Controlled Antenatal Thyroid Screening II: effect of treating maternal sub-optimal 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 Ludgate[1]

[1]School of Medicine, Cardiff University, UK. [2]Centre for Trials Research, Cardiff University, UK. [3]St David's Hospital, Cardiff & Vale University Health Board, UK. [4]Neuroscience and Mental Health Research Institute, Cardiff University, UK.

Key words: pregnancy; cognition; subclinical hypothyroidism; subclinical hyperthyroidism; hypothyroidism; hyperthyroidism.

Correspondence & requests for reprints to: Prof Marian Ludgate, Division of Infection and Immunity, School of Medicine, Cardiff University, Cardiff, Wales UK, ludgate@cardiff.ac.uk phone +44 2920 745457

Funded by The Charles Wolfson Trust, Action Medical Research (project code GN2033)/The Henry Smith Charity (20122759 GN 2033, grants to ML) and American Thyroid Association (grant to PT).

Declaration of interests: All authors declare no competing interests.

Word Count 4042

Abstract

Context & Objective: The Controlled Antenatal Thyroid Screening (CATS) study investigated treatment for suboptimal gestational thyroid function (SGTF) on childhood cognition and found no difference in 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-GTF mothers.

Design, Setting & Participants: One examiner, blinded to participant group, assessed children's IQ (WISC-IV), long-term memory and motor function (NEPSY-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 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 OR=1.15 (95% 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 FT4 >97.5th centile of the entire CATS-I cohort revealed no significant effect on 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 due to the similar proportion of IQ<85 in children of women with normal-GTF and SGTF.

Précis

Cognitive assessments of children aged 9 from the first CATS study confirms no effect of treatment for maternal SGTF on IQ<85 and no IQ difference when compared with children from euthyroid mothers.

Introduction

Triiodothyronine (T3) 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). Prior to this thyroid hormones in the fetal brain are solely of maternal origin (4,5). Thyroid dysfunction occurs in around 2.5% of pregnancies (6) and severe hypothyroidism during the first two trimesters may result in irreversible neurological 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 T4 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), amongst others.

The Controlled Antenatal Thyroid Screening (CATS) study commenced in 2002 (CATS-I) and was the first randomised controlled trial (RCT) to investigate the effect of screening and treatment for hypothyroidism during pregnancy on child cognition (28). Women (n=21,846) were recruited at a median gestation of 12 weeks, 3 days; (ten UK centres and one in Turin, Italy). Mothers were defined as having suboptimal

gestational thyroid function (SGTF) if their FT4 was <2.5th percentile and/or TSH >97.5th percentile as assessed during the CATS study and half were treated with 150µg thyroxine daily. Offspring born to SGTF mothers had their IQ assessed at age 3 years 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 and colleagues 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 (29). 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 SGTF mothers 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 thyroxine 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 'over-treatment' (defined as maternal FT4 above the 97.5th percentile of the CATS-I UK cohort) on IQ scores. Finally we analysed 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). Briefly 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 >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. 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 below). All 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=15 737), and named 'normal-GTF'; a random sample of 4,000 from this group was also invited to participate, again by letter (figure 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 centre or in their homes. One psychologist (CH) undertook all of the CATS-II assessments to allow good consistency

and was unaware of participant group. Ten percent of assessments were double scored (RP) to ensure accuracy (32). IQ was measured using the Wechsler Intelligence Scale for Children (WISC) fourth edition UK version which generated a full-scale IQ (FSIQ) calculated equally from four sub-domains: verbal comprehension IQ (VCIQ), perceptual reasoning IQ (PRIQ), working memory IQ (WMIQ) and processing speed IQ (PSIQ). Additional cognitive assessments (8,22) 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 delayed- MDD, and list memory- LM), working memory (memory for designs- MD, and narrative memory- NM) and fine motor coordination (fingertip tapping dominant hand- FTDH, and fingertip tapping non-dominant hand, FTNDH). 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 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 10/WSE03/33) and Cardiff & 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 centre visits. Missing data were largely due to non-response 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 6 points in mean IQ (31) or 80% power with a 5% two-sided significance level to detect a 1.97

increase in odds of IQ < 85 in untreated SGTF assuming mean IQ to be 100 with a SD of 15 (32). 240 participants (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 her child's IQ.

Analyses

The data were analysed in SPSS version 20 and STATA version 12 in accordance with the pre-specified statistical plan (32).

The primary analysis assessed the odds of 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 FSIQ<85 were also compared between the three groups. Univariate analysis was followed by multivariate analysis to adjust for key potential covariates in 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 centre) 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 1 signifies most deprived and 5 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 non-propotional model. Using 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, then a multivariate analysis of covariance (MANCOVA, adjusted for the six covariates) for mean scores. The additional cognitive assessments were also compared by a MANCOVA, and an analysis of covariance (ANCOVA) for the LM subtest (reduced dataset due to 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 (FT4 >2·5th and TSH >97·5th percentiles), isolated hypothyroxinaemia (FT4 <2·5th and TSH <97·5th percentiles), and overt hypothyroidism (FT4 <2·5th and TSH >97·5th percentiles). These were calculated by MANCOVAs (IQs, additional cognitive assessments, and LM) to include interactions between the three groups; normal-GTF, and whether maternal SGTF was treated or not.

Finally, we explored differences between participants taking account of those we defined as ‘overtreated’ i.e the treated SGTF group whose FT4 values were above the 17.7pmol/L threshold established by the 97.5th percentile at recruitment in the UK

CATS sample. We compared over-supplementation to child FSIQ <85 first, followed by analyses for mean scores, all adjusted for the same covariates detailed above.

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 (figure 1). Smaller groups completed the additional cognitive assessments (please see supplementary 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 3.57mU/L, respectively and mean FT4s were 14.12, 11.92, and 11.79pmol/L, respectively (table 1). The CATS-I and CATS-II SGTF samples were largely unbiased (statistics presented in supplementary table 1).

Significant differences between 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 to the treated SGTF group, though only by 0.8 years. Similarly, 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 FSIQ <85 between the normal-GTF and merged SGTF groups (fully adjusted OR=1.15 (95% CI 0.52, 2.51) p=0.731). This non-significant 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, 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, Chi p for the trend = 0.11).

Secondary analyses

Do data fit a proportional or non-proportional 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 between 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 to the treated SGTF children (OR=0.99 (95%

CI 0.38, 2.52) p=0.98); this was due to 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 (B=0.43 (95% CI -0.68, 1.56) p=0.442) and FT4 (B= 0.33 (95% CI -0.25, 0.91) p=0.270) on FSIQ in the fully adjusted model.

Analysis of women with SGTF by dividing FSIQ score into quintiles did not reveal any benefit of treatment in the fully adjusted model (p=0.98) with no evidence of a non-proportional effect (p=0.75) (data not shown).

Does treatment for SGTF affect any subdomain?

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

Sensitivity Analysis

As CATS-II followed the UK sample we analysed 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 versus untreated in the CATS-II subset of CATS-I (n=212).

Pearson correlations to assess how associated the scores were from the WPPSI-III and the WISC-IV for the treated 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 breast feeding contributing only an additional 1%.

Exploratory analyses

Different types of abnormal thyroid function (subclinical hypothyroidism, isolated hypothyroxinema) were also explored using MANCOVA. No significant differences were found in the mean IQ scores, IQ<85 or 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 IQ<85 was apparent in 0% of the treated but 10% of the untreated groups. These analyses are presented in supplementary table 2.

Over-supplementation

Finally, we explored differences between 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 (0.08-0.99); 30 weeks mean FT4 15.56 (2.50), median TSH 0.27 (0.03-0.84). The threshold for high FT4 was established by the 97·5th percentile recruitment in the UK CATS sample (17.7pmol/L); one-third of the treated SGTF had FT4 >17.7pmol/L.

We compared over-supplementation to child FSIQ <85 first, followed by analyses for mean scores, all adjusted for the same covariates detailed above. There was no significant effect on child's IQ<85 and no difference between mean IQ scores of the groups or additional cognitive assessments, including the LM subtest (p=0.875, p=0.765, p=0.951, respectively), data not shown.

Of note, we observed no detrimental effect of over-supplementation 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 for 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 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 (29). Of interest Haddow et al (8) reported that mean FSIQ scores and FSIQ scores <85 were not significantly different 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 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 age 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 levothyroxine therapy. The fetus relies wholly on maternal thyroxine delivery up until about 14-18 weeks gestation when its own thyroid gland becomes functional (7). Fetal brain development begins immediately after conception and therefore treatment initiated at 12-13 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 towards the end of the first trimester (median of 12 weeks and 3 days). (33) Similarly, thyroxine 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 in order to overcome these limitations.

A further consideration in the CATS study design is that the starting dose of levothyroxine administered may have been too high and therefore adverse outcomes in women who were over-treated may have masked any benefits of treatment. The CATS-I study was the first RCT to investigate the effects of treatment for SGTF in pregnancy and hence there were no previous studies for guidance. Furthermore, there is no universal consensus on thyroxine 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 bi-phasic effect of FT4 on cognition, with children of women with both low and high thyroxine levels displaying

lower IQs and smaller grey matter and cortex volumes (35) (31), we observed no significant difference in the proportion of IQ<85 at age 9 in children of over-treated mothers compared with the rest. Furthermore we did not find any detrimental effect on IQ<85 in children of such women when we analysed 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 (D2) genotype in her child would impact cognition. The hypothesis followed reports that Thr92Ala reduced conversion of thyroxine to tri-iodothyronine (37). We genotyped 426 CATS children finding 73 alanine homozygotes; when a mother had low FT4 during pregnancy and the child had the

357 homozygous alanine D2 genotype, treatment appeared to reduce the odds of
358 FSIQ<85 (reduced OR from 5.72 to 1.85), though this was non-significant and included
359 only a small number of the participants (data not shown).

360 Our study has some limitations although throughout all analyses adjustments were
361 made to control for extraneous effects. 1. The CATS-II power calculation was based
362 on an IQ difference of 6 points, as found by Haddow et al in offspring of women with
363 overt hypothyroidism. We studied women with less severe thyroid dysfunction and
364 thus the study may have been underpowered to detect more subtle cognitive variation.
365 2. This was exacerbated by the recruitment challenges we faced from the outset, with
366 the main problem due to participants having re-located since participating in CATS-I
367 and not responding to invitation. As the study developed, the recruitment process
368 evolved and rates improved but extending the data collection period would have taken
369 the children closer to puberty and its complications. 3. There were some differences
370 noted between the three groups raising the possibility of bias. However, significantly
371 older normal-GTF children than those from the SGTF groups should not have affected
372 the results since both assessment tools used have scores age-corrected in three
373 month intervals. Similarly differences in maternal age at recruitment and place of child
374 assessment were both covariates controlled for in the analyses.

375 In conclusion, results obtained in the current follow-up study have shown no effect of
376 thyroxine supplementation in women with SGTF on child IQ at age 9. These findings
377 support those of the original CATS-I study and a recent large RCT. Our data are
378 consistent with the lack of treatment effect being due to the similar proportion of IQ<85
379 in children of normal-GTF and SGTF mothers rather than the age of cognitive
380 assessment or the relatively high dose of thyroxine supplementation. However, future
381 large randomised trials, with thyroxine interventions at a much earlier stage of

pregnancy (or pre-conception), may still be warranted, since the benefits of treatment may not be fully realised unless treatment is commenced early.

Contributors

CH collected the data, was involved in writing the report and analysed the data with PNT. SC, RP, KM, LZ, MG, AB, OO, IM, MSD, JG, CD, JHL, and AR contributed to study design, data analyses and writing the report. ML designed and managed the project, supervised analyses and contributed to the report.

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.

References

1. Bernal J, Nunez J. Thyroid hormones and brain development. European Journal of Endocrinology 1995; 133:390-398

2. Ahmed OM, El-Gareib AW, El-bakry AM, Abd El-Tawab SM, Ahmed RG. Thyroid hormones states and brain development interactions. International Journal of Developmental Neuroscience 2008; 26:147-209

3. Zimmermann MB. Iodine deficiency. Endocrine Reviews 2009; 30:376-408

4. De Escobar GM. The role of thyroid hormone in fetal neurodevelopment. Journal of Pediatric Endocrinology and Metabolism 2001; 14:1453-1462

5. Contempré B, Jauniaux E, Calvo R, Jurkovic D, Campbell S, Morreale De Escobar G. Detection of thyroid hormones in human embryonic cavities during the first trimester of pregnancy. Journal of Clinical Endocrinology and Metabolism 1993; 77:1719-1722

6. Lazarus JH, Premawardhana LDKE. Screening for thyroid disease in pregnancy. Journal of Clinical Pathology 2005; 58:449-452

7. Rovet JF, Willoughby KA. Maternal thyroid function during pregnancy: Effects on the developing fetal brain. Maternal Influences on Fetal Neurodevelopment: Clinical and Research Aspects2010:55-77.

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. The New England journal of medicine 1999; 341:549-555

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. Clinical endocrinology 2010; 72:825-829

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. Journal of Clinical Endocrinology and Metabolism 2011; 96:3234-3241

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

12. Smit BJ, Kok JH, Vulsma T, Briet JM, Boer K, Wiersinga WM. Neurologic development of the newborn and young child in relation to maternal thyroid function. Acta Paediatrica, International Journal of Paediatrics 2000; 89:291-295

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. Journal of Clinical Endocrinology and Metabolism 2010; 95:4227-4234

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. Journal of Clinical Endocrinology and Metabolism 2014; 99:2383-2390

15. 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

16. 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

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 : official journal of the American Thyroid Association 2009; 19:511-519

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. Journal of Clinical Endocrinology & Metabolism 2012; 97:E22-28

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. Journal of Thyroid Research 2011; 2011

20. Grau G, Aguayo A, Vela A, Aniel-Quiroga A, Espada M, Miranda G, Martinez-Indart L, Martul P, Castaño L, Rica I. Normal intellectual development in children born from women with hypothyroxinemia during their pregnancy. Journal of Trace Elements in Medicine and Biology 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. The Journal of clinical endocrinology and metabolism 2009; 94:497-503

22. Willoughby KA, McAndrews MP, Rovet JF. Accuracy of episodic autobiographical memory in children with early thyroid hormone deficiency using a staged event. Developmental Cognitive Neuroscience 2014; 9:1-11

23. Willoughby KA, McAndrews MP, Rovet J. Effects of early thyroid hormone deficiency on children's autobiographical memory performance. Journal of the International Neuropsychological Society 2013; 19:419-429

24. Willoughby KA, McAndrews MP, Rovet JF. Effects of maternal hypothyroidism on offspring hippocampus and memory. Thyroid 2014; 24:576-584

25. Pharoah POD, Connolly KJ. Relationship between maternal thyroxine levels during pregnancy and memory function in childhood. Early Human Development 1991; 25:43-51

26. Ishaik G, Asztalos E, Perlman K, Newton S, Frisk V, Rovet J. Hypothyroxinemia of prematurity and infant neurodevelopment: A pilot study. Journal of Developmental and Behavioral Pediatrics 2000; 21:172-179

27. Henrichs J, Ghassabian A, Peeters RP, Tiemeier H. Maternal hypothyroxinemia and effects on cognitive functioning in childhood: How and why? Clinical endocrinology 2013; 79:152-162

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. The New England journal of medicine 2012; 366:493-501

29. 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

30. N M. IQ and human intelligence. Oxford University Press.

31. 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

32. 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

33. Daniels GH, Dayan CM. Thyroid Disorders. Oxford, UK: Health Press.

34. 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. Human reproduction update 2011; 17:605-619

35. Jouyandeh Z, Hasani-Ranjbar S, Qorbani M, Larijani B. Universal screening versus selective case-based screening for thyroid disorders in pregnancy. Endocrine 2014; 48:116-123

36. Reid Sally M, Middleton P, Cossich Mary C, Crowther Caroline A, Bain E. Interventions for clinical and subclinical hypothyroidism pre-pregnancy and during pregnancy. Cochrane Database of Systematic Reviews 2013(5). http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD007752.pub3/abstract.

37. Castagna MG DM, 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. Journal of Clinical Endocrinology and Metabolism 2017; 102:1623-1630

Legends to Figures and Tables

Table 1: Characteristics of the cohort. Data are expressed as median (Interquartile range, IQR), mean (standard deviation, SD) or the Number (N) of participants (percentage, %). Socioeconomic status is based on a social deprivation score with 1 being the most deprived. Child's language describes whether the child speaks English (Engl) both at home and in school, Welsh in both locations, a combination of English and Welsh or an additional language. GTF=gestational thyroid function; SGTF=suboptimal gestational thyroid function; CATS=controlled antenatal thyroid screening.

Table 2: Logistic regressions for odds of IQ below 85. Data are expressed as Odds Ratios (OR) with 95% confidence intervals (95% CI). SGTF=suboptimal gestational thyroid function. FSIQ=full scale intelligence quotient. VCIQ=verbal

comprehension intelligence quotient. PRIQ=perceptual reasoning intelligence quotient. WMIQ=working memory intelligence quotient. PSIQ=processing speed intelligence quotient. Model 1=unadjusted. Model 2=adjusted for child gender. Model 3=adjusted for model 2 and whether the mother breastfed over one month, and mother 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.

Table 3: Mean scores for IQs. Data expressed as means (standard deviations) of group, or the Number (N) of participants (percentage, %) having IQ<85. GTF=gestational thyroid function; SGTF=suboptimal gestational thyroid function; WISC=Wechsler intelligence scale for children fourth edition UK; VCIQ=verbal comprehension intelligence quotient, PRIQ=perceptual reasoning intelligence quotient; WMIS=working memory intelligence quotient; PSIQ=processing speed intelligence quotient; FSIQ=full scale intelligence quotient; NEPSY=developmental neuropsychological assessment second edition; MD=memory for designs; MDD=memory for designs delayed; FTDH=fingertip tapping dominant hand; FTNDH=fingertip tapping non-dominant hand; NM=narrative memory; LM=list memory. *reduced dataset

Figure 1: Flow-chart of recruitment to the Controlled Antenatal Thyroid Screening (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.

536 TABLE 1:

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 (mlU/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
Thyroxine 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/Socio-economic 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 one 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 centre	112 (48%)	27 (23%)	19 (19%)			
Child's language				0.950	0.364	0.541
English school/home	180 (78%)	95 (80%)	85 (87%)			
Welsh school/Engl home	42 (18%)	20 (17%)	11 (11%)			

Welsh school/home	7 (3%)	3 (2%)	1 (1%)
Engl school/other home	2 (1%)	1 (1%)	1 (1%)
Welsh school/other home	1 (1%)	0	0

537

538 Table 2:

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

539

540

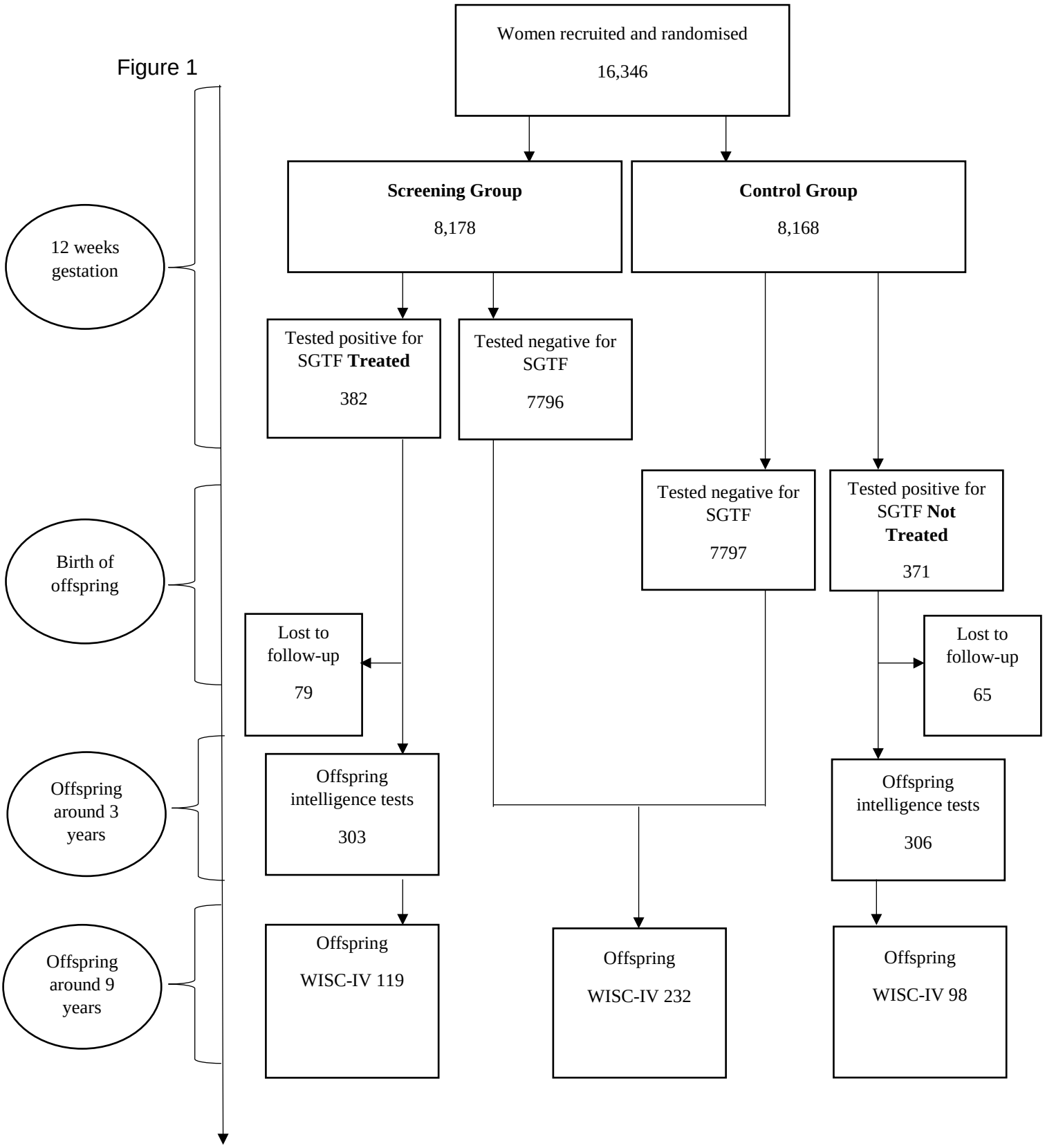
541 Table 3:

	Groups			
Cognitive assessment	Normal-GTF N=232	Merged SGTF N=217	Treated SGTF N=119	Untreated SGTF N=98
WISC				
VCIQ	99·81 (11·26)	98·60 (11·42)	97·56 (9·95)	99·86 (12·93)
<85	28 (12%)	30 (14%)	19 (16%)	11 (11%)
PRIQ	105·37 (12·30)	104·55 (12·87)	104·49 (12·26)	104·63
	11 (5%)	18 (8%)	7 (6%)	(13·64)
<85				11 (11%)
WMIQ	99·91 (11·24)	99·81 (12·72)	99·73 (13·28)	99·90 (12·07)
<85	18 (8%)	24 (11%)	14 (12%)	10 (10%)
PSIQ	103·66 (12·75)	102·39 (12·73)	103·16 (12·71)	101·45 (12·75)
	22 (9%)	18 (8%)	8 (7%)	10 (10%)
<85				
FSIQ	103·10 (11·68)	102·01 (12·59)	101·76 (12·04)	102·31
	15 (6%)	21 (10%)	10 (8%)	(13·28)
<85				11 (11%)
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)	11·06 (2·76)	11·02 (2·78)	11·12 (2·74)

	N=170	N=146	N=78	N=68
LM*	10·93 (2·84)	10·62 (2·86))	10·63 (3·13)	10·60 (2·54)

542

Figure 1


Women recruited and randomised
16,346
Screening Group
8,178
Control Group
8,168
12 weeks gestation
Tested positive for SGTF Treated
382
Tested negative for SGTF
7796
Tested negative for SGTF
7797
Tested positive for SGTF Not Treated
371
Birth of offspring
Lost to follow-up
79
Lost to follow-up
65
Offspring around 3 years
Offspring intelligence tests
303
Offspring intelligence tests
306
Offspring around 9 years
Offspring
WISC-IV 119
Offspring
WISC-IV 232
Offspring
WISC-IV 98