P=0.010, η_p^2=0.036). The untreated OH group had significantly higher scores compared to the normal GTF for ADHD Total (P=0.022 (0.945, 16.876)).

When exploring the effect of treatment to the 17.7pmol/L threshold for maternal FT4, there was a sustained difference at the multivariate level (V=0.203, F(27,726)=1.948, P=0.003 η_p^2=0.068). At the univariate level SDQ Conduct (F(3,248)=4.402, P=0.005, η_p^2=0.051), ADHD Hyperactivity (F(3,248)=5.665, P=0.001, η_p^2=0.064), and ADHD Total (F(3,248)=3.487, P=0.016, η_p^2=0.040) were significantly different. The over-treated OH group had significantly higher scores for SDQ Conduct (P=0.038 (0.072, 4.273) and ADHD Hyperactivity (P=0.020 (0.357, 6.597) compared to the normal GTF. Again, for ADHD Total, the untreated OH had significantly higher scores than the normal GTF (P=0.045 (0.117, 7.693).

Children born to mothers who were untreated for OH in pregnancy, had higher scores for the Child ADHD Questionnaire, optimal treatment improved this with some means moving closer to those of the normal GTF (table 4). Over-treatment for OH had a significant effect here on domains found with the entire SGTF sample; SDQ Conduct and ADHD Hyperactivity. Caution is advised as the sample sizes are small for OH.

Table 4 | Questionnaire scores for the normal gestational thyroid function (GTF) group, and treated and untreated overt hypothyroid group. Figures are mean values (standard deviation).

Questionnaire domain	Normal GTF	Treated overt hypothyroid	Untreated overt hypothyroid	Optimally treated	Over-treated
	N=237	N=10	N=9	N=6	N=4
Strengths and difficulties questionnaire					
Emotion	2.32 (2.24)		2.78 (1.64)	2.67 (2.94)	3.25 (3.30)
Conduct	1.21 (1.50)		2.44 (2.07)	0.83 (0.98)	3.25* (2.99)
Hyperactivity	3.09 (2.52)		5.11 (3.37)	4.33 (2.16)	4.25 (4.35)
Peer problems	1.58 (1.88)		1.33 (1.22)	1.33 (1.97)	0.50 (1.00)
Total difficulties	8.20 (6.01)	10.00 (8.51)	11.67 (7.16)	9.17 (7.08)	11.25 (11.41)
Prosocial	8.78 (1.78)		8.44 (1.33)	9.17 (1.33)	8.50 (3.00)
Inattention	6.07 (5.67)		11.44 (6.77)	7.32 (5.44)	9.25 (9.98)
Hyperactivity	2.05 (2.29)		4.11 (3.48)	4.00 (2.61)	5.25* (4.19)
Impulsivity	3.13 (2.79)		5.22 (3.31)	2.17 (1.83)	5.25 (4.50)
Total	11.25 (9.53)	15.99 (12.67)	20.78* (12.10)	13.48 (8.29)	19.75 (18.32)
Social communication questionnaire	4.28 (3.70)	6.70 (8.08)	4.00 (2.78)	6.17 (5.85)	5.72 (8.21)

*Significantly higher (P<0.05) group results (from columns left to right)
SGTF=suboptimal gestational thyroid function, ADHD=attention deficit hyperactivity disorder
Untreated overt hypothyroid (1) Child ADHD questionnaire (ADHD) Inattention, higher than Normal GTF, (2) ADHD Total, higher than Normal GTF, (3) Strengths and difficulties questionnaire Conduct, higher than Normal GTF.
Treated overt hypothyroid ADHD Hyperactivity and SDQ Conduct, higher than Normal GTF. Untreated overt hypothyroid ADHD Total, higher than Normal GTF.

Exploratory analysis; treatment by target range

The following analyses explored potential limits of treatment for minimising specific neurodevelopmental symptoms in the child. Analyses were again MANCOVAs with the same four covariate adjustments.

As can be seen in table 5, the means of the treated group decreased with decreasing FT4 threshold values. This confirms the main conclusions from the article that over-treatment could have an impact for the child's neurodevelopment; as well as the threshold of 17.7pmol/L in pregnancy being the most 'detrimental'. As thresholds decreased, interestingly, those who were treated up to the thresholds increased questionnaire scores. This supports a theory of 'U-curved effects'(2); demonstrating that higher and lower values of maternal FT4 can affect the child. A further possible mechanism could be that those born to mothers with the lowest FT4 had more of an effect to the group mean with decreasing groups with lowering FT4 thresholds. SDQ Peer problems being significantly higher in the normal GTF group were repeatedly found compared to above thresholds for maternal FT4. At the whole group level of the treated SGTF, TSH was positively correlated to SDQ Peer problems at 30 weeks' gestation. As the FT4 levels were high, TSH would have been lowering, and supporting the correlation, the lower the TSH level, the lower SDQ Peer problem scores will have been.

Over 17pmol/L in pregnancy

For the first threshold of 17pmol/L (n=53), there was a significant difference between the questionnaire scores at the multivariate level (V=0.114, F(27,1320)=1.922, P=0.003, η_p^2=0.038). Univariate analyses identified difference for SDQ Peer problems (F(3,446)=3.942, P=0.009, η_p^2=0.026) and ADHD Overactivity (F(3,446)=4.229, P=0.006, η_p^2=0.028). Bonferroni corrected pairwise comparisons evidenced that the normal GTF had a higher group mean than those born to mothers with FT4 values >17pmol/L (P=0.040, 0.019 to 1..363), but the effect was inversed for ADHD Overactivity (P=0.003, 0.324 to 2.340).

Over 16pmol/L in pregnancy

The lower threshold 16pmol/L (n=70) also yielded a difference at the multivariate level for questionnaire scores (V=0.127, F(27,1320)=2.162, P=0.001, η_p^2=0.042). The domains with significant differences were, SDQ Peer problems (F(3,446)=4.755, P=0.003, η_p^2=0.031) and ADHD Overactivity (F(3,446)=3.851, P=0.010, η_p^2=0.025). Significance persisted following Bonferroni corrections, and akin to FT4 values >17pmol/L throughout pregnancy, SDQ Peer problems were higher in the normal GTF compared to those mothers with FT4 values >16pmol/L SGTF (P=0.015, 0.088 to 1.289), with those >16pmol/L having higher scores compared to the normal GTF group for ADHD Overactivity (P=0.006, 0.226 to 2.034).

Over 15pmol/L in pregnancy

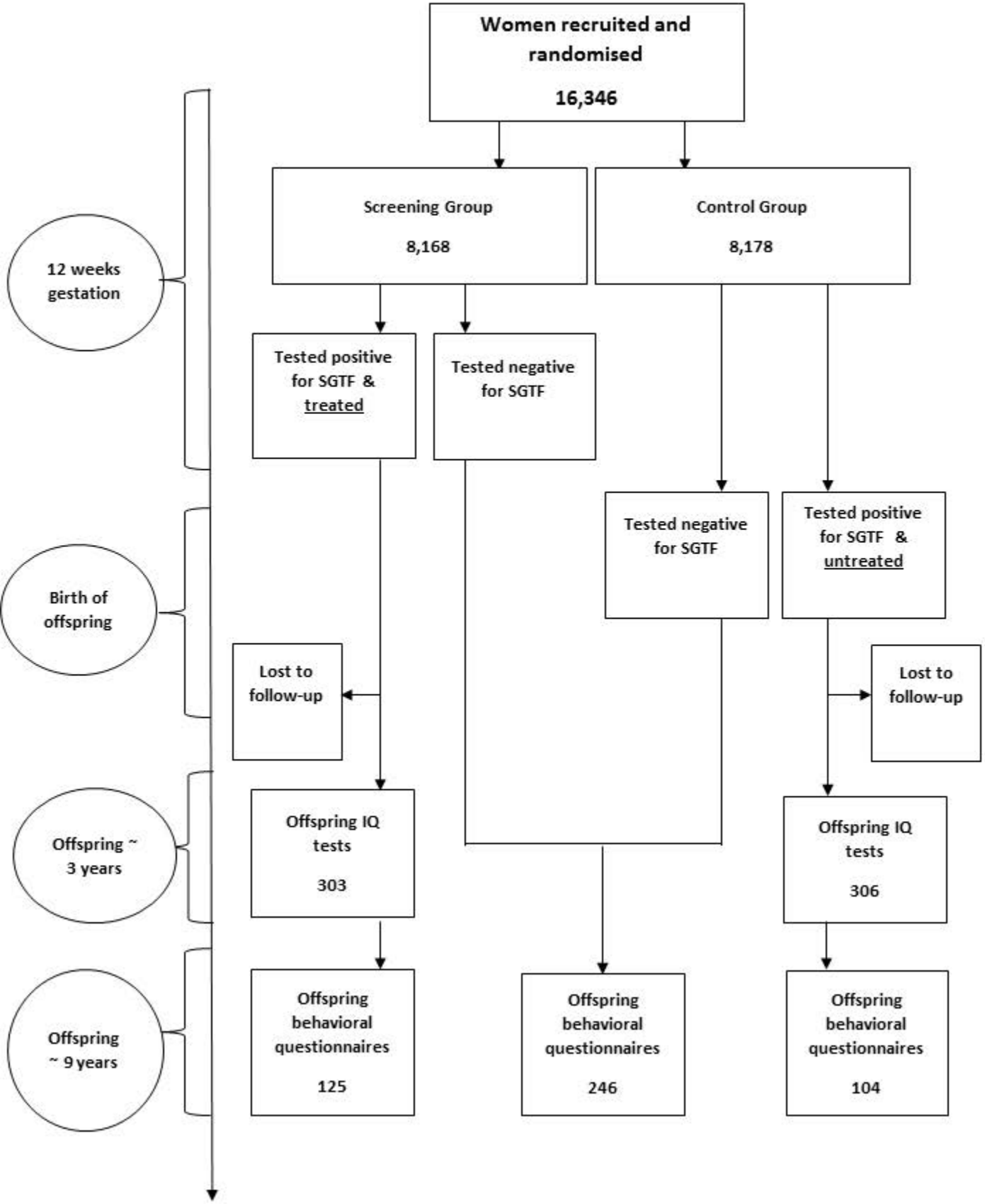
The lowered threshold of 15pmol/L (n=87) was used for the following analysis. The MANCOVA identified a difference between the groups (V=0.115, F(27,1320)=1.941, P=0.003, η_p^2=0.038). Univariate and pairwise comparison analyses identified differences for SDQ Peer problems (F(3,446)=5.871, P=0.001, η_p^2=0.038, normal GTF higher than those with maternal FT4 values >15pmol/L SGTF, P=0.009, 0.113 to 1.220), and SCQ (F(3,446)=2.872, P=0.036, η_p^2=0.019, children born to mothers with FT4 <14.9pmol/L scored higher than the normal GTF P=0.044, 0.032 to 4.211).

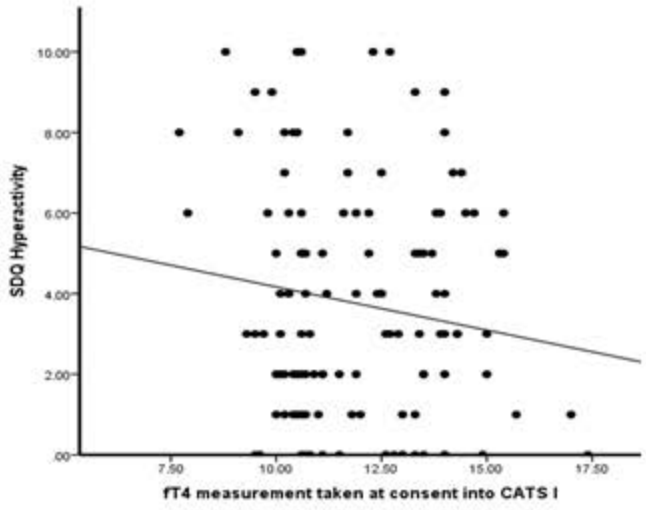
Table 5| Mean questionnaire scores for thresholds of FT4 in pregnancy. Figures are mean values (standard deviation).

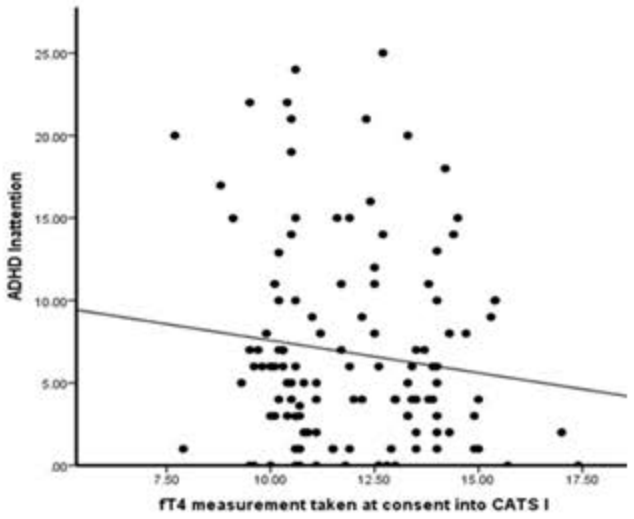
Questionnaire domain			Treatment thresholds					
	Normal GTF N=237	Untreated SGTF N=100	>17pmol/L N=53	<16.9pmol/L N=64	>16pmol/L N=70	<15.9pmol/L N=47	>15pmol/L N=87	<14.9pmol/L N=30
Strengths and difficulties questionnaire								
Emotion	2.32 (2.24)	2.24 (2.01)	2.58 (1.96)	2.28 (2.13)	2.46 (2.08)	2.35 (2.02)	2.34 (2.02)	2.62 (2.16)
Conduct	1.21 (1.50)	1.27 (1.42)	1.60 (1.84)	1.59 (1.81)	1.43 (1.73)	1.84 (1.93)	1.44 (1.70)	2.03 (2.09)
Hyperactivity	3.09 (2.52)	3.51 (2.74)	3.91 (3.03)	3.48 (2.80)	3.73 (3.06)	3.60 (2.67)	3.61 (3.00)	3.87 (2.62)
Peer problems	1.58 (1.88)	1.16 (1.28)	0.92 (1.03)	1.48 (1.96)	0.91 (1.21)	1.70 (2.02)	0.93 (1.33)	2.10 (2.07)
Total difficulties	8.20 (6.01)	8.18 (5.42)	9.02 (6.15)	8.83 (6.65)	8.53 (6.28)	9.49 (6.61)	8.33 (6.25)	10.62 (6.64)
Prosocial	8.78 (1.78)	8.73 (1.72)	8.89 (1.70)	8.72 (1.81)	9.09 (1.55)	8.47 (2.00)	9.05 (1.52)	8.23 (2.25)
Child ADHD questionnaire								
Inattention	6.07 (5.67)	5.83 (5.25)	7.21 (6.51)	6.24 (5.76)	6.79 (6.44)	6.52 (5.62)	6.34 (6.19)	7.65 (5.83)
Overactivity	2.05 (2.29)	2.53 (2.73)	3.34 (3.45)	2.41 (2.62)	3.17 (3.41)	2.32 (2.35)	2.79 (3.25)	2.93 (2.42)
Impulsivity	3.13 (2.79)	3.32 (2.85)	3.92 (3.67)	3.55 (2.98)	3.91 (3.56)	3.43 (2.89)	3.69 (3.39)	3.80 (3.08)
Total	11.25 (9.53)	11.67 (9.50)	14.47 (12.78)	12.20 (10.30)	13.87 (12.66)	12.27 (9.56)	12.83 (12.03)	14.38 (9.89)
Social communication questionnaire	4.28 (3.70)	4.09 (2.89)	4.84 (5.06)	5.41 (4.81)	4.72 (4.75)	5.80 (5.13)	4.66 (4.82)	6.57 (4.99)

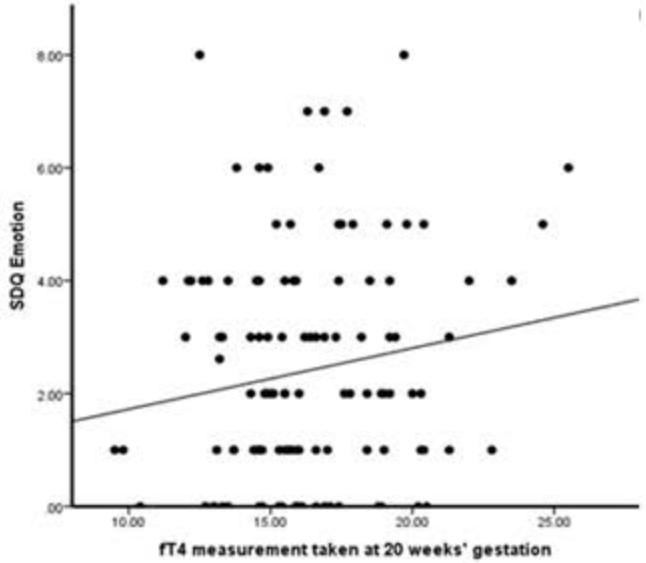
SGTF=suboptimal gestational thyroid function, ADHD=attention deficit hyperactivity disorder

1. Tunbridge W, Evered D, Hall R, Appleton D, Brewis M, Clark F, et al. The spectrum of thyroid disease in a community: the Whickham survey. Clin Endocrinol. 1977;7(6):481-93.
2. Korevaar T, Muetzel R, Medici M, Chaker L, Jaddoe VWV, de Rijke YB, et al. Association of maternal thyroid function during early pregnancy with offspring IQ and brain morphology in childhood: a population-based prospective cohort study. The Lancet Diabetes & Endocrinology. 2016;4(1):35-43.


Women recruited and randomised
16,346
Screening Group
8,168
Control Group
8,178
12 weeks gestation
Tested positive for SGTF & treated
Tested negative for SGTF
Tested negative for SGTF
Tested positive for SGTF & untreated
Birth of offspring
Lost to follow-up
Lost to follow-up
Offspring ~ 3 years
Offspring IQ tests
303
Offspring IQ tests
306
Offspring ~ 9 years
Offspring behavioral questionnaires
125
Offspring behavioral questionnaires
246
Offspring behavioral questionnaires
104


SDQ Hyperactivity
10.00
8.00
6.00
4.00
2.00
.00
7.50
10.00
12.50
15.00
17.50
fT4 measurement taken at consent into CATS I


ADHD Inattention
25.00
20.00
15.00
10.00
5.00
.00
7.50
10.00
12.50
15.00
17.50
fT4 measurement taken at consent into CATS I


SDQ Emotion
8.00
6.00
4.00
2.00
.00
10.00
15.00
20.00
25.00
fT4 measurement taken at 20 weeks' gestation

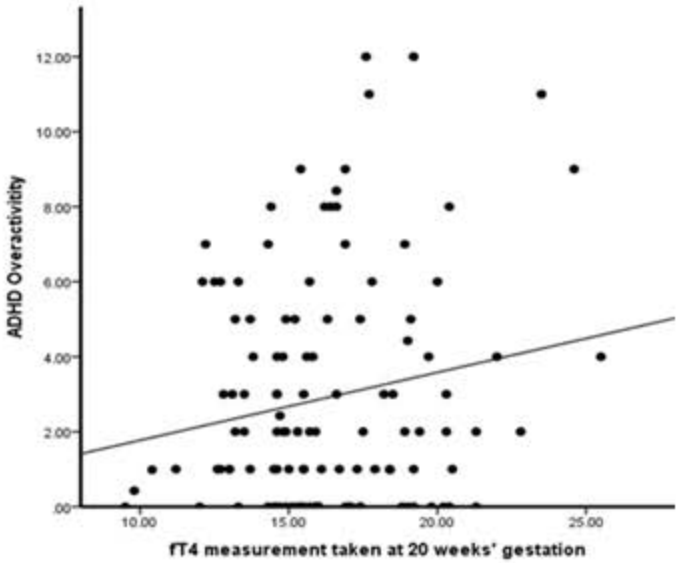

ADHD Overactivitity
12.00
10.00
8.00
6.00
4.00
2.00
.00
10.00
15.00
20.00
25.00
fT4 measurement taken at 20 weeks' gestation

Table 1 | Characteristics of the cohort. Values are means (standard deviations) unless stated otherwise.

Characteristics	Normal-GTF	Treated SGTF	Untreated SGTF	Normal-GTF vs. Treated SGTF (P)	Normal-GTF vs. Untreated SGTF (P)	Treated SGTF vs Untreated SGTF (P)
	N=246	N=125	N=104			
TSH at study entry (mIU/L)†	1..2 (0.7-1.8)	4.1 (1.9-5.1)	3..6 (1.1-4..5)	0.001	0.001	0.003
FT4 at study entry (pmol/L)	14.1 (1..7)	12.0 (1.9)	11.7 (1.9)	0.001 (95% CI 1.675 to 2.633)	0.001 (95% CI 1.867 to 2.887)	1.000
Age of mother at study entry (years)	31.7 (5.2)	30.2 (5.1)	30.9 (4.8)	0.025 (0.138, 2.824)	0.588	0.879
Social deprivation† quintile (mean)	4 (3-5) (3.7)	5 (3-5) (3.8)	4 (2-5) (3.4)	0.687	0.491	0.162
Child breastfed for over one month (%)*	65.3	58.9	56.3	0.226	0.113	0.697
Male children (%)	50.0	52.0	48.1	0.716	0.742	0.554
Age of children†	9..8 (9.0-10.3)	9.6 (9.1-9.8)	9..4 (9.0-9..9)	0.003	0.017	0.996

†Median and interquartile ranges presented.
*missing per group, n=1
Confidence intervals presented for continuous data only. Current free thyroxine reference ranges for adults during pregnancy per trimester are as follows: first trimester 10.5-18.3pmol/L, second trimester 9.5-15.9pmol/L, third trimester 8.6-13.7pmol/L (serum TSH and FT4 were measured with the use of Immunochemiluminescence, ADVIA Centaur, Siemens Healthcare Diagnostics).
SGTF= suboptimal gestational thyroid function, TSH=thyroid stimulating hormone, FT4=free thyroxine

Table 2 | Group total scores for the questionnaires. Figures are mean values (standard deviation)				**p**
Questionnaire total scores	Normal-GTF	Treated SGTF	Untreated SGTF	
	N=237	N=117	N=100	
Strengths and difficulties questionnaire*	8.20 (6.01)	8.92 (6.40)	8.17 (5.42)	0.53
Child ADHD questionnaire	11.24 (9.53)	13.23 (11.49)	11.67 (9.50)	0.22
Social communication questionnaire	4.28 (3.70)	5.15 (4.91)	4.09 (2.89)	0.08
*Total difficulties score				

Table 3 | High questionnaire scores from the normal gestational thyroid function (GTF), treated, and untreated suboptimal GTF groups. Figures are numbers (percentages).

Questionnaire domain	Normal GTF	Treated SGTF	Untreated SGTF	Optimally treated	Over-treated
	N=237	N=117	N=100	N=77	N=40
Strengths and difficulties questionnaire					
Emotion					
High*/very high	39 (16.4)	19 (16.3)	13 (13.0)	11 (14.3)	8 (20.0)
Conduct					
High/very high	20 (8.5)	18 (15.4)^	7 (7.0)	9 (11.7)	9 (22.5)^^^
Hyperactivity					
High/very high	15 (6.3)	16 (13.7)^	10 (10.0)	10 (13.0)	5 (12.5)
Peer problems					
High/very high	39 (16.4)	13 (11.1)	7 (7.0)	12 (15.6)	1 (2.5)
Total difficulties					
High/very high	28 (11.8)	16 (13.7)	8 (8.0)	10 (13.0)	6 (15.0)
Prosocial					
Low/very low	27 (11.4)	13 (11.1)	10 (10.0)	8 (10.4)	5 (12.5)
Child ADHD questionnaire					
Inattention					
>2SD†	11 (4.8)	9 (7.8)	3 (3.0)	5 (6.5)	4 (10.0)
Hyperactivity					
> 2SD	9 (3.9)	12 (10.2)^	5 (5.0)	5 (6.5)	7 (17.5)^^
Impulsivity					
>2SD	6 (2.5)	7 (6.1)	3 (3.0)	3 (3.9)	4 (10.0)
Social communication questionnaire					
Total >15	4 (1.7)	7 (6.0)^	1 (1.0)	4 (5.2)	3 (7.5)^^

SGTF=suboptimal gestational thyroid function normal-GTF = normal gestational thyroid function

*high = 90-95th percentile, as calculated by the SDQ (scores >90th percentile are clinically significant); very high>95th percentile, as calculated by the SDQ

†SDs calculated by Thapar and colleagues

^p<0.05 treated versus normal-GTF

^^p<0.05 over-treated versus untreated SGTF

^^^ p< 0.01 over-treated versus untreated SGTF