

Controlled Antenatal Thyroid Screening II: Effect of Treating Maternal Suboptimal Thyroid Function on Child Cognition

Charlotte Hales,¹ Peter N. Taylor,¹ Sue Channon,² Ruth Paradice,³ Kirsten McEwan,² Lei Zhang,¹ Michael Gyedu,¹ Ameen Bakhsh,¹ Onyebuchi Okosieme,¹ Ilaria Muller,¹ Mohd S. Draman,¹ John W. Gregory,¹ Colin Dayan,¹ John H. Lazarus,¹ D. Aled Rees,⁴ and Marian Ludgate¹

¹School of Medicine, Cardiff University, Cardiff, Wales CF14 4XN, United Kingdom; ²Centre for Trials Research, Cardiff University, Cardiff, Wales CF14 4XN, United Kingdom; ³St David's Hospital, Cardiff and Vale University Health Board, Cardiff, Wales CF11 9XB, United Kingdom; and ⁴Neuroscience and Mental Health Research Institute, Cardiff University, Cardiff, Wales CF24 4HQ, United Kingdom

Context and Objective: The Controlled Antenatal Thyroid Screening (CATS) study investigated treatment of suboptimal gestational thyroid function (SGTF) on childhood cognition and found no difference in intelligence quotient (IQ) at 3 years between children of treated and untreated SGTF mothers. We have measured IQ in the same children at age 9.5 years and included children from normal gestational thyroid function (normal-GTF) mothers.

Design, Setting, and Participants: One examiner, blinded to participant group, assessed children's IQ (Wechsler Intelligence Scale for Children, Fourth Edition UK), long-term memory, and motor function (Developmental Neuropsychological Assessment II) from children of 119 treated and 98 untreated SGTF mothers plus children of 232 mothers with normal-GTF. Logistic regression explored the odds and percentages of an IQ < 85 in the groups.

Results: There was no difference in IQ < 85 between children of mothers with normal-GTF and combined SGTF, i.e., treated and untreated (fully adjusted odds ratio [OR] = 1.15 [95% confidence interval (CI) 0.52, 2.51]; $P = 0.731$). Furthermore, there was no significant effect of treatment [untreated OR = 1.33 (95% CI 0.53, 3.34); treated OR = 0.75 (95% CI 0.27, 2.06) $P = 0.576$]. IQ < 85 was 6.03% in normal-GTF, 7.56% in treated, and 11.22% in untreated groups. Analyses accounting for treated-SGTF women with free thyroxine > 97.5th percentile of the entire CATS-I cohort revealed no significant effect on a child's IQ < 85 in CATS-II. IQ at age 3 predicted IQ at age 9.5 ($P < 0.0001$) and accounted for 45% of the variation.

Conclusions: Maternal thyroxine during pregnancy did not improve child cognition at age 9.5 years. Our findings confirmed CATS-I and suggest that the lack of treatment effect may be a result of the similar proportion of IQ < 85 in children of women with normal-GTF and SGTF. (*J Clin Endocrinol Metab* 103: 1583–1591, 2018)

Tri-iodothyronine and thyroxine (T₄) are essential for early brain development, and maternal thyroid hormones are required by the fetus until its own thyroid starts to function, which can be as late as 18 weeks gestation (1–3). Before this, thyroid hormones in the fetal brain are solely of maternal origin (4, 5). Thyroid dysfunction occurs in ~2.5% of pregnancies (6), and severe hypothyroidism during the first two trimesters may result in irreversible neurologic deficits, although the effect of more modest variation in thyroid hormone levels is unclear. Later in pregnancy, the fetus may be better able to compensate for any lack of maternal thyroid hormones, but compensation is likely to be incomplete until the fetal thyroid is fully functional at term (7).

Several studies reported that higher levels of maternal thyroid-stimulating hormone (TSH) during pregnancy may be associated with a negative impact on the child's intelligence (8–11), but this was not confirmed by others (12, 13). Likewise, findings for low maternal thyroxine levels are contradictory with some (9, 13–17) but not all (10, 18–21) studies, providing evidence of lowered intelligence in the children. As well as intelligence quotient (IQ) and general cognition, further deficits for offspring following exposure to underactive maternal thyroid function have been identified, including memory (15, 22–25) and motor difficulties (8, 9, 16, 26, 27), among others.

The Controlled Antenatal Thyroid Screening (CATS) study commenced in 2002 (CATS-I) and was the first randomized controlled trial (RCT) to investigate the effect of screening and treatment of hypothyroidism during pregnancy on child cognition (28). Women ($n = 21,846$) were recruited at a median gestation of 12 weeks, 3 days [in 10 centers in United Kingdom (UK) and 1 center in Turin, Italy]. Mothers were defined as having suboptimal gestational thyroid function (SGTF) if their free thyroxine (FT₄) was <2.5th percentile, and/or TSH was >97.5th percentile, as assessed during the CATS study, and one-half was treated with 150 µg levothyroxine (LT₄) daily. Offspring born to mothers with SGTF had their IQ assessed at age 3 years, and no difference was found between those whose mothers were treated (mean IQ 99.2) or untreated (100.0) during pregnancy ($P = 0.40$). Similar results were obtained in a recent study from Casey *et al.* (29), who reported no beneficial effect on offspring cognition up to age 5 of treating mothers with subclinical hypothyroidism or hypothyroxinemia at 16.7 or 17.8 weeks mean gestation, respectively. The young age of the children, when tested in these large RCTs, might explain the reported lack of treatment effect. IQ evaluations below age 5 may serve as a general indicator of cognitive function but may not be best suited as a longer-term measure of cognitive function (30). Therefore, the primary aim of CATS II was to

measure the children's cognitive function at age 9 years using a more in-depth battery of tests. Furthermore, neither of these trials compared the IQs of children from euthyroid mothers with those of mothers with SGTF to elucidate whether there is a deficit requiring treatment. Our second aim addressed this point by assessing cognitive function in children from mothers with normal gestational thyroid function (normal-GTF). The dose of LT₄ used in the CATS study was relatively high, and recent reports suggest adverse effects of cognition from both too much and too little thyroid hormone (31). Consequently, we explored a possible effect of "overtreatment" (defined as maternal FT₄ above the 97.5th percentile of the CATS-I UK cohort) on IQ scores. Finally, we analyzed the correlation between cognitive assessments undertaken at age 3 and 9 years, as this will be invaluable when designing future studies.

Methods

Study design and population

The original CATS study was previously described in detail (28). In brief, CATS-I recruited 21,846 women (excluding history of thyroid disease, twin pregnancies, maternal age <18 years, or gestational age > 15 weeks and 6 days), predominantly in the UK, at their first antenatal hospital appointment. Participants were randomized either to screen (treated) or control (untreated) groups, the former having their thyroid function tested immediately, and the latter after their child was born. If the mother's FT₄ was <2.5th percentile, and/or TSH was >97.5th percentile, then they were classified as having SGTF, with percentiles calculated from the CATS cohort. Women in the screen group with SGTF were treated with LT₄ (starting dosage 150 µg) for the remainder of their pregnancies. The primary outcome was children's IQ at age 3 from the screen and control groups.

CATS-II included only UK participants for logistical reasons ($n = 16,346$). The target sample size was informed by prior power calculations (see later in text). All CATS mothers from the UK SGTF-treated and SGTF-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 $n = 15,737$) and named "normal-GTF"; a random sample of 4000 from this group was also invited to participate, again by letter (Fig. 1).

Cognitive assessments

CATS-II IQ and additional cognitive assessments were conducted when children were aged 7.00 to 10.92 years (32), either in the research center or in their homes. One psychologist (C. Hales) undertook all of the CATS-II assessments to allow good consistency and was unaware of the participant group. Ten percent of assessments were double scored (by R. Paradice) to ensure accuracy (32). IQ was measured using the Wechsler Intelligence Scale for Children, Fourth Edition UK (WISC-IV), which generated a full-scale IQ (FSIQ), calculated equally from four subdomains: verbal comprehension IQ (VCIQ), perceptual reasoning IQ (PRIQ), working memory IQ (WMIQ), and processing speed IQ (PSIQ). Additional cognitive assessments (8, 22)

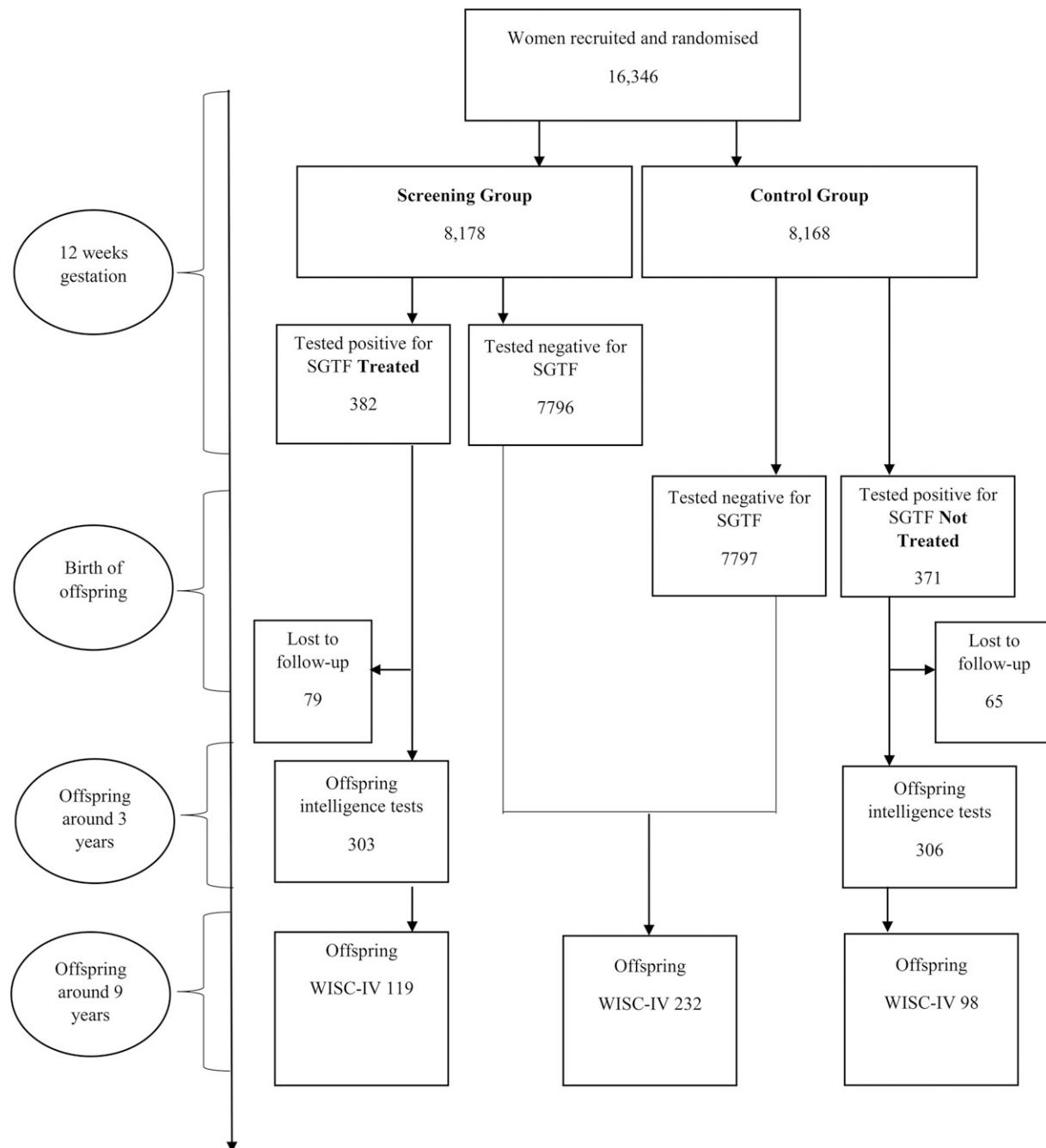


Figure 1. Flow chart of recruitment to the CATS study illustrates initial recruitment for CATS-I, when child IQ was assessed at 3 years of age, and the follow-up study, CATS-II, in which child IQ was assessed at 9 years of age. WISC-IV, Wechsler Intelligence Scale for Children, Fourth Edition UK.

were administered to some children (those not too tired following WISC administration), using the Developmental Neuropsychological Assessment (NEPSY), Second Edition (details can be found in the Supplemental Information). These assessments tested long-term memory [memory for designs (MD) delayed and list memory (LM)], working memory (MD and narrative memory), and fine-motor coordination (fingertip-tapping dominant hand and fingertip-tapping nondominant hand). As the normal-GTF group means for both assessments were close to the anticipated values (WISC-IV IQ: 100; additional NEPSY assessments: 10), the authors conclude that there was no selection bias in which children completed all assessments in CATS-II.

CATS-II was approved by the Wales Research Ethics Committee 2 (reference no. 10/WSE03/33) and Cardiff and Vale University Health Board. Written and informed consent was obtained from all mothers, both in CATS-II and initially during their pregnancies; child assent was obtained during the research center visits. Missing data were largely a result of nonresponse to invitation.

Sample-size justification

Samples of 120 participants from the treated (CATS-I screen) and untreated SGTF (CATS-I control) groups would have 90% power to detect a difference of six points in mean IQ or 80% power with a 5% two-sided significance level to detect a 1.97

increase in odds of $\text{IQ} < 85$ in untreated SGTF, assuming the mean IQ is 100 with a standard deviation (SD) of 15 (32). Participants (240; 1.5%) from the normal-GTF group (CATS-I normal thyroid function in test and screen groups) were required to assess whether maternal SGTF influenced their children's IQ.

Analyses

The data were analyzed in SPSS, version 20, and STATA, version 12, in accordance with the prespecified statistical plan (32).

The primary analysis assessed the odds of $\text{FSIQ} < 85$ in the normal-GTF and the merged SGTF group. An interaction term for treatment of SGTF was then added, all using logistic regression. Mean IQ differences and percentages with $\text{FSIQ} < 85$ were also compared among the three groups. Univariate analysis was followed by multivariate analysis of covariates (MANCOVA) to adjust for key potential covariates in the following four models:

Model 1: crude

Model 2: adjusted for child sex

Model 3: adjusted for model 2 and age of mothers at birth of offspring and whether the child was breastfed

Model 4: adjusted for model 3 and schooling (Welsh- or English-medium school attended), place of assessment (home or research center), and socioeconomic status (calculated from postcode social deprivation scores obtained from <https://statswales.wales.gov.uk/Catalogue/Community-Safety-and-Social-Inclusion/Welsh-Index-of-Multiple-Deprivation/WIMD-2014> for Wales and <http://apps.opendatacommunities.org/showcase/deprivation> for England); a score of one signifies most deprived and five, least deprived.

Step-wise analysis of covariates was performed only for binary outcomes, but all six covariates were included in continuous analyses.

Secondary analyses explored several aspects. We assessed using likelihood ratio tests, whether response to treatment best fitted a proportional or nonproportional model. With the use of the normal-GTF to the untreated SGTF data, we could investigate whether maternal TSH influenced FSIQ.

We also compared subdomain IQs (VCIQ, PRIQ, WMIQ, and PSIQ) in the treated and untreated SGTF groups to explore the effect of treatment, initially, by logistic regressions for scores < 85 and then a MANCOVA adjusted for the six covariates for mean scores. The additional cognitive assessments were also compared by a MANCOVA and an analysis of covariance for the LM subtest (reduced dataset as a result of late introduction).

Sensitivity analyses comprised comparison of CATS-I and CATS-II VCIQ, PRIQ, and FSIQs using Pearson correlations.

As exploratory analyses, within the broad term of SGTF, we investigated subclinical hypothyroidism ($\text{FT4} > 2.5\text{th}$ and $\text{TSH} > 97.5\text{th}$ percentiles), isolated hypothyroxinemia ($\text{FT4} < 2.5\text{th}$ and $\text{TSH} < 97.5\text{th}$ percentiles), and overt hypothyroidism ($\text{FT4} < 2.5\text{th}$ and $\text{TSH} > 97.5\text{th}$ percentiles). These were calculated by MANCOVAs (IQs, additional cognitive assessments, and LM) to include interactions among the three groups, normal-GTF, and whether maternal SGTF was treated.

Finally, we explored differences among participants, taking account of those we defined as "overtreated," *i.e.*, the treated

SGTF group whose FT4 values were above the 17.7-pmol/L threshold established by the 97.5th percentile at recruitment in the UK CATS sample. We compared oversupplementation with child FSIQ < 85 first, followed by analyses for mean scores, all adjusted for the same covariates detailed previously.

Supplemental exploratory analyses can be found in the Supplemental Information: subclinical hypothyroidism, isolated hypothyroxinemia, and overt hypothyroidism.

Results

Group characteristics

In CATS-I, 16,346 women were UK-based and provided the prospective cohort for CATS-II. There were 382 treated and 371 untreated for SGTF; of these, 303 treated and 306 untreated SGTF offspring completed IQ testing at age 3.2 years. No data were collected from the normal-GTF group.

In CATS-II, IQ assessment occurred in a total of 449 children at a mean age of 9.5 years: 119 treated SGTF, 98 untreated SGTF, and 232 from the normal-GTF group (Fig. 1). Smaller groups completed the additional cognitive assessments (see Supplemental Data for explanations): 110 treated SGTF, 85 untreated SGTF, and 215 normal-GTF.

At recruitment into CATS-I, CATS-II mothers from normal-GTF, treated, and untreated SGTF groups had median TSHs of 1.16, 4.09, and 3.57 mU/L, respectively, and mean FT4s were 14.12, 11.92, and 11.79 pmol/L, respectively (Table 1). The CATS-I and CATS-II SGTF samples were largely unbiased (statistics presented in Supplemental Table 1).

Significant differences among the CATS-II participant groups are detailed in Table 1. As anticipated, maternal FT4 and TSH, at recruitment into CATS-I, were higher (FT4) and lower (TSH) in the normal-GTF compared with both SGTF groups. Maternal TSH was higher in the treated compared with untreated SGTF CATS-II mothers. Mean maternal age at consent into CATS-I was higher in the normal-GTF compared with the treated SGTF group, although only by 0.8 year. Likewise, a difference between the groups was seen in those from the SGTF groups, being more likely to opt for participation from their home rather than attending the research clinic. The children in the normal-GTF group were significantly older (by just 4 months) than the SGTF groups.

Primary analysis

There was no significant difference for odds of $\text{FSIQ} < 85$ between the normal-GTF and merged SGTF groups (fully adjusted odds ratio [OR] = 1.15 [95% confidence interval (CI) 0.52, 2.51]; $P = 0.731$). This nonsignificant finding was sustained when an interaction term for treatment was included, although treatment improved FSIQ [untreated, fully adjusted OR = 1.33 (95% CI 0.53,

Table 1. Characteristics of the Cohort

Characteristics	Groups					
	Normal-GTF (n = 232)	Treated SGTF (n = 119)	Untreated SGTF (n = 98)	Normal-GTF vs Treated SGTF, P	Normal-GTF vs Untreated SGTF, P	Treated SGTF vs Untreated SGTF, P
Thyrotropin at CATS-I consent, mIU/L	1.16 (0.66–1.83)	4.09 (1.79–5.09)	3.57 (1.18–4.49)	0.001	0.001	0.007
T4 at CATS-I consent, pmol/L	14.12 (1.76)	11.92 (1.93)	11.79 (1.88)	0.001	0.001	1.000
Maternal age at CATS-I consent, years	31.85 (5.16)	30.26 (5.08)	31.05 (4.88)	0.018	0.578	0.767
Social deprivation/ socioeconomic status	4.00 (3–5; mean 3.71)	4.00 (3–5; mean 3.78)	4.00 (2–5; mean 3.37)	0.807	0.359	0.161
1	26 (11%)	15 (13%)	16 (16%)			
2	27 (11%)	12 (10%)	14 (14%)			
3	36 (15%)	15 (13%)	17 (17%)			
4	43 (18%)	19 (16%)	20 (20%)			
5	100 (43%)	58 (49%)	31 (32%)			
Child breastfed over 1 month, n (%)	150 (65%)	72 (60%)	56 (57%)	0.445	0.198	0.616
Child Characteristics						
Male children, n (%)	177 (50%)	65 (55%)	49 (50%)	0.457	0.943	0.497
Child age at participation	9.83 (9.00–10.33)	9.58 (9.08–9.94)	9.50 (9.00–9.94)	0.001	0.024	0.710
Where child was assessed				0.001	0.001	0.554
Home	120 (52%)	92 (77%)	79 (81%)			
Research center	112 (48%)	27 (23%)	19 (19%)			
Child's language						
English school/home	180 (78%)	95 (80%)	85 (87%)	0.950	0.364	0.541
Welsh school/English home	42 (18%)	20 (17%)	11 (11%)			
Welsh school/home	7 (3%)	3 (2%)	1 (1%)			
English school/other home	2 (1%)	1 (1%)	1 (1%)			
Welsh school/other home	1 (1%)	0	0			

Data are expressed as median (interquartile range), mean (SD), or the number (n) of participants [percentage (%)]. Socioeconomic status is based on a social deprivation score, with one being the most deprived. Child's language describes whether the child speaks English, both at home and in school; Welsh, in both locations; a combination of English and Welsh; or an additional language.

3.34); treatment OR = 0.75 (95% CI 0.27, 2.06); $P = 0.576$. Table 2 displays the FSIQ regression models.

The percentages with IQ < 85 were 6.03% in normal-GTF, 7.56% in treated, and 11.22% in untreated SGTF groups (Table 3; χ^2 P for the trend = 0.11).

Secondary analyses

Do data fit a proportional or nonproportional model?

Mean child FSIQs per group were 103.10 (SD 11.68), 101.76 (12.04), and 102.31 (13.27) for the normal-GTF, treated, and untreated SGTF groups, respectively (Table 3). There was no difference among the mean FSIQ scores of the three participant groups ($P = 0.678$). There was no significant difference for odds of the normal-GTF children having higher FSIQs compared with the treated SGTF children [$OR = 0.99$ (95% CI 0.38, 2.52); $P = 0.98$]. This was a result of a mean IQ difference of only 0.79 between the groups.

Does maternal TSH predict FSIQ?

Analysis of the relationship between FSIQ and thyroid status in normal-GTF and untreated SGTF revealed no clear association between TSH [Beta coefficient = 0.43 (95% CI –0.68, 1.56); $P = 0.442$] and FT4 [Beta coefficient = 0.33 (95% CI –0.25, 0.91); $P = 0.270$] on FSIQ in the fully adjusted model.

Analyses of women with SGTF, by dividing the FSIQ score into quintiles, did not reveal any benefit of treatment in the fully adjusted model ($P = 0.98$), with no evidence of a nonproportional effect ($P = 0.75$; data not shown).

Does treatment of SGTF affect any subdomain?

No differences were found among subdomain IQ scores < 85 (see Table 2 for sub-IQ regression models) or for mean subdomain IQ scores for VCIQ, PRIQ, WMIQ, and PSIQ among the groups ($P = 0.193$). The mean scores of the additional cognitive assessments were also compared, with no difference identified among the three participant groups ($P = 0.732$, LM $P = 0.266$; Table 3).

Sensitivity analysis

As CATS-II followed the UK sample, we analyzed the CATS-I UK-only cohort ($n = 609$) and revealed IQ < 85 in 14% treated and 17% untreated; the difference was not significant. Furthermore, there was no significant difference in percentage IQ < 85 treated vs untreated in the CATS-II subset of CATS-I ($n = 212$).

Pearson correlations to assess how associated the scores were from the Wechsler Preschool and Primary Scale of Intelligence, Third Edition, and the WISC-IV for the treated

Table 2. Logistic Regressions for Odds of IQ Below 85

IQs	Models	Merged SGTF to Normal-GTF OR (95% CI)	P Interaction	OR Untreated (95% CI)	OR Treatment (95% CI)	P Treatment Interaction
FSIQ	1	1.58 (0.78, 3.21)	0.206	1.97 (0.86, 4.50)	0.65 (0.26, 1.63)	0.355
	2	1.57 (0.77, 3.19)	0.217	1.98 (0.86, 4.55)	0.63 (0.25, 1.59)	0.325
	3	1.38 (0.66, 2.86)	0.389	1.77 (0.75, 4.16)	0.61 (0.23, 1.58)	0.308
	4	1.15 (0.52, 2.51)	0.731	1.33 (0.53, 3.34)	0.75 (0.27, 2.06)	0.576
VCIQ	1	1.08 (0.57, 2.03)	0.820	0.89 (0.38, 2.09)	1.38 (0.55, 3.48)	0.491
	2	1.07 (0.57, 2.02)	0.833	0.89 (0.38, 2.09)	1.36 (0.54, 3.44)	0.506
	3	0.99 (0.52, 1.88)	0.968	0.82 (0.34, 1.93)	1.38 (0.54, 3.53)	0.491
	4	0.93 (0.47, 1.83)	0.834	0.70 (0.29, 1.73)	1.62 (0.62, 4.20)	0.317
PRIQ	1	1.82 (0.84, 3.94)	0.130	2.54 (1.06, 6.07)	0.49 (0.18, 1.33)	0.156
	2	1.82 (0.84, 3.94)	0.131	2.54 (1.06, 6.07)	0.49 (0.18, 1.33)	0.156
	3	1.60 (0.73, 3.53)	0.238	2.31 (0.95, 5.62)	0.46 (0.17, 1.28)	0.132
	4	1.35 (0.59, 3.09)	0.482	1.78 (0.69, 4.56)	0.56 (0.19, 1.58)	0.268
WMIQ	1	1.48 (0.78, 2.81)	0.232	1.35 (0.60, 3.04)	1.17 (0.50, 2.77)	0.715
	2	1.47 (0.77, 2.79)	0.241	1.35 (0.60, 3.05)	1.15 (0.49, 2.73)	0.742
	3	1.33 (0.69, 2.57)	0.393	1.21 (0.53, 2.78)	1.18 (0.49, 2.84)	0.713
	4	1.26 (0.63, 2.53)	0.513	1.04 (0.43, 2.50)	1.42 (0.57, 3.52)	0.449
PSIQ	1	0.79 (0.36, 1.71)	0.550	0.88 (0.33, 2.32)	0.81 (0.25, 2.61)	0.729
	2	0.78 (0.36, 1.69)	0.524	0.88 (0.33, 2.33)	0.79 (0.24, 2.53)	0.688
	3	0.75 (0.34, 1.63)	0.463	0.85 (0.32, 2.27)	0.77 (0.24, 2.49)	0.664
	4	0.75 (0.33, 1.68)	0.482	0.82 (0.20, 2.24)	0.85 (0.26, 2.77)	0.783

Data are expressed as OR, with 95% CIs. Model 1, unadjusted; model 2, adjusted for child's sex; model 3, adjusted for model 2 and whether the mother breastfed >1 month and mother's age at time of study consent during pregnancy; model 4, adjusted for model 3 and where the child was assessed, child's language spoken at school and home, and social deprivation score.

and untreated SGTF groups found that all scores were positively correlated ($P < 0.0001$). Furthermore, age 3 IQ predicts 45% of the variation in age 9 IQ with other variables, such as breastfeeding contributing only an additional 1%.

Exploratory analyses

Different types of abnormal thyroid function (sub-clinical hypothyroidism, isolated hypothyroxinemia) were also explored using MANCOVA. No significant differences were found in the mean IQ scores (IQ < 85) or

Table 3. Mean Scores for IQs

Cognitive Assessment	Groups			
	Normal-GTF (n = 232)	Merged SGTF (n= 217)	Treated SGTF (n = 119)	Untreated SGTF (n= 98)
WISC				
VCIQ	99.81 (11.26) 28 (12%)	98.60 (11.42) 30 (14%)	97.56 (9.95) 19 (16%)	99.86 (12.93) 11 (11%)
<85				
PRIQ	105.37 (12.30) 11 (5%)	104.55 (12.87) 18 (8%)	104.49 (12.26) 7 (6%)	104.63 (13.64) 11 (11%)
<85				
WMIQ	99.91 (11.24) 18 (8%)	99.81 (12.72) 24 (11%)	99.73 (13.28) 14 (12%)	99.90 (12.07) 10 (10%)
<85				
PSIQ	103.66 (12.75) 22 (9%)	102.39 (12.73) 18 (8%)	103.16 (12.71) 8 (7%)	101.45 (12.75) 10 (10%)
<85				
FSIQ	103.10 (11.68) 15 (6%)	102.01 (12.59) 21 (10%)	101.76 (12.04) 10 (8%)	102.31 (13.28) 11 (11%)
<85				
NEPSY	n = 215	n = 195	n = 110	n = 85
MD	10.36 (2.92)	9.69 (3.13)	9.63 (3.27)	9.76 (2.96)
MDD	10.34 (2.65)	9.86 (2.84)	9.77 (2.79)	9.98 (2.92)
FTDH	12.24 (1.60)	11.94 (1.45)	11.90 (1.41)	12.01 (1.52)
FTNDH	12.51 (1.37)	12.24 (1.41)	12.21 (1.39)	12.31 (1.46)
NM	11.56 (2.76) n = 170	11.06 (2.76) n = 146	11.02 (2.78) n = 78	11.12 (2.74) n = 68
LM ^a	10.93 (2.84)	10.62 (2.86)	10.63 (3.13)	10.60 (2.54)

Data expressed as means (SD) of group or the number (n) of participants [percentage (%)] having IQ < 85.

Abbreviations: FTDH, fingertip-tapping dominant hand; FTNDH, fingertip-tapping nondominant hand; MDD, MD delayed; NM, narrative memory.

^aReduced dataset.

additional assessments between children of treated and untreated mothers. Similar results were obtained in the offspring of a small number of women with overt hypothyroidism, identified during participation in CATS, although $\text{IQ} < 85$ was apparent in 0% of the treated but 10% of the untreated groups. These analyses are presented in Supplemental Table 2.

Oversupplementation

Finally, we explored differences among participants, taking account of those in the treated SGTF group with raised FT4 values [20 weeks mean FT4 16.19 (2.83), TSH median 0.33 (0.08 to 0.99); 30 weeks mean FT4 15.56 (2.50), median TSH 0.27 (0.03 to 0.84)]. The threshold for high FT4 was established by the 97.5th percentile recruitment in the UK CATS sample (17.7 pmol/L); one-third of the treated SGTF had $\text{FT4} > 17.7 \text{ pmol/L}$.

We compared oversupplementation with child FSIQ < 85 first, followed by analyses for mean scores, all adjusted for the same covariates detailed previously. There was no significant effect on a child's IQ < 85 and no difference among mean IQ scores of the groups or additional cognitive assessments, including the LM subtest ($P = 0.875$, $P = 0.765$, and $P = 0.951$, respectively; data not shown).

Of note, we observed no detrimental effect of oversupplementation on IQ < 85 in children of such women in CATS-I for whom we had information on FT4 levels after therapy was initiated (UK cohort, $n = 609$).

Discussion

We revisited the effects of treatment of SGTF on cognition in the CATS children at an average age of 9.5 years. Our results confirm those of CATS-I, in that we saw no significant differences in FSIQ < 85 or mean IQ scores in the children of treated and untreated women. Our results also confirm those of Casey and colleagues (29), who reported no beneficial effect on offspring cognition up to age 5 of treating mothers with subclinical hypothyroidism or hypothyroxinemia at 16.7 or 17.8 weeks mean gestation, respectively. Of interest Haddow *et al.* (8) reported that mean FSIQ scores and FSIQ scores < 85 were not significantly different in comparing children born to mothers who were treated or not ($P = 0.20$ and $P = 0.90$, respectively), although the study was retrospective, and the treatment groups were small. In contrast to our findings, however, the study by Haddow *et al.* (8) showed that the IQ of children born to untreated mothers was significantly lower than those of control children.

One criticism of CATS-I was that cognitive assessments were conducted in children at too young an age for differences to be evident. Our current findings indicate that this

may not be the case, as we found that IQ scores at ages 3 and 9 were strongly correlated in the two CATS studies with FSIQ at age 3, predicting 45% of the variability in FSIQ at age 9 and with other factors contributing very little.

The design of the CATS-I study has also been questioned in relation to the timing of initiation of LT4 therapy. The fetus relies wholly on maternal T4 delivery, up until ~ 14 to 18 weeks gestation, when its own thyroid gland becomes functional (7). Fetal brain development begins immediately after conception and, therefore, treatment initiated at 12 to 13 weeks may have missed the early, critical phase of brain development. The CATS study participants were recruited during their first scheduled visit to the antenatal clinic, which generally fell toward the end of the first trimester (median of 12 weeks and 3 days) (33). Likewise, LT4 supplementation in the study by Casey *et al.* (29) was started even later, and thus, future trials would benefit from recruiting women at a much earlier stage of pregnancy to overcome these limitations.

A further consideration in the CATS study design is that the starting dose of LT4 administered may have been too high, and therefore, adverse outcomes in women who were overtreated may have masked any benefits of treatment. The CATS-I study was the first RCT to investigate the effects of treatment of SGTF in pregnancy, and hence, there were no previous studies for guidance. Furthermore, there is no universal consensus on T4 supplementation dose, even for the treatment of women with overt hypothyroidism who become pregnant. Of note, guidelines for the management of thyroid function during pregnancy recommend assay of TSH alone, and indeed, treatment in CATS-I was monitored and adjusted based on TSH levels. As a result, approximately one-third of the treated mothers achieved a high FT4, which was accompanied by a switch from a positive correlation between FT4 and age 9 cognition at recruitment to a negative correlation after treatment (Supplemental Information). However, in contrast to a study illustrating a biphasic effect of FT4 on cognition, with children of women with both low and high FT4 levels displaying lower IQs and smaller gray matter and cortex volumes (31), we observed no significant difference in the proportion of IQ < 85 at age 9 in children of overtreated mothers compared with the rest. Furthermore, we did not find any detrimental effect on IQ < 85 in children of such women when we analyzed the age 3 cognition data in CATS-I (UK-only cohort).

CATS-II included children from normal-GTF women and found no difference in IQ measures between these and children from SGTF mothers, whether treated or not. This confirmed previous studies reporting no effect of low thyroid function on offspring intelligence or cognition

(10, 12, 13, 18–21) and may, to some extent, explain the absence of treatment benefits observed in the trial. However, our results contradict many animal studies, possibly because the thyroid abnormalities in the CATS mothers are mild when compared with models induced, *e.g.*, by thyroidectomy. The lack of agreement on the effects of FT4 on cognition in observational studies is the result of varying definitions of SGTF, the lack of universal pregnancy-specific reference ranges for thyroid function tests, and the application of various tools to measure cognition in children across the age spectrum. Hence, it is not surprising that the benefits of universal screening during pregnancy on cognition remain hotly debated, although other adverse pregnancy outcomes have been well reported (such as pre-eclampsia, miscarriage, and preterm birth) (34–36). In our protocol paper (32), one of the secondary analyses planned to investigate whether the combination of low maternal FT4 during pregnancy and the presence of an adverse deiodinase 2 genotype in her child would impact cognition. The hypothesis followed reports that Thr92Ala reduced conversion of T4 to triiodothyronine (37). We genotyped 426 CATS children, finding 73 alanine homozygotes; when a mother had low FT4 during pregnancy, and the child had the homozygous alanine deiodinase 2 genotype, treatment appeared to reduce the odds of FSIQ < 85 (reduced OR from 5.72 to 1.85), although this was nonsignificant and included only a small number of the participants (data not shown).

Our study has some limitations, although throughout all analyses, adjustments were made to control for extraneous effects. The CATS-II power calculation was based on an IQ difference of six points, as found by Haddow *et al.* (8) in offspring of women with overt hypothyroidism. We studied women with less severe thyroid dysfunction, and thus, the study may have been underpowered to detect more subtle cognitive variation. This was exacerbated by the recruitment challenges that we faced from the outset—the main problem was a result of participants having relocated since participating in CATS-I and not responding to invitation. As the study developed, the recruitment process evolved, and rates improved, but the extension of the data-collection period would have taken the children closer to puberty and its complications. There were some differences noted among the three groups, raising the possibility of bias. However, significantly older normal-GTF children than those from the SGTF groups should not have affected the results, as both assessment tools used have scores age corrected in 3-month intervals. Likewise, differences in maternal age at recruitment and place of child assessment were both covariates controlled for in the analyses.

In conclusion, results obtained in the current follow-up study have shown no effect of LT4 supplementation in

women with SGTF on child IQ at age 9. These findings support those of the original CATS-I study and a recent large RCT. Our data are consistent with the lack of treatment effect being a result of the similar proportion of IQ < 85 in children of normal-GTF and SGTF mothers, rather than the age of cognitive assessment or the relatively high dose of LT4 supplementation. However, future large, randomized trials, with LT4 interventions at a much earlier stage of pregnancy (or preconception), may still be warranted, as the benefits of treatment may not be fully realized unless treatment is commenced early.

Acknowledgments

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 to the CATS project.

Financial Support: This work was funded by the Charles Wolfson Trust, Action Medical Research (Project Code GN2033)/The Henry Smith Charity (20122759 GN 2033; grants to M.L.), and the American Thyroid Association (grant to P.N.T.).

Author Contributions: C.H. collected the data, was involved in writing the report, and analyzed the data with P.N.T. S.C., R.P., K.M., L.Z., M.G., A.B., O.O., I.M., M.S.D., J.W.G., C.D., J.H.L., and D.A.R. contributed to study design, data analyses, and writing the report. M.L. designed and managed the project, supervised analyses, and contributed to the report.

Correspondence and Reprint Requests: Marian Ludgate, PhD, Division of Infection and Immunity, School of Medicine, Cardiff University, Cardiff, Wales CF14 4XN, United Kingdom. E-mail: ludgate@cardiff.ac.uk.

Disclosure Summary: The authors have nothing to disclose.

References

1. Bernal J, Nunez J. Thyroid hormones and brain development. *Eur J Endocrinol.* 1995;133(4):390–398.
2. Ahmed OM, El-Gareib AW, El-Bakry AM, Abd El-Tawab SM, Ahmed RG. Thyroid hormones states and brain development interactions. *Int J Dev Neurosci.* 2008;26(2):147–209.
3. Zimmermann MB. Iodine deficiency. *Endocr Rev.* 2009;30(4): 376–408.
4. Morreale de Escobar G. The role of thyroid hormone in fetal neurodevelopment. *J Pediatr Endocrinol Metab.* 2001;14(Suppl 6): 1453–1462.
5. Contempré B, Jauniaux E, Calvo R, Jurkovic D, Campbell S, de Escobar GM. Detection of thyroid hormones in human embryonic cavities during the first trimester of pregnancy. *J Clin Endocrinol Metab.* 1993;77(6):1719–1722.
6. Lazarus JH, Premawardhana LD. Screening for thyroid disease in pregnancy. *J Clin Pathol.* 2005;58(5):449–452.
7. Rover JF, Willoughby KA. Maternal thyroid function during pregnancy: effects on the developing fetal brain. In: Zimmermann AW, Connors SL, eds. *Maternal Influences on Fetal Neurodevelopment: Clinical and Research Aspects.* New York, NY: Springer Science + Business Media; 2010:55–77.
8. 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. *N Engl J Med.* 1999;341(8):549–555.
9. 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. *Clin Endocrinol (Oxf).* 2010;72(6):825–829.
 10. 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. *J Clin Endocrinol Metab.* 2011;96(10):3234–3241.
 11. Klein RZ, Sargent JD, Larsen PR, Waisbren SE, Haddow JE, Mitchell ML. Relation of severity of maternal hypothyroidism to cognitive development of offspring. *J Med Screen.* 2001;8(1):18–20.
 12. Smit BJ, Kok JH, Vulsmma T, Briët JM, Boer K, Wiersinga WM. Neurologic development of the newborn and young child in relation to maternal thyroid function. *Acta Paediatr.* 2000;89(3):291–295.
 13. 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. *J Clin Endocrinol Metab.* 2010;95(9):4227–4234.
 14. 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. *J Clin Endocrinol Metab.* 2014;99(7):2383–2390.
 15. Suárez-Rodríguez M, Azcona-San Julián C, Alzina de Aguilar V. Hypothyroxinemia during pregnancy: the effect on neurodevelopment in the child. *Int J Dev Neurosci.* 2012;30(6):435–438.
 16. Pop VJ, Brouwers EP, Vader HL, Vulsmma T, van Baar AL, de Vijlder JJ. Maternal hypothyroxinaemia during early pregnancy and subsequent child development: a 3-year follow-up study. *Clin Endocrinol (Oxf).* 2003;59(3):282–288.
 17. 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(5):511–519.
 18. Craig WY, Allan WC, Kloza EM, Pulkkinen AJ, Waisbren S, Spratt DI, Palomaki GE, Neveux LM, Haddow JE. Mid-gestational maternal free thyroxine concentration and offspring neurocognitive development at age two years. *J Clin Endocrinol Metab.* 2012;97(1):E22–E28.
 19. Chevrier J, Harley KG, Kogut K, Holland N, Johnson C, Eskenazi B. Maternal thyroid function during the second half of pregnancy and child neurodevelopment at 6, 12, 24, and 60 months of age. *J Thyroid Res.* 2011;2011:426427.
 20. Grau G, Aguayo A, Vela A, Aniel-Quiroga A, Espada M, Miranda G, Martínez-Indart L, Martul P, Castaño L, Rica I. Normal intellectual development in children born from women with hypothyroxinemia during their pregnancy. *J Trace Elem Med Biol.* 2015;31:18–24.
 21. Oken E, Braverman LE, Platek D, Mitchell ML, Lee SL, Pearce EN. Neonatal thyroxine, maternal thyroid function, and child cognition. *J Clin Endocrinol Metab.* 2009;94(2):497–503.
 22. Willoughby KA, McAndrews MP, Rovet JF. Accuracy of episodic autobiographical memory in children with early thyroid hormone deficiency using a staged event. *Dev Cogn Neurosci.* 2014;9:1–11.
 23. Willoughby KA, McAndrews MP, Rovet J. Effects of early thyroid hormone deficiency on children's autobiographical memory performance. *J Int Neuropsychol Soc.* 2013;19(4):419–429.
 24. Willoughby KA, McAndrews MP, Rovet JF. Effects of maternal hypothyroidism on offspring hippocampus and memory. *Thyroid.* 2014;24(3):576–584.
 25. Pharoah POD, Connolly KJ. Relationship between maternal thyroxine levels during pregnancy and memory function in childhood. *Early Hum Dev.* 1991;25(1):43–51.
 26. Ishaik G, Asztalos E, Perlman K, Newton S, Frisk V, Rovet J. Hypothyroxinemia of prematurity and infant neurodevelopment: a pilot study. *J Dev Behav Pediatr.* 2000;21(3):172–179.
 27. Henrichs J, Ghassabian A, Peeters RP, Tiemeier H. Maternal hypothyroxinemia and effects on cognitive functioning in childhood: how and why? *Clin Endocrinol (Oxf).* 2013;79(2):152–162.
 28. 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. *N Engl J Med.* 2012;366(6):493–501.
 29. Casey BM, Thom EA, Peaceman AM, Varner MW, Sorokin Y, Hirtz DG, Reddy UM, Wapner RJ, Thorp JM Jr, Saade G, Tita AT, Rouse DJ, Sibai B, Iams JD, Mercer BM, Tolosa J, Caritis SN, VanDorsten JP; Eunice Kennedy Shriver National Institute of Child Health and Human Development Maternal–Fetal Medicine Units Network. Treatment of subclinical hypothyroidism or hypothyroxinemia in pregnancy. *N Engl J Med.* 2017;376(9):815–825.
 30. Mackintosh N. *IQ and Human Intelligence.* 2nd ed. Oxford, UK: Oxford University Press; 2011.
 31. Korevaar TI, Muetzel R, Medici M, Chaker L, Jaddoe VWV, de Rijke YB, Steegers EAP, Visser TJ, White T, Tiemeier H, Peeters RP. 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.* 2016;4(1):35–43.
 32. Hales C, Channon S, Taylor PN, Draman MS, Muller I, Lazarus J, Paradice R, Rees A, Shillabeer D, Gregory JW, Dayan CM, Ludgate M. The second wave of the Controlled Antenatal Thyroid Screening (CATS II) study: the cognitive assessment protocol. *BMC Endocr Disord.* 2014;14(1):95.
 33. Daniels GH, Dayan CM. *Fast Facts: Thyroid Disorders.* Oxford, UK: Health Press; 2005.
 34. Jouyandeh Z, Hasani-Ranjbar S, Qorbani M, Larijani B. Universal screening versus selective case-based screening for thyroid disorders in pregnancy. *Endocrine.* 2015;48(1):116–123.
 35. van den Boogaard E, Vissenberg R, Land JA, van Wely M, van der Post JAM, Goddijn M, Bisschop PH. Significance of (sub)clinical thyroid dysfunction and thyroid autoimmunity before conception and in early pregnancy: a systematic review. *Hum Reprod Update.* 2011;17(5):605–619.
 36. Reid SM, Middleton P, Cossich MC, Crowther CA, Bain E. Interventions for clinical and subclinical hypothyroidism pre-pregnancy and during pregnancy. *Cochrane Database of Syst Rev.* 2013;31(5):CD007752.
 37. Castagna MG, Dentice M, Cantara S, Ambrosio R, Maino F, Porcelli T, Marzocchi C, Garbi C, Pacini F, Salvatore D. DIO2 Thr92Ala reduces deiodinase-2 activity and serum-T3 levels in thyroid-deficient patients. *J Clin Endocrinol Metab.* 2017;102(5):1623–1630.