CARDIFF UNIVERSITY

# Cognition and Behaviour of Children Born to Mothers with an Underactive Thyroid; Data from the Controlled Antenatal Thyroid Screening Study

**Charlotte Emily Hales**

Thyroid Research Group

Centre for Endocrine and Diabetes Sciences

School of Medicine

**A dissertation submitted to Cardiff University in accordance with the requirements of the degree of Doctor of Philosophy (Ph.D.)**

**February 2016**

# Summary

## Background and aims

Underactive thyroid function during pregnancy and its effects on offspring intelligence, general cognition and behaviour have long been researched and reported on. Some of the differences found for the offspring are so apparent, that is has warranted authors to suggest universal thyroid function screening during pregnancy. The current study was the world's first randomised controlled trial to investigate the effects of treatment for suboptimal gestational thyroid function (SGTF) on offspring. The aims of this thesis are, 1) to re-analyse the intelligence scores for the offspring at age 3 years, 2) to clarify any SGTF effects by cognitive testing and behavioural questionnaires completed at offspring age 9, and 3) compare and contrast wave one and two findings.

## Methods and analysis

1) Data was previously collected for this analysis. Treated and untreated SGTF groups were compared by unadjusted and adjusted models. 2) I conducted the intelligence, additional cognitive testing, and collection of the completed behavioural questionnaires. All data were analysed separately (per chapter) by multivariate analysis models. 3) Comparison of intelligence results were explored by correlations and a repeated measures multivariate analysis.

## Results and conclusions

Re-analysis of the age 3 intelligence scores revealed that the untreated SGTF group performed worse compared to the treated SGTF group ($p = .008$ for scores below 85). No age 9 differences in intelligence or additional cognitive tests were found. The behavioural questionnaires revealed that treatment for SGTF may have had a detrimental effect for the offspring. Intelligence score comparisons revealed no differences between the groups. These results suggest that any intelligence effects from the mother not being treated for SGTF may be present at age 3 but have disappeared by age 9. However, treatment for SGTF appeared to significantly increase behaviour problems for offspring at age 9; though not clinically significant.

## Author's Declarations

**DECLARATION**

This work has not been submitted in substance for any other degree or award at this or any other university or place of learning, nor is being submitted concurrently in candidature for any degree or other award.


Signed ………………………… (candidate)     Date …………………………

**STATEMENT 1**

This thesis is being submitted in partial fulfilment of the requirements for the degree of PhD.

Signed ………………………… (candidate)     Date …………………………

**STATEMENT 2**

This thesis is the result of my own independent work/investigation, except where otherwise stated.

Other sources are acknowledged by explicit references.  The views expressed are my own.

Signed ………………………… (candidate)     Date …………………………

**STATEMENT 3**

I hereby give consent for my thesis, if accepted, to be available for photocopying and for inter-library loan pending submission of the study's cognition and behaviour publications. The title and summary will also be made available to outside organisations pending publication submissions.

Signed ………………………… (candidate)     Date …………………………

**STATEMENT 4: PREVIOUSLY APPROVED BAR ON ACCESS**

I hereby give consent for my thesis, if accepted, to be available for photocopying and for inter-library loans **after expiry of a bar on access previously approved by the Academic Standards & Quality Committee.**

Signed ………………………… (candidate)     Date …………………………

# Acknowledgments

Firstly, I would like to extend my sincere gratitude to my supervisors Prof. Marian Ludgate, Dr. Sue Channon and Dr. Kirsten McEwan, for their expert guidance, comments, and for keeping me positive throughout this degree. Their flexibility in thinking and working truly enabled me to achieve the highest possible participant numbers, whilst believing in me as a valued member of the research team.

I also wish to acknowledge Dr. Peter Taylor for his valued input, dedication and enthusiasm of the project.

Finally, I wish to thank my husband, Alex, for supporting me throughout all of my degrees; as always, I promise this will be the last.

For my Grandfather,


Alfons Komander

# Contents

# Figures

# Tables

## Glossary (abbreviations)

**ADHD**        Attention Deficit Hyperactivity Disorder

**ADI**         Autism Diagnostic Interview

**ADI-R**       Autism Diagnostic Interview Revised

**ANCOVA**      Analysis of Covariance.

**ANOVA**       Analysis of Variance

**ASCs**        Autism Spectrum Conditions

**ASQ**         Autism Screening Questionnaire

**CATS**        Controlled Antenatal Thyroid Screening

**CYP IAPT**    Children's and Young People's Improving Access to Psychological Therapies

**D2**          Deiodinase 2

**DSM**         Diagnostic and Statistical Manual of the American Psychiatric Association

**FS**          Full Scale

**FTDH**        Fingertip Tapping Dominant Hand

**FTNDH**       Fingertip Tapping Non-dominant Hand

**g**           General Intelligence

**gc**          Crystallised Intelligence

**gf**          Fluid Intelligence

**GSH**         Gestational Subclinical Hypothyroidism

**GTF**         Gestational Thyroid Function

**IQ**          Intelligent Quotient

**LM**          List Memory and List Memory Delayed

**LTM**         Long Term Memory

**MANCOVA**     Multivariate Analysis of Covariance

**MANOVA**      Multivariate Analysis of Variance

| | |
|---|---|
| **MD** | Memory for Designs |
| **MDD** | Memory for Designs Delayed |
| **NEPSY-II** | Developmental Neuropsychological Assessment, Second Edition |
| **NM** | Narrative Memory |
| **PR** | Perceptual Reasoning |
| **PS** | Processing Speed |
| **RCT** | Randomised Control Trial |
| **RTH** | Resistance to Thyroid Hormone |
| **SCQ** | Social Communication Questionnaire |
| **SD** | Standard Deviation |
| **SDQ** | Strengths and Difficulties Questionnaire |
| **SEN** | Special Educational Needs |
| **SGTF** | Suboptimal Gestational Thyroid Function |
| **T3** | Triiodothyronine |
| **T4** | Thyroxine |
| **THOP** | Transient Hypothyroxinaemia of Prematurity |
| **TPO** | Thyroid Peroxidase |
| **TPO-Ab** | Thyroid Peroxidase Antibodies |
| **TSH** | Thyroid Stimulating Hormone |
| **UK** | United Kingdom |
| **US** | United States |
| **VC** | Verbal Comprehension |
| **WISC** | Wechsler Intelligence Scale for Children, UK fourth edition |
| **WM** | Working Memory |
| **WPPSI** | Wechsler Preschool and Primary Scale of Intelligence, third edition |

**WPPSI-R**        Wechsler Preschool and Primary Scale of Intelligence-Revised

1. Controlled Antenatal Thyroid Screening Study II; cognitive and behavioural data

## 1.1.    General introduction

### 1.1.1.    Chapter Overview

The aim of this thesis is to add to the knowledge base of psychological effects for children born to mothers who have had an underactive thyroid during their pregnancies. The focus of this first chapter is to review the literature pertaining to maternal thyroid function and its impact on the offspring. Firstly there is a brief overview of thyroid function, how this can change during pregnancy and subsequently may become deficient. The potential effects to offspring intelligence are discussed, then memory, motor coordination, language and hearing difficulties, as well as behaviour and brain morphology. This literature review closes with a discussion of universal screening for treatment of an underactive thyroid during pregnancy.

### 1.1.2.    Thyroid Function

The thyroid gland is located in the neck, is often described as a butterfly shape, and forms two lobes connected by an isthmus. The thyroid is the main site for iodine uptake in the body (1). Thyroid peroxidase (TPO) (produced by the thyroid gland) incorporates iodine into two hormones, thyroxine (T4) and triiodothyronine (T3). These thyroid hormones are involved in the regulation of basal metabolic rate and macronutrient metabolism (2, 3). In the central nervous system, thyroid hormones also regulate cell migration, differentiation and myelination (4). T3 is the active form of the hormone and has a shorter half-life (one day) than T4 (5). T3 binds to three thyroid hormone receptors, one that is largely restricted to the pituitary, and the other two are widely distributed throughout the body. T4 is converted to T3 in most tissues by two enzymes; deiodinase type 1 and type 2; deiodinase type 3 converts T4 to the inactive form of T3. Specific transporters identified in carrying thyroid hormones across cell lines, are organic anion co-transporting polypeptide 1C1 for T4 across the blood-brain barrier, and mono-carboxylate transporter 8 for moving T3 into brain neurones (6). Free T4 is unbound thyroxine in the bloodstream which is available for uptake and use by cells, whereas T4 is also circulating in the bloodstream, but is bound to proteins by thyroxine-binding globulin. T3 and T4 are stored in the form of thyroglobulin and are released when thyroid stimulating hormone (TSH), from the pituitary, stimulates the thyroid. TSH maintains circulating levels of thyroid hormones, iodine uptake and thyroid growth (7) and its secretion is regulated by T4/T3 in a negative feedback loop. Replacement of thyroid hormones with T4 only, provides an individual with a long-lasting store of thyroid hormone that is gradually converted to T3 (5).

Thyroid dysfunction occurs when there is an excess (hyperthyroidism) or limited levels of thyroid hormones (hypothyroidism). This thesis is concerned with the latter only, and hyperthyroidism will not be discussed. Deficient thyroid function has a higher prevalence in females than males, with a proposed ratio of 4:1 (8). Chronic autoimmune thyroiditis is the most common cause of spontaneous hypothyroidism in iodine-sufficient regions (5). If TPO antibodies are present (Ab+), this might also inhibit the function of TPO and (9) and can be viewed as a 'precursor' for future underactive thyroid problems. Iodine deficiency is also a prevalent catalyst to thyroid dysfunction and is discussed below. Some hypothyroid symptoms include fatigue, weight gain, muscle cramps and/or joint pain. These symptoms can also be viewed by some as common characteristics to pregnancy, which is one of the reasons an underactive thyroid is often overlooked during pregnancy.

As mentioned, iodine deficiency can also be seen as a precursor for hypothyroid issues, as it is essential for the production of T4 and T3 (10). The general population of the United Kingdom (UK) is iodine deficient (11) (which may be due to diet), and an iodine deficiency disrupts the metabolism of thyroid hormones (10) as 70-80% of an individual's iodine is located in the thyroid gland (1). The recommended iodine intake for adults is 150-300 μg, hypothyroidism is a risk when iodine intake falls below 50 μg/day (5). When iodine supplies are severely inadequate, TSH increases as a compensation mechanism. With the cumulative demands of thyroid hormones during pregnancy (12, 13), iodine requirements also increase for the mother.

### 1.1.2.1.    *Thyroid function in pregnancy*

Thyroid dysfunction occurs in around 2.5% of pregnancies (14). T3 and T4 are essential for early brain development, and maternal thyroid hormones are required by the foetus, before it can produce its own (15-17). Thyroid physiology of the mother is altered during pregnancy due to an increase in thyroid-binding globulin and increased thyroid hormone requirements from the foetus (18). During normal pregnancy, the high-oestrogen environment causes the concentration of thyroxine-binding globulin to rise. As a result, total T4 and T3 concentrations increase in early pregnancy, peak in mid-pregnancy, and consequently remain elevated throughout the remainder of the pregnancy (5). At the beginning of second and third trimesters, T4 and T3 concentrations are 30-100% higher than before pregnancy (12, 13), as blood volume also increases. Inadequate supply of thyroid hormones from the mother could disrupt the brain development of the foetus (19). Mono-carboxylate transporter 8 is expressed in the placenta (6), and T3 uptake is mediated by L-type amino acid transporter 1 in the placental cell line (20). Severe maternal hypothyroidism during the

first two trimesters may result in irreversible neurological deficits, whereas later in pregnancy the foetus may be able to compensate for any lack of maternal thyroid hormones, but not achieving full function until term (21). The tiny gland first appears at the base of the tongue of the foetus at around 20 days gestation, and migration to the neck is complete by seven weeks (22). Thyroid hormones in the foetal brain are known to be of maternal origin (23, 24). At about 12 weeks gestation, the foetal thyroid begins to trap iodine and by 18-20 weeks it is working almost to full capacity (5). Rovet (22) illustrates proposed brain development timings for different neuropsychological aspects for the foetus (based upon previous work by Rovet (21, 25)). If thyroid hormones are disrupted during certain phases of gestation, it is concluded that these could then affect the proposed domains. Evidence of the importance of thyroid hormones for the foetus can be illustrated by congenital hypothyroidism, the foetus can intake the necessary hormones during gestation, but following birth it is reliant upon its own. Untreated neonates with congenital hypothyroidism display effects to growth, cognitive difficulties and language deficits (26). Early treatment for their hypothyroidism can suppress some of the effects, but mild disruptions to cognition may persist (27); new-born screening occurs for the baby at around 8-14 days old (28).

### 1.1.2.1.1. Iodine Deficiency

During pregnancy, the iodine requirement also increases because of the high renal iodide clearance by the mother and foetal thyroid requirement (5). Iodine is essential for neurodevelopment in utero (29). The World Health Organisation recommends an additional 50 µg of iodine supplementation for pregnant women (compared to adults), if suspected of being iodine deficient. In women with chronic iodine deficiency during pregnancy, their depleted iodine stores are not able to compensate for enlarged iodine demands leading to increased risk of maternal goitre and hypothyroidism (30).

In places with iodine-poor diets, cognitive deficits associated with maternal thyroid dysfunction in offspring have been studied for more than a century (31). It is recognised that a low maternal urinary iodine concentration of < 150 µg/L has led to poorer scores for verbal functioning for the offspring compared to those born to mothers whom had concentrations > 150 µg/L (32, 33), likewise deficits for neurobehavioural performance delays (34) and psychomotor development (35) have also been identified. However, there may be an inverted 'U' associated with treatment for iodine deficiency during pregnancy, as Murcia et al. (36) found that if mothers were over-supplemented with iodine, their children's scores were rated lower on Bayley Scales of development (also confirmed by (37)).

### 1.1.2.1.2.   Gestational subclinical hypothyroidism

Subclinical hypothyroidism, defined as an elevated level of TSH with normal circulating levels of T4 and T3 (38), affects 3-6% of the UK population (39, 40). Subclinical hypothyroidism is more prevalent in females, (8%) than males (3%), and cases diagnosed increase with age, 10-20% of women and 5-10% of men over 65 years old are reported to have an elevated TSH (5). Gestational subclinical hypothyroidism (GSH) in pregnancy is defined as a TSH concentration higher than the upper limit of the pregnancy related reference-range with normal T4 (and, if measured, normal T3). The upper limit of TSH is now defined as 2.5 mIU/l in the first trimester and 3 mIU/l in the second and third trimesters (41, 42). Subclinical hypothyroidism is a biochemical diagnosis as symptoms may be mild, non-specific and mimic typical symptoms occurring in pregnancy (43).

### 1.1.2.1.3.   Maternal Hypothyroxinaemia

Women with euthyroid function displaying a transient and mild decrease in T4 hormone levels during pregnancy without a rise in TSH (44), are reported to have maternal hypothyroxinaemia (45). Maternal hypothyroxinaemia has been identified in 4-10% of pregnant women (46-48). Though not as common as GSH, this lowering of T4 is attracting interest in recent studies, and evidence is mounting for adverse effects on the offspring (34, 49-53).

### 1.1.3.   Intelligence

There is some evidence that neuropsychological and intellectual development of offspring can be adversely affected by GSH (49, 54-57) or an iodine deficiency during pregnancy (10, 32, 58). The suggested mechanism for these effects of iodine deficiency and GSH is that although the brain is very dependent on thyroid hormones for normal development, active secretion of thyroid hormone in the foetus does not start until about 18-20 weeks gestation so the foetus is dependent on the mothers' circulating hormones for growth and development up until this point (59).

### 1.1.3.1.   What is 'intelligence'?

There is no widely accepted definition of intelligence. There is an identified link between how broad the definition should be to how broad the intelligence domain should be (60). Further to this, Boeck also adds (pp. 6):

> "…must a definition be based on research or should research be based on a definition?"

Intelligence is stated as requiring a flexibility in responding to challenging situations and actively shaping our environment (61). It is argued that intelligence could be referred to the

commonly coined '*g*' (for 'general-intelligence'), or could be described as being driven by domain-specific faculties (62-65). Intelligence has been defined as an individual's ability to understand and reason correctly with concepts and solve problems (66, 67).

Intelligence testing can be viewed as a desire to quantify intelligence, as assessment batteries attempt to measure this fluid concept. As Jensen (pp. 76) (68) stated, "Intelligence…is what intelligence tests measure". Intellect, as measured by Intelligence Quotient (IQ) tests, has been shown to predict a range of life outcomes such as academic performance, job performance, years in education, quality of life and even physical health (69-77). IQs measured on standardised tests are normally distributed with an average IQ score falling in the range of 90-109; with a percentile ranking of 50 for an IQ of 100 (78). Very young children are frequently assessed by development scales, and the most recent Bayley Scale of Development can assess from as young as 16 days (79). These scales are primarily used to identify any possible developmental delays for the child (80), and are not necessarily IQ measurements.

Genes are also acknowledged to play a role in an individual's intelligence; heritability accounts for around 50% of intelligence, therefore the environment is accepted to construct the rest of the variance (81). It has been identified that the environmental effects on intelligence are important in childhood, but are negligible in adulthood (82). There appears to be an overlap between gene interplay and cortical thickness that could influence intelligence (83). Different brain morphology is apparent in individuals with differing intelligence levels; those with superior intelligence levels show more intense and prolonged cortical thickening followed by more rapid thinning (84). This thickening and thinning has been suggested to occur in an extended sensitive period, during which the brain is responsive to environmental input (85). More recent cortical thickness studies support McGue (82), that heritability increases in childhood and adolescence, whilst environmental influences decrease in importance (86-89).

### 1.1.3.2. *Underactive thyroid function in pregnancy and the effect on offspring intelligence*

The number of studies investigating the impact of an untreated GSH on an offspring's IQ are growing, but the findings are equivocal. In one retrospective study, untreated GSH was shown to lower an offspring's IQ by a mean of 7 points (54), and of the 48 GSH offspring nine children had an IQ of < 85 compared to only six of the 124 matched control children (7-9 year old offspring assessed by a Wechsler Intelligence Scale for Children- third edition).

Interestingly, in this study by Haddow and colleagues there was a small group of women that were treated for their GSH (n = 14), but significant differences were only found between the untreated and matched controls (see Table 1 and Table 2 for further details).

Table 1
*Summary of Results from Haddow et al.*

| IQs (WISC-III) | *p* (treated GSH vs. untreated GSH) | *p* (Untreated GSH vs. matched controls) |
|---|---|---|
| Verbal IQ | .30 | .006 |
| Performance IQ | .30 | .01 |
| Full scale IQ | .20 | .005 |

Notes. Adapted from Haddow et al. (1999). Maternal thyroid deficiency during pregnancy and subsequent neuropsychological development of the child. The New England Journal of Medicine, 341 (8), pp. 549-555. Table extracted and edited from pp. 553.
IQ=intelligent quotient, WISC-III=Wechsler intelligence scale for children-third edition, GSH=gestational subclinical hypothyroidism.

Li et al. (49) detected this impact on IQ in children as young as 25- 30 months using Bayley Scales of development (compared to controls, mean intelligence scores were found to be significantly lower *p* = 0.008). However, even though Li et al. included a large sample (n = 1,268) with serum samples taken at 16-20 weeks gestation, there were only 18 mothers identified with GSH, with others being euthyroid. Klein et al. (90) found evidence for offspring cognitive deficits for increasing maternal TSH values measured at 17 weeks gestation (124 GSH mothers from a sample of 25,000 women). It was found that the higher the percentile ranking of TSH, the lower the IQ measurement would be for the offspring. Furthermore, IQs below 1 standard deviation (SD) were more frequent in children born to mothers with GSH compared to controls with gestational euthyroid function (*p* = .006). In a large population cohort in China, it has also been identified that GSH is associated with poorer neurodevelopment and also poor vision of offspring (91). Finally, Smit et al. (56) identified effects of GSH (identified during the first trimester) on offspring at ages 6 and 12 months measured by Bayley Scales, but no significance was achieved with the child at 2 years of age. Again, caution is advised as Smit et al.'s research was on a very small number of mother-child pairs consisting of 20 in total; specifically, only seven identified as having GSH and six with normal gestational euthyroid function. Henrichs et al. (53) included a much larger sample of women (n = 3,659) and also found no effect of GSH (as well as measuring for hypothyroxinaemia at 13 weeks gestation) for offspring language impairment, verbal and nonverbal functioning.

Maternal hypothyroxinaemia has also been linked to lower intelligence for the offspring. In a study by Ghassabian et al., hypothyroxinaemia was measured around 18 weeks, and was defined as T4 in the lowest 5th percentile of the cohort. IQ was then measured in the children at age 6 from the large sample (n = 3,727). It was found that nonverbal IQ was 4.3 points lower in the maternal hypothyroxinaemic children ($p$ = .001), compared to children born to mothers who had normal thyroid function (52). Henrichs et al. (53) confirmed similar findings and identified a nonverbal cognitive developmental delay for offspring born to such women. Higher percentile cut-off as a definition for maternal hypothyroxinaemia has also been found to be detrimental to the children. Suárez-Rodriguez et al. (51) had a cut-off for maternal hypothyroxinaemia at the 10th percentile (measured at 37 weeks gestation) and found using McCarthy Scales of Children Abilities when offspring were age 38 to 60 months, that the general cognitive index score was lower compared to controls ($p$ < .01). However, the research included a very small sample size of only 70 children being assessed. Also, Pop et al. (50) measured thyroid function at 12 weeks gestation and classified women as being hypothyroxinaemic when their T4 was below the 10th percentile (with normal TSH). The offspring were followed up at two stages, 1 year of age (63 case, 62 control) and 2 years of age (57 case, 58 control). Even with the small sample sizes, there were significant differences on measurements by Bayley Scales for mental processing and motor function (all $p$'s < .02). Li et al. (49) also found differences using Bayley Scales of Infant Development and found in their small sample (study details as above and in Table 2) that children born to hypothyroxinaemic mothers performed 9.30 points lower ($p$ = .004) compared to children born to euthyroid mothers during their pregnancies. The final study to discuss was by Berbel et al. (34), they had a sample (n = 345) of three groups of women, those with normal thyroid function, those classed as having maternal hypothyroxinaemia at 12-14 weeks gestation and those with the same classification at full-term. It was found that the offspring in the latter two groups compared to the normal group had lower cognitive function measured at 18 months old ($p$ < .05 and $p$ < .001).

As with GSH, the evidence is conflicting on the impact maternal hypothyroxinaemia has on the offspring's cognitive function. Su et al. (91) identified in their research that GSH had a detrimental cognitive effect for the offspring, but for those children born to mothers with hypothyroxinaemia during their pregnancies, no difference was found compared to maternal euthyroid controls. Bayley Scales were again a popular choice of assessment tool for the research as young children were being assessed. Craig et al. (92) assessed using Bayley Scales in a large sample (n = 5,734) and found that of the 198 children born to mothers with

gestational hypothyroxinemia, there were no significant differences to those born to euthyroid mothers ($p$ = .14). Chevrier et al. (93) assessed children (n = 287) with Bayley Scales at 6, 12 and 24 months of age and found no difference compared to offspring born to euthyroid women. The children in the study also underwent a Wechsler Preschool and Primary Scale of Intelligence (94) at 60 months of age and again, found no differences. Grau et al. (95) recently assessed neuropsychological outcome in offspring born to mothers measured to have hypothyroxinaemia (T4 below the 10th percentile) at the end of their first trimester. Children were assessed at two time points, 1 year of age (n = 455) and between 6-8 years (n = 289) with a Wechsler intelligence scale. No differences were found between the case children compared to the controls, furthermore there was no difference in intelligence when compared to free T4 in each trimester. Finally, Oken et al. (96) used a visual recognition memory test, which was stated to be as good a predictor of development as Bayley Scales, in children (n = 500) aged 6 months and 3 years of age and also found no difference between those born to mothers who were hypothyroxinaemic during their pregnancies and those with normal thyroid function.

So in summary, the evidence is conflicting as to whether an underactive thyroid during pregnancy could have a negative impact on a child's intelligence and cognition. There seems a great discontinuity between studies on a number of aspects: TSH and T4 classification cut-off points, when the mothers have their thyroid function measured during their pregnancy, the age of the offspring assessment and also the type of assessments all vary from study to study. The following sections explore the evidence concerning potential areas of specific impairments to the offspring; memory function, motor coordination, language delay, hearing ability and behavioural observations.

### 1.1.4.    Further cognitive effects

In addition to effects on intelligence, effects on specific areas of cognitive function for offspring born to mothers who had gestational deficient thyroids are also reported. As with the intelligence research, evidence was conflicting and studies tended to focus on either GSH or maternal hypothyroxinemia. Four domains are to be covered in relation to human studies (selected due to their evidence); memory, motor coordination, language delays and reading ability, and the offspring's hearing.

#### 1.1.4.1.    Memory

It has been proposed that memory is affected in a child born to a mother who has underactive thyroid function. Memory deficits are suggestive of possible hippocampal

damage in the brain (97, 98): the hippocampus has been shown to require an adequate supply of thyroid hormone during development (15) and insufficient exposure to thyroid hormones in utero has been shown to affect hippocampal structure and function (99, 100). It is suggested that the insufficiency will affect the transcription of specific thyroid hormone dependent genes resulting in hippocampal functional abnormalities (15, 27). Episodic autobiographical memory is the recall of past events using many different types of episodic details, such as emotions, who, what was present etc. (101).

The hippocampus was also implicated in episodic autobiographical memory retrieval (102), and this memory system has been evidenced to be impaired for those children born to mothers with GSH. Willoughby et al. (103) had a small group (n = 17) of children aged 10-14 years of age born to mothers whom were treated for GSH. Compared to the controls, the GSH children performed significantly worse and relayed less details of the event, including perceptual and place details. Also in an exploratory analysis, severe thyroid hormone deficiency in the third trimester was associated with lower proportion accuracy scores compared to the controls. This confirmed previous findings by Willoughby et al.'s (104) that controls performed better on autobiographical memory tasks. More recently in 2014, Willoughby et al. (105) identified that individuals (n = 24) born to mothers who had GSH and were treated for such, scored significantly lower on memory indices compared to euthyroid controls (n = 30). All of these studies used children born to mothers that were treated for GSH. It would have been interesting if there was a sample in all of the small studies that included children from mothers that were untreated.

Table 2

*Ten Summarised Studies Frequently Referred to Throughout This Thesis*

| Reference (Country) | Gestational age measurement | Participants / controls | T4 | TSH | Treatment | Offspring assessment age | Offspring assessment domain / tool | Blinded testing | Results (*p*) | Adjustments |
|---|---|---|---|---|---|---|---|---|---|---|
| Berbel et al. (34) (Spain) | 1) 4-6 wks & f/t 2) 12-14 wks & f/t 3) f/t | 1) 12 2) 19 / 13 | Free T4 Cases: 1) 0-10th percentile, (0.71-0.82 ng/dL) & > 20th percentile at f/t (>0.91 ng/dL) 2) 0-10th percentile Controls: > 20th percentile | Normal range for all groups: 0.38-4.80 µIU/mL | Iodine from time of consent for all groups, 200 microg KI per day | 18 months | Cognition, motor, language / Brunet-Lezine scale | Yes | Cognition: <.05 Motor: <.05 Language: >.05 | None |
| Craig et al. (92) (America) | 2nd-3rd trimester | 99 / 99 (matched) | Free T4 Cases: <3rd percentile (0.92 ng/dL). | Normal range for both groups: | None | 2 years | Cognition, motor, language / Bayley Scale of Infant | Yes | Unadjusted; cognition and motor: =.05 Language: | Gestational age, child age, ,maternal |

| | | | Controls: 10-90th percentile (1.00-1.34 ng/dL). | 0.26-3.34 mLU/liter | | | Development-III | | >.05. Adjusted: all >.05 | weight and education |
|---|---|---|---|---|---|---|---|---|---|---|
| Haddow et al. (54) (America) | 2nd-3rd trimester | 1) 48 2) 14 / 124 | T4 Cases: <7.75 µg/dL Controls: >threshold | Cases: >98th percentile Controls: <threshold | 1) None 2) Treated from time of consent with thyroid hormone. | 7-9 years | Intelligence / WISC-III. Language / Test of Language Development-II & The Peabody Individual Achievement Test-R. Motor / The Development Test of Visual-motor Integrations & The Grooved Pegboard | Yes | Intelligence: 1) <.05 2) >.05 Language: 1) <.05 2) >.05 Motor: 1) <.05 2) >.05 | None |
| Henrichs et al. (53) (Netherlands) | 13.3 weeks | Total: 3659 (cases not specified) | Free T4 Cases: <10th (<11.76 pmol/liter) and < 5th | Normal range for all groups: 0.03-2.5 mU/liter | None | A) 18 months (n=3411) B) 30 months (n=2819) | A) Language / McArthur Communicative Development Inventory | N/A | Both free T4 percentiles; Language and nonverbal | Maternal age & education, prenatal distress, prenatal smoking, |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | (<10.96 pmol/liter) percentiles. Controls: > 10th percentile (11-25 pmol/liter) | | | | B) Cognition / Parent Report of Children's Abilities (verbal and nonverbal) Language / Language Development Survey. | | cognition: <.05 TSH: >.05 | birth weight, gestational age at sampling, child ethnicity. |
| Klein et al. (90) (America) | 17 weeks | 1) 28 2) 20 / 124 (matched) | Not disclosed | Cases: 1) 98-99.85th percentile 2) > 99.85th percentile Controls: < 98th percentile | None | 8 years old | Intelligence /WISC-III | Not disclosed | 1) >.05 2) <.05 | Socioeconomic status, parental education and occupation |
| Li et al. (49) (China) | 16-20 weeks | 1) 18 2) 19 / 142 (matched) | T4 Cases: 1) 2.5th-97.5th percentile (101.79-218.49 nmol/l) 2) <2.5th percentile | 1) >97.5th (>4.21 mIU/l) 2) 2.5th-97.5th percentile (0.12-4.21 mIU/l) Controls: | None | 25-30 months old | Cognition & Motor / The Bayley Scale of Infant Development | Yes | Cognition and motor 1) & 2) <.05 | None |

| Study | Gestational age | Cases / Controls | Hormone & thresholds | Normal range / thresholds | | Age at follow-up | Outcome / Test | | Results | Confounders |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | (<101.79 nmol/I) Controls: 2.5th-97.5th percentile | 2.5th-97.5th percentile | | | | | | |
| Pop et al. (50) (Netherlands) | 12 weeks | 1) 63 / 62 (matched) 2) 57 / 58 (matched) | T4 Cases: <10th percentile (<12.14 nmol/I) Controls: >10th percentile | Normal range for groups: 0.15-2.0 mIU/I | None | 1) 1 year old 2) 2 years old | Cognition and motor / The Bayley Scale of Infant Development | Yes | 1) & 2) cognition & motor: <.05 | None |
| Saurez-Rogriguez et al. (51) (Spain) | 37 weeks | 37 / 33 (matched) | T4 Cases: <10th percentile. Controls: 9.5-23.9 pmol/I | Normal range for groups: 0.2-5 µU/mL | None | 3-5 years old | Cognition, memory, motor/ McArthuy Scales of Children's Abilities | N/A | Cognition & memory: <.05, motor: >.05 | None |
| Smit et al. (56) (Netherlands) | <20 weeks | 7 / 6 | T4 Cases and controls: if 1st trimester 7.4-24.2 pmol/I. If 2nd trimester | Cases: 1st trimester >2.0 µU/mL. 2nd trimester >2.3 µU/mL Controls: below thresholds | None | 6, 12 & 24 months | Cognition / The Bayley Scale of Infant development | Yes | 6 & 12 months, <.05. 24 months, >.05 | Maternal ethnicity and education. |

14

| Su et al. (91) (China) | <20 weeks | 1) 41 2) 43 / 845 | 5.1-14.3 pmol/l T4 Cases: 1) 5-95th percentile 2) <5th percentile Controls: 5-95th percentile | Cases: 1) >95th percentile 2) 5th-95th percentile Controls: 5th-95th percentile | None | A) before 6 months B) 42 days & 3 months | A) Cognition / The Bayley Scale of Infant Development B) Hearing / Auditory Brain Stem Response | Not disclosed | 1) Cognition: <.05, Hearing: >.05 2) Cognition & Hearing: >.05 | Maternal age and body mass index |
|---|---|---|---|---|---|---|---|---|---|---|

Note. T4=thyroxine, TSH=thyroid stimulating hormone, WISC-III=Wechsler intelligence scale for children third edition.

There are a few studies that have identified specific working (short term) memory deficits for the offspring; rather than those in the long term memory system, episodic and autobiographical. By using a test investigating a string of numbers, it was identified that children born to mothers who had hypothyroxinaemia during their pregnancies performed significantly worse compared to controls (106). Suárez-Rodriguez et al. (51) explored differences in their sample of children born to mothers with gestational hypothyroxinaemia and found that memory was recorded as significantly lower ($p < .01$) compared to controls (see Table 2 for further study details). Finally, an interesting study investigating T4 from the umbilical cord shortly after birth, identified that those whom had low levels of T4 performed better on the memory domain of the McCarthy Rating Scales at age 5 and a half years (n = 542) (107). This latter study is in direct contrast to most other research available, as it implies a positive affect for the child when the mother has an underactive thyroid during her pregnancy. A reason for this may be due to T4 being measured post-delivery from the umbilical cord, whereas other studies discussed in this chapter relate to serum samples collected during gestation.

### 1.1.4.2. Motor Coordination

Motor coordination of children born to women whom had an underactive thyroid during their pregnancies has also been identified in the literature. It was first identified in the 1960s (108) that 12-29 weeks of pregnancy could be a critical period for the visuospatial system including some aspects of motor coordination, and that these are affected by thyroid hormone insufficiency. Fine and gross motor skills appear to be sensitive to thyroid hormone after 16 weeks of gestation (90, 109) and declining levels of T4 in the third trimester result in poor motor skills of preterm babies (110).

GSH has been linked to a poorer motor coordination for the offspring. In Li et al.'s (49) study motor coordination was poorer in those born to mothers who had GSH compared to controls ($p < .001$) (see Table 2 for further study details). Haddow et al. (54) found that children born to mothers who were treated for their GSH, showed no significant difference between motor scores at ages 7-9 years ($p = .30$). However, when Haddow et al. compared the untreated to the matched controls, significance was reached ($p = .04$) with the untreated GSH group performing worse; caution is advised however, as significance was only reached for fine motor coordination of the non-dominant hand of the children (see Table 2 for further study details). Radetti et al. (111) conducted thyroid screening between the 8th and 10th gestational week (n = 691) and identified eight women with GSH who were treated swiftly following

diagnosis. The children were assessed for their psychomotor abilities at 9 months of age, but no differences between the case and control children were found.

Maternal hypothyroxinaemia has also been linked to reduced performance in psychomotor skills (112). Pop et al. (50) found that those born to mothers who were hypothyroxinaemic during their pregnancies performed worse for motor coordination compared to controls ($p$'s < .02, see Table 2 for further study details). Li et al.'s (49) study also included a subset of mothers who were hypothyroxinaemic, and confirmed Pop et al.'s findings ($p$ = .007, see Table 2). Pop et al. (113) assessed 220 children at 32 weeks of age by Bayley Scales and found that there were no differences to controls. The children were assessed again at the later time point of 10 months, and those born to mothers who had gestational hypothyroxinaemia performed worse on psychomotor measurements at this older age point.

Some of the pregnancy studies have revealed no significant effects of maternal hypothyroxinaemia on the offspring. Berbel et al. (34), Craig et al. (92) and Suárez-Rodriguez et al. (51) did not identify a difference for motor function between the case and controls in their studies (see Table 2).

### 1.1.4.3.    *Language Delay and Reading Ability*

A language and reading impairment may also be measureable in children born to mothers who had an underactive thyroid function during their pregnancies. Reading ability has been shown to be sensitive to thyroid hormone levels after 16 weeks gestation (90, 109).

Specifically, GSH has been shown to affect the offspring's language and reading ability. Henrichs et al. (53) found that those offspring born to mothers who had GSH were more likely to be at a higher risk for an expressive language delay compared to controls born to euthyroid mothers. Li et al. (49) confirmed findings and identified that the child's language capabilities would be affected by GSH. Furthermore, Haddow et al. (54) found that between the treated to untreated GSH children there were no differences for language ability ($p$ = .90), however when the untreated GSH were compared to control children, significance was achieved ($p$ = .02) with the untreated performing worse (see Table 2 for further study details of all three studies).

Maternal hypothyroxinaemia has also been linked to a language delay for the offspring (44), however most of the research suggests that there is no deficit for this domain. Berbel et al. (34) and Craig et al. (92) did not identify this link in their samples (see Table 2 for details). Noten et al. (114) recently assessed language in 5 year old children (n = 1,196) between those

born to mothers who had maternal hypothyroxinaemia, and euthyroid controls measured at around 13 weeks gestation, no differences were found between the groups (a 1.61 increased odds of poorer arithmetic was identified for the case group however). However, Li et al. (49) did identify a language deficit for children born to mothers with gestational hypothyroxinaemia compared to controls (see Table 2 for study details).

### 1.1.4.4.    Hearing Ability

There is little evidence that low thyroid function (especially during pregnancy) could affect the offspring's hearing. Cognition and hearing deficits have been shown to both occur in iodine-deficient areas (115-117). Furthermore, maternal hypothyroidism brought on in iodine-deficient areas has been linked to deaf-mutism and low cognitive function for the offspring (118, 119). Hearing difficulties have consistently been reported to be linked to individuals with congenital hypothyroidism (120-123).

As mentioned, TPO is an enzyme that plays a role in the production of thyroid hormones, if antibodies are present (Ab+), these could inhibit the function of TPO (9) and can be viewed as a 'precursor' for future underactive thyroid problems. Wasserman et al. (124) investigated TPO-Ab+ in mothers in their third trimester of pregnancy. The children were assessed using a Wechsler Intelligence Scale for Children at 7 years of age and their auditory levels were evaluated at age 8. The children's hearing was reported to be significantly worse in the case group compared to the controls, and these hearing difficulties were associated with their IQ levels (this hearing association was identified Wasserman et al. (31)).

Su et al. (91) identified a link between thyroid function during pregnancy and offspring hearing deficits, but this was only significant in the children born to women who had hyperthyroidism during their pregnancies (See Table 2 for details). Radetti et al. (111) had a small sample of children born to mothers who were treated for their GSH, and assessed at 9 months of age for audiology proficiency. No differences were found compared to the controls; furthermore, maternal T4 was identified as not being associated to audiological outcome for the offspring.

### 1.1.5.    Behavioural Observations

### 1.1.5.1.    Attention Deficit Hyperactivity Disorder

Attention Deficit Hyperactivity Disorder (ADHD) is characterised by symptoms of inattention, hyperactivity and impulsivity (125, 126) with prevalence rates ranging from 1-5.29% worldwide (127, 128), with an increasing trend (129). The cause of ADHD is still unclear and

it appears to be an accumulation of factors which contribute to an individual displaying ADHD (130). Twin studies have shown a high heritability for ADHD of around 71-90% (131-133). Adoption studies allow a focus on separation of the environment and genetics by investigating the degree of similarities between individuals with ADHD and their biologically related or adopted relatives; studies have found a genetic, inherited contribution to ADHD (134-137). There have been specific studies searching for candidate genes for ADHD (138-140), but no definitive genetic link has yet been identified (130, 141). However, we must be mindful that genetics and the environment are intertwined and cannot exclude environmental influences when discussing special educational needs (SEN) (142).

### 1.1.5.1.1.  ADHD and the Thyroid

It has long been evidenced that some individual's with ADHD may have a thyroid disruption of their own. Hauser at al. (143) found that in individual's with resistance to thyroid hormone (RTH) (defined as mutations in thyroid receptor ß-gene and characterised as a reduced responsiveness of the peripheral and pituitary tissues to the action of thyroid hormone), were significantly more likely to exhibit symptoms of ADHD compared to those without RTH (significant differences were applicable to adult and child groups $p < .001$). However, others have struggled to find a link between RTH and ADHD (144-146). Furthermore, there have also been identified links between an individual's lower concentrations of T4 and ADHD-predominantly Inattentive, but not ADHD Combined type (147).

There is conflicting evidence related to gestational thyroid function and the possible ADHD consequences for the offspring. Andersen et al. (19) assessed at the population level (n = 857,014) in Danish nationwide registers between 1991-2004 and identified maternal hypothyroidism posed no association to ADHD to the offspring. However, this included mothers who were treated for their underactive thyroid and no consistent T4 or TSH classifications were used, as the population was nationwide. Some studies with a strict GSH definition have also found no association. Modesto et al. (45) recently identified in a large population sample from the Netherlands that GSH measured at around the end of the first trimester (mean = 13.9 weeks gestation) was not associated with higher ADHD scores. Pakkilla et al. (148) identified that for female offspring of mothers with GSH, inattention and total ADHD symptoms increased with increases of maternal TSH concentrations; but the results were not replicated with boys. In contrast, Ghassabian et al. (44) reviewed findings from the Generation R study and identified that maternal GSH was related to ADHD symptoms in the offspring. Prior to this, Ghassabian et al. (149) investigated TPO-Ab+ during early pregnancy and found that, if present, there was an increased risk of externalising

problems in preschool children: in particular, ADHD problems. The significant effect remained when maternal TSH was controlled for, the authors concluding that TSH has an effect on offspring ADHD.

Further conflicting evidence was identified for individuals who experienced maternal hypothyroxinaemia. In Pakkilla et al.'s (148) study, no association to maternal low T4 (or TPO-Ab+) was identified to offspring ADHD symptoms. In the Generation R cohort, Ghassabian et al. (44) identified that for those born to mothers who had gestational hypothyroxinaemia there were no offspring behaviour difficulties at age 3 but by age 6, the children were significantly more likely to display ADHD symptom behaviours compared to controls ($p = .03$); however, this significance disappeared after controlling for maternal age and maternal education. Modesto et al. (45) found in their study that maternal hypothyroxinaemia was associated with higher ADHD symptom scores for the child at age 8 years compared to those born to normal GTF mothers. Furthermore, treatment for hypothyroxinaemia made little influence to the results. Mathew et al. (150) investigated maternal hypothyroxinaemia (measured at a mean of 13.6 weeks gestation) in 3,873 mother-child pairs from the Generation R study in the Netherlands. ADHD was assessed around age 8 years of age in the offspring by Conners Parent Rating Scale- revised short form, it was also identified that maternal hypothyroxinaemia was associated with higher scores for ADHD; these results remained unchanged when TPO-Ab were excluded. Finally, in iodine deficient areas, hypothyroxinaemic mothers during their pregnancies reported children with an abnormally high frequency of ADHD (151).

### 1.1.5.2.    Autism Spectrum Conditions

Autism Spectrum Conditions (ASCs) is an umbrella term used to describe Autism, Asperger's Syndrome and Pervasive Developmental Disorders Not Otherwise Specified. There is some evidence of prenatal and early-life 'critical periods' for possible susceptibility to developing ASCs and other neurodevelopmental disruptions (152, 153).

The most common traits of those with ASCs is some form of communication difficulty (154). There may also be an over dependence on routines, fixations on items and/or being highly sensitive to changes in a particular environment (see the Diagnostic and Statistical Manual of the American Psychiatric Association (DSM) fifth edition, 2013 (155) for further details). The terminology ASCs was preferred for this thesis, rather than the more commonly used 'Autism Spectrum Disorders', as it was less stigmatising and it reflects that these individuals may have cognitive strengths rather than just viewing a person with a 'disorder' which can

imply a negative label of a SEN (156). The prevalence in the UK for ASCs was last noted from the 2011 Census (157) to be 1.1% (n = 695,000).

There is much about ASCs that remains unknown with many questions still unanswered; why is the prevalence higher in boys than in girls? Why is this umbrella term so wide? Why is there no 'one size fits all model'? Are the reasons for this developmental disorder genetic? Numerous studies have focused on this latter question (158-163). The environment is now also coming under scrutiny as its links to ASCs are becoming clearer in the recent literature. Additionally, there is a growing body of evidence linking ASCs and non-genetic factors such as the environment, for example, proximity to air pollution (164). Moreover, there is evidence of maternal influences on the risk of ASCs with affects from prenatal maternal antidepressant use (165), maternal infections (166) and low periconceptional folic acid intake (167).

### 1.1.5.2.1. ASCs and the Thyroid

Without knowing the cause of autism, it is difficult to begin to understand the complex role the thyroid may play in increasing the risk of offspring developing ASCs or, the role the thyroid may play in individuals who already have a label of an ASC. As mentioned above, it is documented how thyroid hormones are critical for normal human brain development (168), additionally, how thyroid hormones can influence myelination and also gene expression (169), which could help begin to address this possible link between thyroid and the genetic basis identified for ASCs. Hoshiko et al. (170) identified low T4 levels in newborns to be associated with a high risk for ASCs, whereas Soldin et al. (171) found no association between neonatal thyroid hormone levels and ASCs; the dependency of thyroid hormones may then be before this, during pregnancy.

Few studies have investigated thyroid function during pregnancy and the possible impact it can have on offspring developing an ASC. As well as a small group of studies finding a link between ASCs (as below), there are also at least three studies where no link was identified. Firstly, a study published in 1980 suggested there is no link between thyroid hormones and diagnoses of ASCs (172). However, 35 years and multiple studies later, a systematic review and meta-analysis has been conducted that identified a positive association between maternal autoimmune diseases and risk of ASCs in the offspring; though no detail was covered in respect to classifications of the autoimmune conditions included (173). Croen et al. (174) with 407 participants with ASCs and 2,095 controls found that when maternal

autoimmune disorders were present around the time of their pregnancies, these were unlikely to contribute to the offspring developing an ASC.

Brown et al. (175) found that the prevalence of maternal TPO-Ab+ increased the odds of ASC in offspring by 80% ($p$ = .009) compared to mothers negative for this antibody. Adding validity to these findings is the large sample size of the study; 1,132 individuals with ASCs and 967 matched controls.

Román et al. (176) identified in their large study sample of 4,039, that severe maternal hypothyroxinaemia (measured at 13.4 weeks gestation) was significantly ($p$ = .001) associated with an almost four fold increase in odds of the child developing an ASC. Transient hypothyroxinaemia of prematurity (THOP) is defined as low thyroid hormone levels and normal TSH during a critical period of brain development in infants born prematurely (177). Korzeniewski et al. (178) identified that individual born with THOP were at a 2.5-fold greater risk of having an ASC. Whilst Román (179) postulates that the presence of THOP may increase the risk of ASCs in the offspring.

Whilst hypothyroxinaemia is linked to normal TSH levels, Yau et al. (180) identified a difference in maternal TSH levels being lower in their ASCs group (n = 78) compared to the gender and age matched controls (n = 149), though not significant ($p$ = .18). It was concluded that the higher the TSH levels the less chance there was of the child having an ASC. However, there were no measurements of T3 or T4; therefore it would not be possible to conclude that the mothers were hyperthyroid during their pregnancy. Further to this, TSH was only measured once during mid pregnancy, and if levels were high this has been associated to decreased odds of the offspring having an ASC; therefore low TSH was concluded to be associated ASCs for the offspring (180). However the study was unable to conclude if the mothers had GSH as no measurements of T3 or T4 were taken, also it was unclear whether the low TSH measurements were a product of high T3/T4 for the mothers. Andersen et al. (19) assessed at the population level (n = 857,014) and identified maternal hypothyroidism posed an increased risk for the offspring to be vulnerable to ASCs. However, this included mothers whom were treated for their underactive thyroid and no consistent T4 or TSH classifications were used; it was unfortunate that those with maternal thyroid deficiencies were not included, but as a retrospective study these individuals would have been difficult to include.

### 1.1.6. Brain Structure and Development

There is emerging evidence in a few studies that brain morphology may be affected by thyroid hormones. Children with congenital hypothyroidism have affected brain regions, for example one study has demonstrated areas of cortical thinning and thickening that were significantly different compared to controls (181). Furthermore, adolescents with congenital hypothyroidism have been shown to have an increased magnitude of hippocampal activation and bilateral activation compared to controls, and this was associated with the severity of hypothyroidism the individual experienced early in life (182). It is also well recognised that T4 and T3 are required for brain and neurological development of the foetus (17), and these hormones may be in insufficient supply during pregnancy.

Corpus callosum development occurs in the brain of the foetus during pregnancy (183-185), and it is vulnerable to early thyroid hormone deficiency (186). Samadi et al. (187) investigated a small sample of women and found that those born to hypothyroid mothers during their pregnancy (some were treated), had smaller anterior and larger posterior sub-regions of the corpus callosum compared to healthy controls. Significantly larger splenium and a smaller genu (at the trend level) was also observed. Samadi et al. concluded that 20% of variance in the genu size was explained by maternal TSH levels; suggesting mothers with a long period of TH insufficiency had smaller genua. One of the limitations of this study, aside from the small sample size, was that there was little detail about the treated underactive thyroid function of the mothers, and no comparison was made between those whom were treated or untreated for the duration of their pregnancies. However, building on the evidence of the importance of maternal TSH levels, Si et al. (188) identified children with 'abnormal brain results' were more likely to be born to mothers with high levels of TSH during their pregnancies, compared to those with a 'normal brain result'. It is difficult to discuss this research further as little detail about which brain regions were viewed as 'abnormal' or in-fact, what specific measurements of the brain were taken in the children.

Some studies have investigated the effect of GSH on offspring brain morphology. Willoughby et al. (103) investigated 68 children (ages 10-14 years), 17 of whom were born to mothers that were treated for GSH. It was found that the controls accurately relayed more details from a staged event; including more perceptual details and more accurate places than the case group. A reason for this, could have been that the children born to the treated GSH mothers had smaller hippocampal volumes, and this finding was specific to the right anterior hippocampal volumes; although the result was non-significant (103). In a second study, Willoughby et al. (105) studied the hippocampus again with a group of 54 children (30

controls and 24 born to mothers with GSH that were treated), the case children showed significantly smaller right and left hippocampal volumes compared to the controls. This was particularly evident in the right posterior and left anterior segments. Similar to the previous research (103), the children born to treated GSH mothers performed significantly lower than controls on memory indices, furthermore, these results correlated to the smaller hippocampal volumes identified (105).

Ghassabian et al. (52) investigated hypothyroxinaemia during pregnancy, and as mentioned above, measured IQ at age 6 in the offspring, but also conducted brain imaging scans on the children when they were around 8 years of age. The scans investigated specific brain volumes, cortical thickness and brain surface area. Even though a significant difference was identified for nonverbal IQ between those born to mothers whom were hypothyroxinaemic during their pregnancies, and those who had euthyroid function, no differences were found between any of the brain measurements. Conversely, T4 has been recognised as having an effect on offspring brain morphology. Korevaar et al. (37) investigated maternal thyroid function at 9-18 weeks gestation, and similarly to Ghassabian et al. (52) (as participants were drawn from the same cohort), IQ was measured at age 6 with brain morphology assessed (n = 646) at around 8 years of age. It was identified that an inverted 'U' was evident for the association of T4 to offspring nonverbal IQ ($p$ = .0044), grey matter ($p$ = .0062) and cortex volume ($p$ = .0011). The finding of the inverted 'U' implies that either a high or low maternal T4 measurement could have equally detrimental effects on brain development for the foetus (37). Furthermore, Korevaar et al. identified that TSH was not associated to brain morphology of the offspring, there was also no association between T4 to white matter, corpus callosum or hippocampal volume (conflicts to the above discussed literature (103, 105, 187)).

### 1.1.7. Should we screen for underactive thyroid function in pregnancy?

If there is a significant consequence of a pregnant mother having a deficient thyroid on the child's development, this could potentially be widespread. To test and treat for low thyroid function in pregancy is reasonably "low-cost" (189, 190), thus the argument of 'should we treat' is debated.

In response to the detrimental findings for the offspring, there are those who propose screening during pregnancy to help determine the circulating thyroid hormone levels in the mothers (189, 191-195). Miscarriages, premature births and gestational hypertension have all been related to overt and mild maternal hypothyroidism (196), and treatment has been

shown to reduce these complications (197). Women at a low risk for thyroid dysfunction during their pregnancy and who received treatment for such had a smaller chance of adverse pregnancy outcomes than matched women who did not receive treatment (197). Jouyandeh et al. (190) executed a systematic review and meta-analysis across 3 databases. Articles were identified that demonstrated universal screening would lead to less miscarriages and adverse pregnancy outcomes. The meta-analysis confirmed that case-finding screening, rather than universal screening missed around 49% of pregnant women who had a thyroid dysfunction. Reid et al. (198) also conducted a systematic review and concluded that the treatment of euthyroid women with TPO-Ab+ showed a reduction in preterm births and also a trend to reduced miscarriages with levothyroxine treatment.

Some suggest only performing thyroid function tests in pregnant women with a previous history of thyroid dysfunction and do not recommend universal screening (199). Furthermore, few have investigated the neuropsychological outcomes of the offspring, rather than obstetric outcomes. Thung et al. (189) investigated a treatment model of hypothyroidism in women based on the IQ outcome of the offspring. It was concluded that screening would be cost effective, as there was a low cost initially from thyroid screening tests and treatment. If IQs were to be improved, money would be saved from the "…large additional lifetime costs that [would be] incurred by individuals with neurodevelopmental impairment" (pp. 267).

Before the decision of universal screening can be made, the evidence needs to be more robust and based on longitudinal large-scale randomised trials including women with treated and untreated GSH and their offspring. The Controlled Antenatal Thyroid Screening (CATS) study (200) was the first randomised controlled trial (RCT) to investigate the effects of treatment of women with deficient thyroid function during their pregnancies on their offspring's neuropsychological abilities. Thyroid function was measured at a mean of 12 weeks gestation with IQ measurements at age 3 and 9 (discussed in this thesis), with the latter age also including additional cognitive assessments. More details of the study can be found in the following chapter.

As mentioned in the above literature, there are varying times when maternal thyroid function has been measured during pregnancy and some believe that thyroid testing should not only be at the start of pregnancy, but perhaps continue until the end of the second trimester (201). However, adverse effects on both mother and child appear more apparent if thyroid dysfunction occurs within the first trimester of pregnancy (202). It is proposed that

there needs to be consistency for reference ranges of TSH and T4 (201, 203), and that this should occur before the decision to make screening universal is taken or not.

### 1.1.8. Conclusions

Thyroid function during pregnancy has been discussed including definitions for GSH and maternal hypothyroxinaemia. There have been a wide variety of studies conducted that have assessed the offspring for any differences compared to those born to mothers who had normal thyroid function during their pregnancies. Throughout this chapter, the literature has been contradictory. For GSH, some intelligence differences were identified for the offspring (49, 54, 56, 90, 91), whilst others struggled to find a difference (53, 56). Similarly, for maternal hypothyroxinaemia, differences were reported (34, 49-53) and also not reported (91, 92, 95, 96). On a more general level of cognitive deficits from a gestational underactive thyroid function, memory difficulties have been identified (51, 103-106), as well as motor (49, 50, 54, 110, 112), language difficulties (49, 53) and hearing difficulties (91, 124); but there are conflicting studies reported in the literature (34, 51, 54, 92, 111, 113, 114). As well as these possible cognitive deficits, behaviour has also been debated in the literature. ADHD difficulties have been identified in those born to mothers with an underactive thyroid compared to those born to those with euthyroid function (44, 45, 143, 148-151), but no differences have been found (19, 45, 144-146, 148). ASCs are similar with differences being identified (19, 170, 175-180) and challenged (171, 172, 174) in the literature. It is difficult to draw conclusions as to why there was conflicting evidence from the studies, as the studies vary on a number of factors: where the sample was from, gestational age, age of offspring at testing, and also assessment conducted on the offspring. The aims formulated from this general introduction can be found at the end of the following chapter.

### 1.1.9. Chapter Summary

This chapter has outlined some of the key areas in the literature relating to thyroid function during pregnancy and the outcomes for the offspring. Specifically, the chapter has covered different types of underactive thyroid function during pregnancy, GSH and maternal hypothyroxinaemia and also briefly explored the literature on iodine deficiency and how this can affect a pregnant woman's thyroid function. The offspring outcomes are varied for such women, the main effect reported in the literature is on the child's intelligence. In addition there are further possible deficits experienced by the child such as effect to memory, motor coordination, language and reading difficulties, and hearing abilities. More recently, studies are emerging of brain morphological differences for children born to women with an underactive thyroid function during their pregnancies. Based on the research outcomes,

arguments were proposed for the treatment or non-treatment of such women. This chapter prepares the ground for the following chapter in which wave one of the controlled antenatal thyroid screening study is described and the results are re-analysed using the UK cohort.

## 1.2. Re-analysis of intelligence at age 3; UK CATS I cohort

### 1.2.1. Chapter Overview

This chapter presents the analysis of the UK cohort CATS I data that I have conducted. This analysis is the precursor of the CATS II analysis, the description of the CATS I study sets the scene for the CATS II data collection and the data demonstrates how these children performed at a younger age. The chapter contains an introduction which describes how participants were recruited into CATS I, the methodology of the first cognitive assessments of the offspring, the statistical analysis I have conducted, the results and the discussion of those results.

### 1.2.2. Introduction

Lazarus et al.'s (200) CATS I work was a benchmark study as it was the first large prospective RCT to investigate the impact of treatment for deficient thyroid function during pregnancy on intelligence of the offspring. The CATS I study contributed to the growing literature of possible effects of thyroid dysfunction on child intelligence, including studies which have found an effect (34, 49-57, 90, 91), and those which found no difference compared to controls (53, 56, 91-93, 95, 96).

Women were invited to take part in CATS I at their first hospital antenatal appointment. The median gestation at recruitment was 12 weeks and 3 days. Women were excluded if they were < 18 years old, had a gestational age of > 15 weeks and 6 days, had a twin pregnancy or a known thyroid disease. A total of 21,846 women were recruited (16,349 women in ten centres in the UK, 5,497 women in one centre in Turin, Italy). At recruitment, blood samples were taken from the participants for measurement of TSH and T4 and women were randomly assigned with the use of a computer-generated block design to either the screening or control group.

Screening group participants had serum samples assayed immediately for a thyroid function result. Women were classified as having suboptimal gestational thyroid function (SGTF) if their TSH concentration was above the 97.5$^{th}$ percentile of the cohort, the T4 below the 2.5$^{th}$ percentile, or both (based on international guidelines (41, 204, 205), and also the same as a large cohort study based in the Netherlands that started recruitment same year as CATS I; (37)). If a screen group participant had a positive result for SGTF, they were treated with levothyroxine at a starting dose of 150 µg per day (recommended amount), treatment was initiated at a median of 13 weeks and 3 days. These participants had their TSH and T4 checked 6 weeks after the start of the therapy, and at 30 weeks gestation; treatment

adjustments were made if required. The women in the screen group with a positive result for SGTF were advised to see their family doctor after delivery of their baby to determine whether levothyroxine therapy should be continued or not. Women in the control group had their bloods (taken at the same time as the screen women) assayed after delivery of their baby. If they had a positive serum result for SGTF were also advised to visit their family doctor to see whether treatment should be initiated or not.

The primary outcome for CATS I was an IQ measurement in the offspring. IQ was measured at a mean age of 3.2 years and only those participants who had a positive serum result were included (i.e. not offspring born to mothers who had a normal thyroid function during their pregnancy, the normal GTF group). The Wechsler Preschool and Primary Scale of Intelligence third edition, UK version (WPPSI-III) (206) was administered by two psychologists at the children's homes. The psychologists were unaware of whether the child was born to either a mother from the screen (treated) or the control (untreated) group. As reported by Lazarus et al., the mean IQ scores for the analysis were corrected to a score of 100. For the current chapter, IQs have been left in their 'uncorrected' form to ease comparisons in chapter 2.2. (IQ comparison between ages 3 and 9; children from the CATS sample). There was a non-significant difference between the groups ($p$ = .40) with the mean treated SGTF full scale IQ being 99.2 compared to the mean untreated SGTF full scale IQ of 100.0. To see whether there was an effect with more children scoring lower IQs, percentages of IQs below 1 SD were also calculated. The treated SGTF group had 12.1% of children scoring below 85 compared to the untreated SGTF group having 14.1% ($p$ = .39).

As these results were from CATS I, and this thesis discusses findings from the second wave of the project, I decided to re-calculate the CATS I findings. The purpose of this was three-fold, firstly it acted as a pilot data set for the statistical analysis I would use for this thesis. Secondly, as CATS II only recruited from the UK, I wanted to establish if there would be any major difference found from excluding the Italian sample. Finally, I was interested to re-run the data with the IQs in their 'uncorrected' form, i.e. not corrected to the mean of 100.

Based on the CATS I results of there being no significant IQ differences between the treated and untreated SGTF groups, the current hypothesis was to fail to reject the null hypothesis; i.e. there would be no difference between IQs of those offspring from the treated or untreated SGTF groups.

As reported by Lazarus et al. (200), between 2002 and 2006, a total of 21,846 women were recruited to the CATS I study (see Figure 1 for the participant flow chart). Within this cohort, 16,349 women were recruited from the UK and 609 mother and child pairs from the treated (n = 302) and untreated (n = 307) SGTF groups were revisited at a mean offspring age of 3.2 years for measurements of their IQ.

As stated, the WPPSI-III (206) generates a full scale IQ, verbal IQ and also a performance IQ. Similar to the CATS II IQ test, the verbal IQ was a measure of "acquired knowledge, verbal reasoning, and comprehension", whilst the performance IQ was "a measure of fluid reasoning, spatial processing, attentiveness to detail, and visual-motor integration" ((94) pp. 135-136). The WPPSI was first developed in 1967 (207), and was an extension to the Wechsler adult and child intelligence scales. The WPPSI-III was published in 2002 (94) with the UK standardised version, as used in CATS I, published in 2003 (206) (and the current fourth UK edition released in 2013 (208)). The WPPSI-III was standardised on a sample of 1,700 children divided into 9 age groups of 200 each (except the 7 years 0 to 7 years 3 months composed of 100 children) (206). There was high internal-consistency coefficients reported with all IQs $r \geq 0.93$ and test-retest coefficients were above $r = 0.86$ for verbal, performance and the full scale IQs (209). The WPPSI-III was also demonstrated to display good validity as it had correlations to the WPPSI-R and Wechsler Intelligence Scale for Children third edition of $r = 0.80$-$0.89$ (206).

The WPPSI-III could be used to assess children between the ages of $\geq$ 2 years 6 months to $\leq$ 7 years 3 months. For children aged between $\geq$ 2 years 6 months to $\leq$ 3 years 11 months a shortened version was administered as children were anticipated to have a shorter attention span at this age. For CATS I, the aim was to test children around the age of 3 years so only the shorted version (25-35 minutes) was used. As can be seen in Figure 2, the children in CATS I completed four subtests for the WPPSI-III. The 'block designs' and 'object assembly' subtests comprised the performance IQ score. For block design, children were required to reproduce patterns made from one or two coloured blocks from a stimulus book. Object assembly required the child to fit puzzle pieces together to form a meaningful whole. The verbal IQ score was made up of the scores from the sub-tests of 'information' and 'receptive vocabulary'. For information, the child had to either point to a picture, or verbally answer a brief question presented orally to them by the examiner; the pictures, questions and concepts were about commonplace objects and events. Receptive vocabulary required the child to look at a group of four images and point to the one that the examiner was orally

describing. The raw scores from the four tests were converted into scaled scores, which in turn were converted to the composite IQs. The verbal and performance IQs equally contributed to the full scale IQ.

Data for the current analysis was retrieved from the CATS I Excel documents as IQs were in their 'uncorrected' form (i.e. not adjusted by 5 points as used for the Lazarus et al. (200) publication).
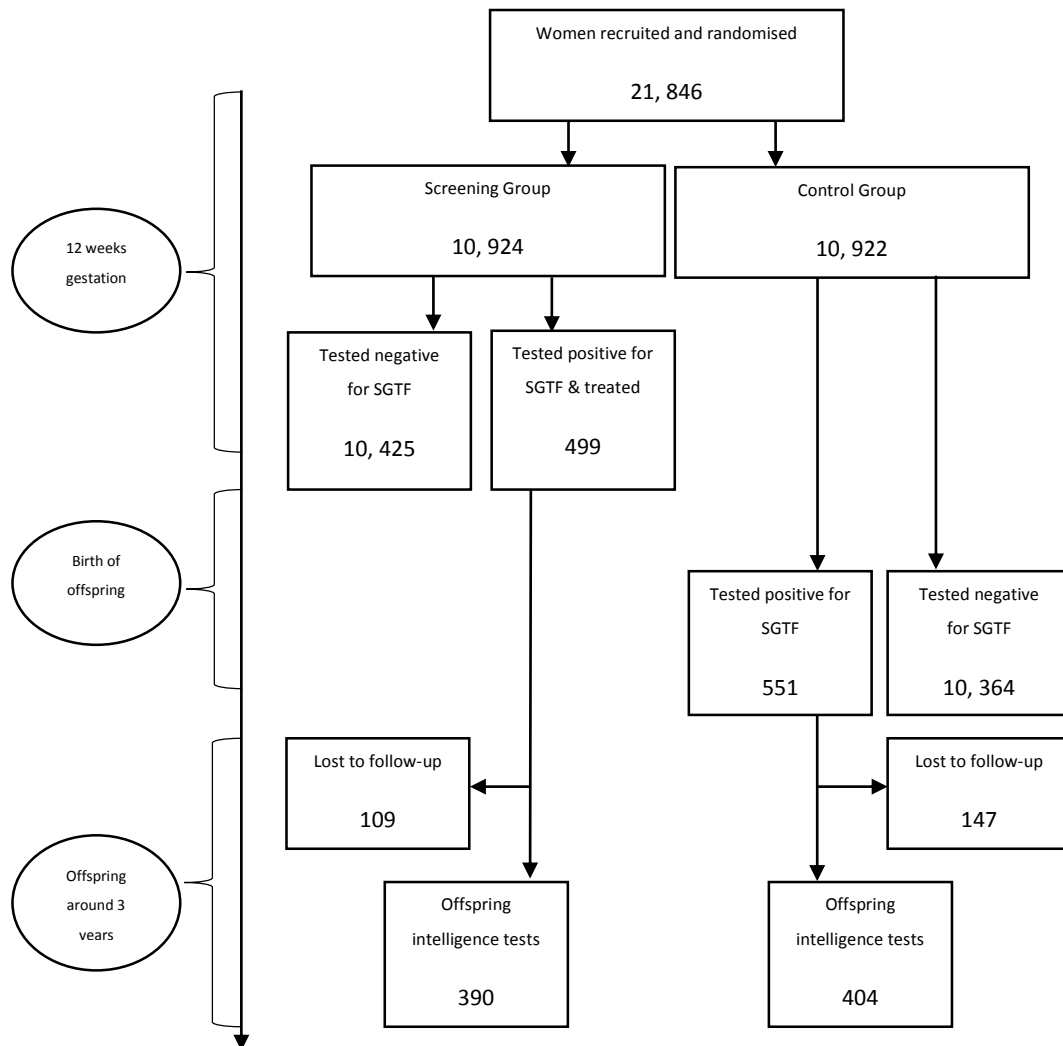


Figure 1: Randomisation and Follow-up of the Study Participants from CATS I

Adapted from Lazarus et al. (2012). Antenatal thyroid screening and childhood cognitive function. The New England Journal of Medicine, 366 (6), pp. 493-501. SGTF=suboptimal gestational thyroid function.

- Receptive Vocabulary
- Information

Verbal IQ

- Block Design
- Object Assembly

Performance IQ

Full Scale IQ
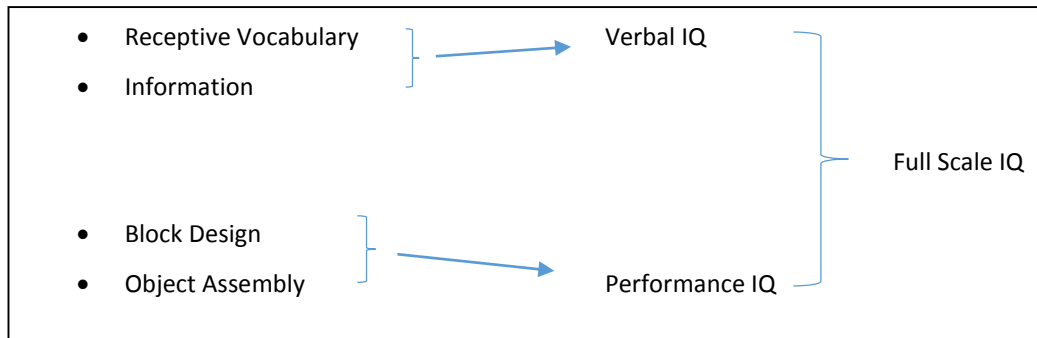
Figure 2: Wechsler Preschool and Primary Scale of Intelligence (WPPSI)-III, UK breakdown of subtests

IQ=intelligent quotient.

### 1.2.4. Statistical Analysis

The statistical analysis for this chapter was executed after the collection of all CATS II data to eliminate the potential bias from un-blinding, a possibility as the CATS I IQs were stored by a CATS ID number which was retained in CATS II.

The CATS I data was cleaned on a blinded dataset to avoid any further bias. No checks of nominal data for input errors was undertaken as all data was cleaned previously as part of the CATS I analysis. IQ cleaning occurred as it was an analysis of the UK only cohort, i.e. potential outliers from the UK cohort may have been 'masked' by the Turin, Italy IQ results. Descriptive statistics were used for the three IQs initially to check the ranges and for missing data. Z-scores were computed to identify any outliers in the data set. It was found for the verbal IQ that three individuals achieved z-scores < -3 and one had a z-score > 3. For performance IQ, only one participant had a z-score < -3, none achieved scores > 3. Finally, for the full scale IQ, two participants had z-scores < -3. Two participants had z-scores below the -3 threshold twice; these participants were removed as they could have skewed the dataset. See Figures 3 and 4 below to show how the histograms changed with the identified outliers removed. In total, there were 609 participants who had completed age 3 IQ assessments that were used for the re-run of the CATS I data analysis, UK cohort only.
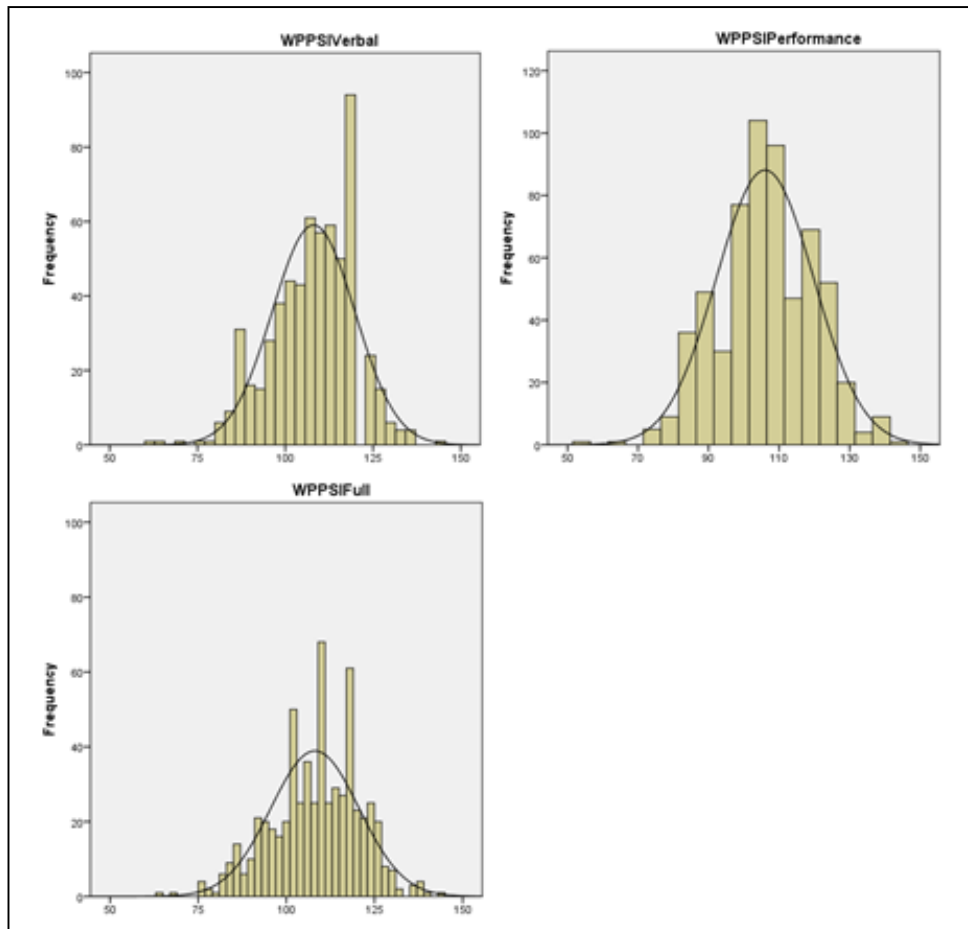
Figure 3: Verbal, performance and full scale Intelligent quotients (IQs) of complete dataset (n=609) of Wechsler Preschool and Primary Scale of Intelligence- third edition, UK (WPPSI-III) in the Controlled Antenatal Thyroid Screening study I

Y axis shows frequency, X axis, IQ scores.
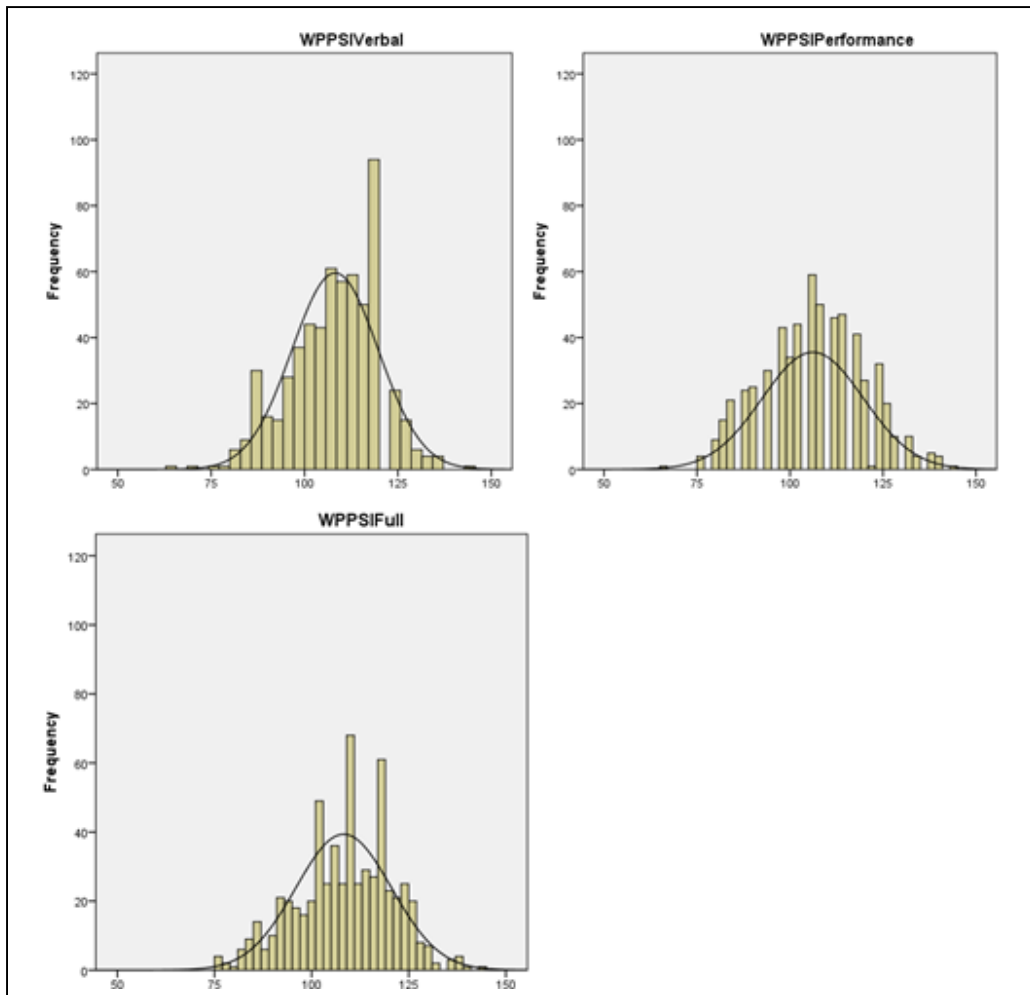
Figure 4: Verbal, performance and full scale Intelligent Quotients (IQs) of dataset (n=607) of Wechsler Preschool and Primary Scale of Intelligence- third edition, UK (WPPSI-III) in the Controlled Antenatal Thyroid Screening study I

Y axis shows frequency, X axis, IQ scores.

All Kolmogorov-Smirnov (and Shapiro-Wilk) normality tests were generated with all p's < .001 for the IQ measures. However, means and medians appeared close and skewness and kurtosis ranges were all within the -1 - +1 range and thus were normal (see Table 3 below). The IQ data were analysed by parametric tests as the variables were continuous, and based on the means, medians, skewness, kurtosis and histograms (see Figures 3 and 4) the data was accepted as being normally distributed.

Table 3

*Means, Medians, Skewness and Kurtosis for Verbal, Performance and Full Scale Intelligent Quotients (IQs)*

| IQ Domain | Mean | Median | Skewness | Kurtosis |
|---|---|---|---|---|
| WPPSI Verbal IQ | 108.18 (11.60) | 110.00 | -.354 | .161 |
| WPPSI Performance IQ | 106.14 (13.62) | 105.00 | -.025 | -.390 |
| WPPSI Full Scale IQ | 108.25 (12.30) | 109.00 | -.220 | -.203 |

Note. Standard deviations appear in parentheses below means.
WPPSI=Wechsler preschool and primary scale of intelligence- third edition, UK version.

The data was analysed using IBM SPSS Statistics version 20. The data collected by the two psychologists was amalgamated into one dataset and no analysis was conducted to investigate possible differences between the two datasets as this was investigated during the CATS I publication analysis, and no differences were found. For the current CATS I re-analysis, comparisons between the treated and untreated SGTF groups were executed by t-tests as it was continuous data. Similar to the CATS I publication (200), the data was analysed firstly in its unadjusted form, i.e. not controlled for any covariates. This was executed in the paper as the study was an RCT. To develop the work, and in keeping with this thesis, the second model of analysis included controlling for three covariates; child gender, mother age at time of consent into CATS I, and a measure of the participants' social deprivation (further details of how this was calculated can be found on page 58). Therefore, model two included a multivariate analysis of covariance (MANCOVA) to examine the variation of these confounders on the dependent variables.

The portion of children with a full scale IQ ≤ 85 (1 SD below the mean of 100 (206)) was also re-assessed by use of chi-square; this was also analysed firstly with the data in its 'uncorrected' form, and secondly by controlling for covariates by a multinomial logistic regression. It was predicted that the screening, treated SGTF group would contain 5% (n = 15) with IQ ≤ 85 and the control, untreated SGTF group would contain 15% (n = 46) with IQ ≤ 85. This prediction was based on the results of Haddow et al.'s work (54), however, for the published CATS I findings, Lazarus et al. (200) adopted Haddow et al.'s percentage calculation of children born to normal thyroid function mothers, for the treated SGTF group. These two analysis designs (t-test and chi-square) were adopted here. The power to detect a difference

in full scale IQ from the UK cohort sample was greater than 95% at the 5% significance level (two-sided test) (200).

Regression models were also used in CATS I to assess the risk of lower IQ based on SGTF classification. As CATS I analysed by TSH and T4 measurements during pregnancy, this allowed this type of analysis; however, I adopted group coding and therefore omitted this specific re-analysis. The CATS I study was an RCT, so covariates did not need to be controlled for (randomisation meant that conditions should have contained participants with similar characteristics), therefore there was only one model of analysis adopted for the publication's regression; unadjusted.

Further analyses were conducted as exploratory investigations of overt hypothyroidism, subclinical hypothyroidism, and maternal hypothyroxinaemia (see appendix 9). These were conducted to explore whether the broad definition of SGTF was potentially 'masking' any significant results or effects of, for example, low maternal T4 with a normal TSH.

### 1.2.5. Results

#### 1.2.5.1. General attendance information

As mentioned in Lazarus et al. (200), there was around a 20% drop-out rate from time of pregnancy for children from the SGTF groups completing the WPPSI-III. In total, 607 (300 treated SGTF and 307 untreated SGTF) children's assessments were included in the current analysis.

#### 1.2.5.2. Analysis

Below in Table 4, are the adjusted and unadjusted group means and SDs for the verbal, performance and full scale IQs. The graph (Figure 5) displays the means achieved by the groups pictorially (unadjusted model); error bars have also been included.

Table 4

*Intelligent Quotient (IQ) Means for Wechsler Preschool and Primary Scale of Intelligence (WPPSI)-III, UK at Age 3*

|  | CATS GROUP | Unadjusted data | | Adjusted Data | |
|---|---|---|---|---|---|
|  |  | N | Mean | N | Mean |
| WPPSI Verbal IQ | Treated SGTF | 300 | 108.68 (10.96) | 300 | 108.68 (10.96) |
|  | Untreated SGTF | 307 | 107.72 (12.20) | 306* | 107.75 (12.21) |
| WPPSI Performance IQ | Treated SGTF | 300 | 106.35 (13.48) | 300 | 106.35 (13.48) |
|  | Untreated SGTF | 307 | 105.94 (13.78) | 306 | 105.96 (13.80) |
| WPPSI Full scale IQ | Treated SGTF | 300 | 108.64 (11.76) | 300 | 108.64 (11.76) |
|  | Untreated SGTF | 307 | 107.88 (12.82) | 306 | 107.91 (12.83) |

Note. *n has dropped due to a participant who resided in Northern Ireland: no social deprivation score available. Standard deviations appear in parentheses below means.
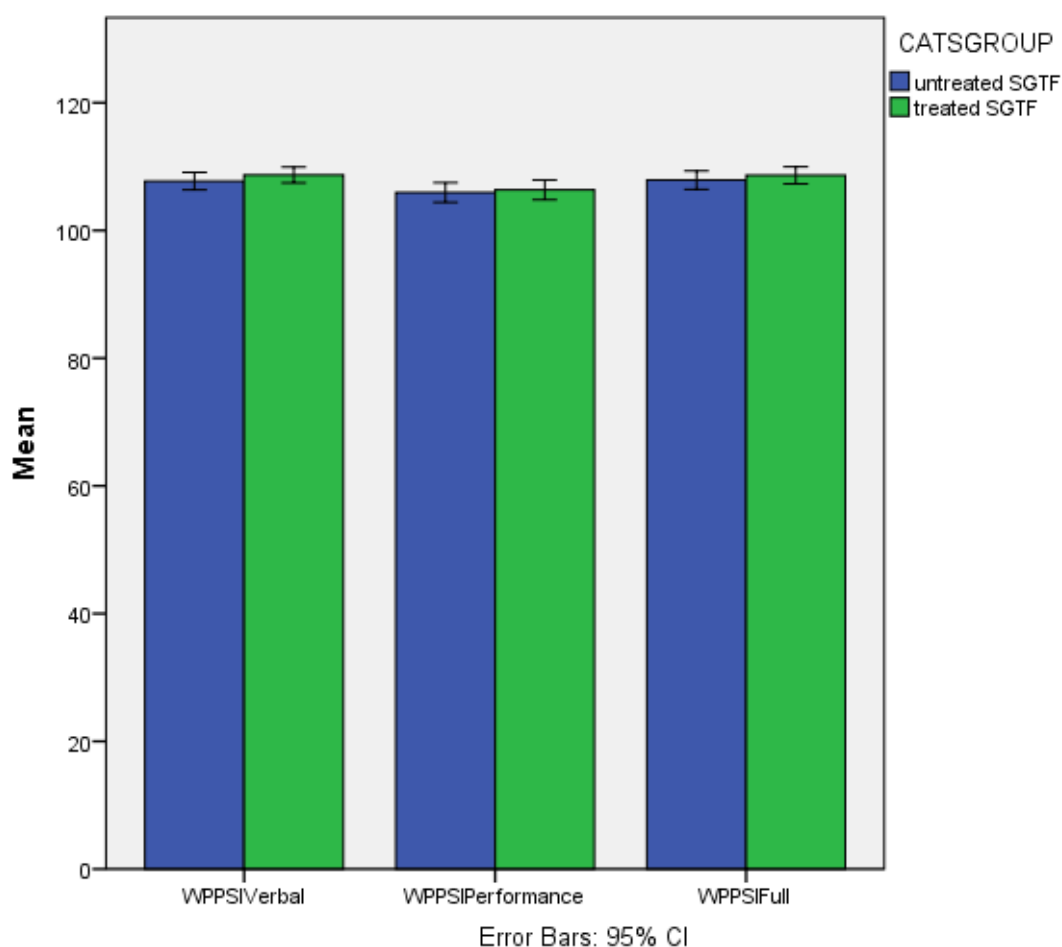


Figure 5: Means per group achieved on the Wechsler Preschool and Primary Scale of Intelligence (WPPSI)-III, UK

SGTF=suboptimal gestational thyroid function, CI=confidence interval. X axis=WPPSI domain, Y axis=mean of scores per group.

For the unadjusted data comparison, on average participants from the treated group had higher IQs than those from the untreated SGTF group (as can be seen in Table 4 above). The difference for verbal IQ was 0.954, 95% CI [-.897, 2.804], and was not significant $t(605) = 1.012$, $p = .312$. The difference for performance IQ was 0.409, 95% CI [-1.765, 2.582], and was not significant $t(605) = .369$, $p = .712$. Finally the difference for full scale IQ was 0.761, 95% CI [-1.201, 2.722], and was also not significant $t(605) = .761$, $p = .447$.

When the data was controlled for covariates (child gender, mother age at time of consent into CATS I, and a measure of the participants social deprivation), the participants from the treated group still had higher IQs than those from the untreated SGTF group (as can be seen in Table 4 above). The MANCOVA also yielded non-significant results, using Roy's largest root (most powerful multivariate statistic (210)), $\Lambda_{ROY} = .002$, $F (3, 599) = .492$, $p = .688$, $\eta_p^2 = .002$. As the multivariate analysis was non-significant, no further investigations of univariate effects were completed.

Percentages of IQ scores $\leq 85$ (unadjusted) were calculated and compared using Pearson Chi-square significance test. Table 5 below shows the significance values, and that full scale IQ was the only IQ that was significantly different between the treated and the untreated SGTF groups ($p = .008$).

Table 5
*Percentage of Intelligent Quotients (IQ) Falling Below 1 Standard Deviation ($\leq 85$)*

|  | IQs $\leq 85$ (%) | | |
|---|---|---|---|
|  | Verbal IQ | Performance IQ | Full scale IQ |
| Treated SGTF | 7 | 25 | 7 |
| (n = 300) | (2%) | (8%) | (2%) |
| Untreated SGTF | 12 | 25 | 21 |
| (n = 307) | (4%) | (8%) | (7%) |
| Pearson Chi-Square | $p$=.265 | $p$=.932 | $p$=.008* |

Note. Scores were from the Wechsler Preschool and Primary Scale of Intelligence (WPPSI)-III, UK and Intelligent Quotients (IQs) from the Controlled Antenatal Thyroid Screening study I UK only cohort. Percentages of scores per group are appear in parentheses below totals. *Significance < .05. SGTF=suboptimal gestational thyroid function.

The significant statistic of full scale IQ between the groups was explored further by adjustments for the covariates child gender, mother age at time of consent into CATS I and social deprivation score. The multinomial logistic regression revealed that children born to

mothers who were not treated for SGTF during their pregnancies were 3.335 times more likely to have a full scale IQ ≤ 85 at age 3 compared to those born to mothers who were treated for SGTF. The regression also revealed that as mother age increased and social deprivation score improved, the chance of achieving an IQ ≤ 85 decreased. It also was apparent that females were less likely to have an IQ ≤ 85 compared to males. See Tables 6 and 7 for further details.

Table 6

Table Displaying the Regression Model's Fit for the Data

| Model | Model Fitting Criteria | Likelihood Ratio Tests | | |
|---|---|---|---|---|
| | -2 Log Likelihood | Chi-Square | *df* | Sig. |
| Intercept Only | 115.821 | | | |
| Final | 86.924 | 28.897 | 4 | .000 |

Note. See improved figure for -2 Log Likelihood.
Df=degrees of freedom.

Table 7

*Main Output from Multinomial Logistic Regression, Full Scale Intelligent Quotient (IQ) ≤ 85*

| full scale IQ ≤ 85 | *B* | Std. Error | Wald | *df* | *Sig.* | Exp(*B*) | 95% Confidence Interval for Exp(*B*) | |
|---|---|---|---|---|---|---|---|---|
| | | | | | | | Lower | Upper |
| Gender | -1.063 | .454 | 5.482 | 1 | .019* | .345 | .142 | .841 |
| Mother age | -.433 | .219 | 3.900 | 1 | .048* | .648 | .422 | .997 |
| Social deprivation | -.348 | .145 | 5.770 | 1 | .016* | .706 | .531 | .938 |
| [Untreated SGTF] | 1.204 | .453 | 7.070 | 1 | .008* | 3.335 | 1.372 | 8.103 |
| [Treated SGTF] | 0 | . | . | 0 | . | . | . | . |

Note. The reference category was full scale Intelligent Quotient (IQ) ≥ 85. *Significance < .05. SGTF=suboptimal gestational thyroid function, B=beta, df=degrees of freedom.

### 1.2.6. Discussion

Based on the CATS I findings, the working hypothesis for this section was that the results would fail to reject the null hypothesis and that there would be no differences between IQ at age 3 of those offspring from either the treated or untreated SGTF groups (UK cohort only). Even though the treated SGTF group performed better than the untreated group for all three IQ scores, the results did not reach significance. Based on the findings for verbal IQ *p* = .312, performance IQ *p* = .712 and full scale IQ *p* = .447, the null hypothesis was not rejected.

Secondly, I investigated whether the treated SGTF group would contain the expected 5% with IQs ≤ 85 and the untreated SGTF would have 15%. For verbal IQ the treated SGTF group had 2% ≤ 85 whilst the untreated SGTF children had 4%. For performance IQ, the treated and untreated SGTF groups both had 8% of individuals with ≤ 85 scores. Finally, for full scale IQ the treated SGTF group had 2% ≤ 85 compared to 7% from the untreated SGTF group. These findings were different to Lazarus et al.'s (200) results for full scale IQ, 12.1% treated SGTF to 14.1% from the untreated SGTF group, as the IQs had been corrected to average around 100 rather than their true results of between 106-109. It was concluded that the 12.1% identified in the complete CATS I cohort (200) was higher than expected as treatment may have been initiated too late (discussed further in 'limitations' page 42). The full scale IQ differences were significant by a Chi-square test ($p$ = .008), when adjusted in a multinomial regression, significance was sustained and it was identified that the untreated SGTF group were 3.335 times more likely to score a full scale IQ ≤ 85 at age 3. However, only three covariates were controlled for and there may have been more effects to have made adjustments for. Furthermore, IQ is a continuous measure, and thus comparisons of means are preferred as binary calculations can simplify the statistical analysis as well as underestimate the extent of variability in a sample (211, 212). It has also been reported that changing continuous data to binary has an increased risk of generating a type one error (213), which may have occurred here.

The results in the current chapter were not surprising as they supported the CATS I findings. CATS I only tested those from the treated and untreated SGTF groups at age 3 of offspring. As CATS I was the first study to investigate the effects of treatment for SGTF, comparing it to studies of euthyroid mothers was important. If there was no difference to IQ of offspring born to a mother who was treated or untreated for SGTF, then would there be any difference for IQ compared to offspring born to a euthyroid mother? Furthermore, if there *was* no difference in childhood IQ from being born to a mother with SGTF compared to normal GTF then, why would treatment have made a difference?

Gestational underactive thyroid function has also been reported to have no effect on neuropsychological outcomes for the offspring. As stated, CATS I included women with either T4 in the lowest 2.5th percentile, TSH in the highest 2.5th percentile, or both. This meant including women with SGTF that were maternally hypothyroxinaemic (low T4), had GSH (high TSH) or had maternal overt hypothyroidism (combination of both). As mentioned, studies have highlighted no cognitive detriments to the child when born to a mother who had GSH (53, 56, 107, 214). Interestingly, Smit et al. (56) identified effects of GSH on offspring at ages

6 and 12 months, but no significance was seen with the child at 2 years of age. This was confirmed by Henrichs et al. (53) who identified no difference between offspring aged 18 and 30 months for language impairment who were born to euthyroid mothers and those that had GSH. Similarly, there is evidence that children born to mothers who had maternal hypothyroxinaemia do not have any cognitive deficits when compared to those born to euthyroid mothers (91-96, 215).

However, there have been articles where a lower thyroid function in pregnancy has been reported to have a detrimental cognitive affect for the child, for GSH (49, 54, 56, 90, 91), and also maternal hypothyroxinaemia (34, 49-53, 57). As CATS I did not assess any of the children born to mothers who had normal GTF, this was important to explore in the second wave of the study as there could have been a difference between the normal GTF and untreated SGTF children; alike to the studies cited here.

There may have been some issues with the WPPSI-III measure itself and reasons as to why the mean IQs were so high follows. To help aid recruitment for the cognitive assessments, children were visited in their homes for the WPPSI-III. This could have affected the IQs scored by the children, because their home environment could have been filled with distractions which are discussed as having an adverse effect for cognitive assessments (216). Within the previous research around thyroid function and childhood intelligence testing, some did assess within the home environment (such as Pop et al. (50)), whereas others did not mention the testing environment (for example Haddow et al. (54)). However, in CATS I distractibility from assessing in the home environment was not an issue as mean scores were higher than anticipated. The WPPSI-III results could have been subject to the 'Flynn Effect'; this is a phenomenon of increasing IQ scores over the years; i.e. as time goes on, people appear to perform better on such tests (217-223). The WPPSI-III used in CATS I was published in 2003 (206), and the assessments were conducted by two psychologists between 2006 and 2010. It would be unlikely that the Flynn Effect would have occurred in the earlier assessments, but it was possible it could have had an effect on the latter assessments. There was also the possibility that those more able to help research studies participated, i.e. more able to give up their time, to comfortably welcome a test examiner into their home, and therefore may possibly have a better social background which could also explain the high mean IQs.

From Table 4 above, it is clear that for verbal, performance and full scale IQs, the treated group achieved higher IQ results compared to the untreated SGTF group. If the IQs were

rounded up or down to full numbers, the verbal and full scale IQs for the treated SGTF group were one point higher; which was not clinically significant, nor a statistically significant result. It would be interesting to see whether with larger groups, this difference may have been bigger, or in-fact statistically significant.

Appendix 9 contains supplementary analyses exploring maternal overt hypothyroidism, subclinical hypothyroidism, and hypothyroxinaemia in the CATS I UK only cohort. No differences were identified between the treated and untreated SGTF groups, concluding that, within this cohort, treatment with levothyroxine was of no benefit to these offspring at age 3 years.

### 1.2.6.1.    Limitations

One of the limitations was that the treatment during pregnancy may have been initiated too late. The CATS I cohort had a median of 13 weeks and 3 days for blood samples from the mothers. Lazarus et al. (200) executed an exploratory analysis to see whether using the mothers who were recruited earlier during their pregnancies would yield different results, but it was non-significant; possibly as the groups were small and thus, underpowered to see any affects (200). One of the reasons for samples being extracted around the end of the first trimester was that women were recruited into CATS I by first appointment at the antenatal clinic in a hospital. To have enabled the study to recruit from earlier in pregnancy would have meant perhaps recruiting women from the first visit to their GP to confirm the pregnancy.

A further limitation that was highlighted by Lazarus et al. (200), was that the IQ testing at age 3 may not have been that reliable and IQ tests in older children have been found to be more accurate (224-226). Specifically, childhood IQ testing at ages 5-12 years generates a good stability into later adult life, thus are perceived as more of a true reflection of an individual's capabilities (86). Furthermore (as discussed below in chapter 1.4., Intelligence measured at age 9; CATS II data), IQ test outcomes are dependent on a number of examinee factors such as motivation, shyness, and rapport with the examiner (227-229).

The final critique would be that the CATS I study recruited women as having an SGTF if their TSH was above the 97.5[th] percentile, if their T4 was below the 2.5[th] percentile, or both (only consisting of 5% of both SGTF groups). This mixture of differing underactive thyroid functions may have clouded the results. Lazarus et al. (200) did investigate this further as an exploratory analysis. Non-significant results were found between the six groups (TSH above the 97.5[th] percentile, T4 below the 2.5[th] percentile or a combination of both, and then

whether the mothers were treated or not). It was argued that the groups were too small however and thus underpowered to find any statistically significant results.

### 1.2.6.2. Conclusions

The current analysis indicated that within the UK only CATS I cohort, there was no significant difference between the mean IQs of children from either the treated SGTF or untreated SGTF mothers at age 3 years. The CATS I published findings, when including the Turin sample, found $p = .40$ compared to the identified $p = .447$ for full scale IQ in the current chapter. This indicated that the Italian sample did not make any difference to the reported findings between the treated and untreated SGTF groups. Conversely, when the IQs $\leq 85$ are considered, for full scale IQ I identified that there was a significant difference between the UK only cohort of CATS I, but the published findings including the Italian sample were non-significant. A regression revealed that those children from the untreated SGTF group were 3.335 times more likely than those from the treated SGTF group, to obtain a full scale IQ $\leq$ 85.

As discussed, one of the major limitations of CATS I was that the children were young, thus a speedy intelligence test was desirable. Also, none of the offspring from the normal GTF were examined. CATS II secured funding to revisit children from both of the SGTF groups, and also some from the normal GTF group. As the children were older (around 7-10 years), a more in-depth intelligence battery was administered as well as the option to explore further cognitive domains. With older children, a measurement of the child's behaviour was also attainable, and the study also collected vast physical measurements (not discussed in this thesis).

### 1.2.7. Thesis aims of section 1: CATS II data

Based on the literature reviewed and the re-analysis of the CATS I findings, the following research aims were developed:

i. By reassessing the children at the older age of 9, would there be a continued non-significant difference identified for intelligence.

ii. Would there also be non-significant findings at age 9 in other potential areas of cognition.

iii. As there were no differences between the treated and untreated SGTF groups at age 3, would any differences to the normal GTF group be measurable; as there is a wealth of studies displaying that an underactive thyroid during pregnancy does not affect a child's cognition (53, 56, 91-93, 95, 96).

iv. Would there also be non-significant differences between the groups that extend beyond cognition, i.e. behaviour.

For the data I collected for CATS II, specific hypotheses grounded in theory were also developed and can be found in the respective results chapters; for intelligence see page 62, for additional cognitive domains see page 85 and for the behaviour see page 103.

### 1.2.8. Chapter Summary

This chapter has covered the CATS I UK only cohort analysis; it contained a study overview and in-depth description of how mothers were recruited into the project during their pregnancies. A brief methodology of the WPPSI-III measure was included in the methods section, as well as how the WPPSI-III was administered. The statistical analysis conducted for this re-run of the data was discussed followed by the main t-test results. The chapter concluded with the discussion of the non-significant results between the treated and untreated SGTF groups. The aim of this chapter was to aid the reader to grasp a better understanding of the project, and how CATS I had developed before I began the data collection in CATS II. The following chapter reviews the protocol for the cognitive and the behavioural data collected for CATS II and discusses the analysis of my data that I adopted for this thesis.

## 1.3. Methods for the cognitive and behavioural data collection for the CATS II study

### 1.3.1. Chapter Overview

The current chapter identifies the methods and methodology that were undertaken for the data collection for CATS II and, in-turn, this thesis. Before I began work on CATS II, the IQ test was pre-selected for data collection, in addition to the child behaviour questionnaires. Some of the questionnaires had already been administered to mothers who participated before my employment. The specific assessment battery used to investigate further cognitive domains was suggested as I began my journey on the project, however I did have direct input into which tests were to be selected from the battery.

CATS II was the follow-on study from CATS I, a large multi-centre RCT that aimed to investigate the possible long term effects of exposure to SGTF (see Appendix 1: Timeline of the CATS project). In CATS II cognitive assessments were administered to children aged between 7 and 10 years to ascertain their overall development. The study aimed to recruit a total of 480 participants from August 2011 over 4 years, and to complete the cognitive assessments in the latter 28 months of the project. Participants were seen either at the research centre (a clinical environment) or visited at their homes. If the participants preferred they had the choice to provide a reduced data set using a postal pack. The study was approved by the Wales Research Ethics Committee 2 and Caldicott Guardian.

I disaggregated this thesis from the CATS II project by adopting a different statistical analysis plan; briefly described in this chapter and more detail can be found in the following relevant results chapters. The different analyses were adopted to illustrate how this body of work was separate from the planned CATS II analysis. The results of the three groups of participants were analysed (normal GTF, treated and untreated SGTF) with no further investigations into pregnancy thyroid function to focus this work on psychological outcomes.

### 1.3.2. Population, eligibility and sample size

Participants were included if they were involved in CATS I and originally recruited in the UK. The CATS I sample did include a subset of participants from Turin, Italy (n = 5,497); these were not included in CATS II. As mentioned CATS II began recruiting before the cognitive assessments commenced as physical aspects of the mother-child pairs were investigated in the UK only. As funds were secured later for the cognitive assessments, it was logistically easier to keep to the UK only and not revisit those from Italy. Potential participants were

approached for CATS II when the child involved in the study was ≥7 years 0 months to ≤ 10 years 11 months. Participants were excluded from the study if they had moved overseas.

With a 5% two-sided significance level and 90% power, a sample of 120 from both treated and untreated SGTF groups would have allowed a detection of a difference of 6 points in mean IQ (assuming mean IQ to be 100 with a SD of 15 (78)) for a statistically significant result. In addition 240 participants (1.5%) from the normal GTF group were randomly selected from the UK cohort of CATS II; 15,744. These participants were used to assess whether there was an interaction with maternal thyroid status on offspring IQ as a baseline comparison. Therefore, three groups were re-visited at offspring age 9; normal GTF, treated and untreated SGTF.

### 1.3.3. Measures

To be able to address the aims of CATS II, an IQ test (a) and additional cognitive tests (b) were administered to the children in the study. Questionnaires (c) were also completed by the mothers to quantify if any behavioural problems were evident in the children.

a) The Wechsler Intelligence Scale for Children- Fourth Edition UK (WISC-IV). The WISC-IV (78) provided subtest and composite scores that represented intellectual functioning across specific cognitive domains: Verbal Comprehension (VC); Perceptual Reasoning (PR); Working Memory (WM); and Processing Speed (PS). The IQs generated from these areas equally contributed to the Full Scale (FS) IQ.
b) Developmental NEuroPSYchological Assessment- Second Edition (NEPSY-II). Possible delays in long term memory (LTM), WM and fine-motor skills were investigated by selected subtests from the NEPSY-II (230): List Memory and List Memory Delayed (combined score) (LTM), Memory for Designs (WM), Memory for Designs Delayed (LTM), Fingertip Tapping (motor) and Narrative Memory (WM).
c) Questionnaires:
   i. Strengths and Difficulties Questionnaire (SDQ) (231)
   ii. Child ADHD Questionnaire (modified by Thapar et al. (232))
   iii. Social Communication Questionnaire (SCQ) (233)
   iv. General CATS questionnaire (administered to gather demographic information)

A brief overview of the study procedure follows the discussion of each measure, more detail regarding the administration of the specific measures can be found in the following three results chapters (pages 62-104). As I was the only examiner and sole person scoring the questionnaires for CATS II, there was good consistency; training was completed prior to cognitive test administration. A randomly selected 10% of completed cognitive assessments were double scored to ensure accuracy by an educational psychologist on the research team. Stability was ensured by reviewing means fortnightly to check no scaled or composite scores

were well above or below the average which may have indicated skewed testing. The cognitive tests were administered in a set order with standardised verbal instructions given to each child. The WISC-IV was administered first, followed by the NEPSY-II. All behavioural questionnaires were completed by the child's mother.

*a)* *1.3.3.1. WISC-IV*

The WISC-IV used for the data collection of CATS II, was the UK standardised fourth edition (78). As mentioned in the General introduction (chapter 1.1.), it was reported in the literature that a deficit to a child's intelligence may be measurable if the mother had SGTF (34, 49-57, 90, 91). The WISC-IV was adopted as the Wechsler scales are the most widely used tests of intelligence in the world (234-238). Furthermore, the WISC-IV generated a FSIQ as a representation of an individual's general intelligence functioning, as well as yielding four domain scores (VCIQ, PRIQ, WMIQ and PSIQ). Thus if there were any specific deficits, the assessment could be interrogated to investigate where these might be.

## 1.3.3.1.1. Development of the WISC-IV

As well as the Wechsler scales being the most widely used intelligence batteries, they have led development and research into intelligence for more than 50 years (239). Conversely, WISCs have been criticised for lacking a firm grounding in theory (234, 240) with studies scrutinising their higher order structure (240, 241).

The WISC-IV began its development over 60 years prior to its use in the CATS II research. The very first of the Wechsler tests was called the Wechsler-Bellevue Intelligence Scale (242), with Form-II released in 1946 (243). Soon after, the first Wechsler intelligence scale for children (ages 5-15 years) was published (244), and subsequently revised in 1974 (245). The third edition was published 17 years later (246), the fourth edition (current research) in 2003 (78) and the most recent, fifth edition, in 2014 (247). Versions II, III, IV and V assessed children within the range of 6 years 0 months to 16 years 11 months.

The WISC-IV included more items than its predecessor to enable it to test lower functioning in younger children; ceilings had also been improved on certain subtests (240). It contained 15 tests, with ten constituting the core battery of the assessment. These ten core subtests were selected for the CATS II data collection and none of the additional five tests were used, to help aid continuity between participants. The WISC-IV was the first to measure on four factor indices: VC, PR, WM and PS. Around 60% of the core subtests from the WISC-IV were new or revised from the third edition (241). The WISC-V has been published in the United States (US), but at time of writing this thesis is awaiting UK standardisation. The fifth edition

is a major revision of the fourth (248), which now includes a larger five-factor model where PR has been divided to measure visual spatial and fluid reasoning as two separate entities (similar to what was suggested by Kaufman et al. (234) in their review of the WISC-IV). Furthermore, the FSIQ from the WISC-V can be calculated by seven core subtests instead of the ten used in CATS II, which is a positive aspect with regard to testing time, but could invite criticism for reliability (i.e. if a child performs badly on one subtest, it will have a higher impact on the FSIQ (249)).

One of the reasons the WISCs are popular intelligence measures, is that all of the editions have undergone extensive norming and standardisation. The WISC-IV used in the current research, was standardised from a stratified sample of 2,200 individuals: 200 children per year of birth within the specified WISC-IV age range, with equal numbers of males and females (78). For the UK standardisation (250); 800 British children in 110 schools were examined within the space of six months concluding that some of the language had to be altered for the UK version. Norming the WISC-IV would have been complex as the battery assessed a wide age range of children with differing levels of ability. Norming was also important to ensure that older children would not get frustrated by answering questions that were too simple, likewise that younger children would begin a subtest at an appropriate level.

### 1.3.3.1.2. Reliability and Validity

The reliability of a test refers to the accuracy, consistency and stability of the test scores across multiple situations (251). Test-retest stability was assessed using a sample of 243 children with 18-27 participants comprising each of the 11 age groups. The WISC-IV was administered twice with test intervals ranging from 13-63 days (mean = 32 days), it was found that there was adequate stability for all age groups (correlations between tests at time one and time two were high $r$ = .80) (78), which demonstrated good consistency of the scores.

The validity of the WISC-IV invited scrutiny due to the two new radical composite scores (WM and PS) and introduction of five new subtests. The three new subtests on the PR domain were added to enhance the measurement of fluid reasoning, with the additional two subtests assisting measurements of the WM and PS domains (78). A number of inter-correlational studies were conducted and displayed that the WISC-IV correlated well to other child Wechsler scales ($r$ = .89, see Kaufman for further details (234)), including its previous version.

*b)* *1.3.3.2. NEPSY-II*

Being the most common IQ test (234-238), the WISC-IV was simple to select for the CATS II study. Deciding on which *additional* tests to administer was grounded in theory (see chapter 1.1., General introduction). As the child cognitive assessments were conducted in one session, it was not feasible to test all possible domain deficits. Based on the literature, the most frequent deficits investigated were memory and motor coordination. Therefore I helped guide the decision process to assess across these domains.

Subtests from the NEPSY–II (230) were selected and administered in an identical manner. The NEPSY-II was chosen for the measure of the additional tests because this assessment battery was commonly adopted in the UK and contained the widest variety of subtests to choose from on a domain level. As well as suggesting the cognitive domains to assess (see Appendix 2: The decision process for additional tests for CATS II), I had direct input into which subtests were selected from the NEPSY-II. One of the NEPSY-II subtests, list memory and list memory delayed (LM), I suggested adding latterly to testing; this was to aid any evidence found for a LTM deficit.

### 1.3.3.2.1. Development of the NEPSY

Korkamn et al. (252) argued the uniqueness of the NEPSY as being solely based on childhood assessment, unlike other neuropsychological measures that appeared as 'add-ons' to adult cognitive scales. Ahmad and Warriner (253) discussed the four main purposes for the development of the NEPSY:

1. To provide an instrument to detect deficits that interfere with learning.
2. To offer a tool for identifying and assessing brain damage and dysfunction and to measure the extent to which they may affect operations and development.
3. To provide researchers and clinicians with a tool for long-term follow-up.
4. An instrument that would deliver an assessment that was standardised, reliable and valid for investigating normal and atypical neuropsychological assessment.

The NEPSY was developed in 1998 by Korkman et al. (252) and originally only assessed children between the ages of 3 and 12. Prior to this, Korkman published a version in Finish for children aged between 5 and 6 years, and the 1998 version followed for a wider age group in English. The first edition NEPSY was used to assess across five domains: attention/executive function, language, visual-spatial processing, sensorimotor and memory and language. The second edition (used in the current research) additionally included the domain of social perception (230): this would aid assessments for individuals suspected of having an ASC.

One of the reasons the NEPSY was so popular was because of the vast norming and standardisation completed. For the first and second editions, 100 children of each age group and of equal gender were used to make a complete cohort of 1000; for the second edition, these were split further by age, half born in the first six months of the year, and the other half in the latter (230). The second edition included many subtests from the first edition. However, some of these were not re-normed during the standardisation phase (including LM among others) (254). The rationale given for this was that these particular tests were not expected to be subject to the Flynn effect; however, there was no empirical basis to support this decision (255).

### 1.3.3.2.2. Reliability and Validity

The reliability of the NEPSY-II was tested and it was identified that fingertip tapping dominant hand (FTDH), fingertip tapping non-dominant hand (FTNDH) and LM were amongst the subtests with the highest reliability coefficient, and the lowest were found on some subtests including memory for designs (MD) spatial and total scores, and memory for designs delayed (MDD) total score (230). It was expected that the lowest reliability tests would mainly belong to the memory and learning domain as they would have been influenced by practice effects. However, some subtests were not included in the reliability testing, and no rationale was provided as to why this would be or why some stability estimates were reused (255).

The validity was assessed in the NEPSY-II by a number of correlational studies. Validity of intellectual functioning was assessed using the WISC-IV (78), Differential Abilities Scales-Second Edition (256) and the Wechsler Nonverbal Scale of Ability (257). Correlations between these scales demonstrated the NEPSY-II to be predictive of cognitive performance for both verbal and nonverbal domains (comparisons between the NEPSY-II and WISC-IV are evaluated further in chapter 2.4., Memory score comparisons; data from the WISC-IV and NEPSY-II assessments). Nine further scales were adopted to test other aspects from the NEPSY-II (see Korkman et al. (230) for details).

### c) 1.3.3.3. Questionnaires

Based on the literature cited in the General introduction (chapter 1.1.), it was apparent that possible deficits from SGTF may stretch beyond the cognition of the child and could also affect different behavioural aspects (19, 44, 45, 143, 148-151, 170, 175-180). The SDQ was adopted for CATS II as it could report across multiple domains. As discussed on pages 19-22, there was some evidence that children born to mothers with SGTF may be more vulnerable to ADHD and ASC symptoms than those born to euthyroid mothers. As a consequence the

Child ADHD Questionnaire (232) and the SCQ (233) were used, selected as measures being widely used, familiar to the research team and for ease of administration to minimise the burden on the mother (e.g. the SCQ takes approximately ten minutes to answer and less than five minutes to score (258)).

The behavioural questionnaires were given to the mothers of the children in CATS II and were completed independently by them. Where possible, participants were asked, after completion, if they had any queries regarding the questionnaires and they were also checked by a member of the CATS II data collection team that they were filled in correctly (see chapter 1.6., Behavioural questionnaires at age 9; CATS II data, for further information about missing data). No feedback was offered to families for questionnaires.

### i. 1.3.3.3.1. Strengths and Difficulties Questionnaire (SDQ)

The Strengths and Difficulties Questionnaires (SDQ) (231) used in the CATS II data collection, is a brief behavioural screening questionnaire. It was developed as an extension of Rutter's parent questionnaire (259, 260) and has been rated as one of the most commonly used tools for measuring psychopathological symptoms in children aged 4-17 years (261) with over 3,000 research citations (262). Reasons for this include its availability in 40 languages (263), it is quick to fill in and cost effective (264). It has been frequently used by the 'Children and Young People's Improving Access to Psychological Therapies' (CYP IAPT) programme delivered by the national health service in England. Multiple versions are available for the parent, teacher and also a self-report form for 11-16 year olds (265); CATS II administered the parent version. As the SDQ is so widely used it is constantly under scrutiny. Recently, Curvis et al. (266) found that younger children (6-10 years) could self-report the assessment tool when items were read aloud to them. This self-reporting could be seen as more accurate as it would more eloquently reflect the child's voice (although changing the mode of delivery would inherently change the nature of the tool).

The SDQ consisted of 25 items grouped into five subscales; hyperactivity-inattention, emotional symptoms, conduct problems, peer problems and (a subscale of strength) pro-social. The five subscales were found to be internationally validated and appeared to have good psychometric properties (263, 267-269).

### ii. 1.3.3.3.2. Child ADHD Questionnaire

The Child ADHD Questionnaire was a combination of two scales that had been modified by Thapar et al. (232), a version of the DuPaul scale (270) and Conners' abbreviated parent questionnaire (271). The questionnaire was originally devised to include 14 items from the

DSM-III revised ((272)), and it was modified and updated to also include four further items to match symptoms from the DSM-IV (125) and ICD-10 criteria (232). There were 18 items for scoring out of a possible 26 questions; this included eight 'dummy' questions from Conners' questionnaire (271), which were not included in the final score. There were three subscales measured by this ADHD tool, inattention (nine items), overactivity (four items) and impulsivity (five items). A total ADHD score was obtained by combining all of the three subscales together, which would generate a score within the range of 0-54. The higher the score obtained on the questionnaire, the higher the presence of ADHD symptoms would be for that individual (232).

The DuPaul scale was rated as having good internal consistency for the subscales and for the total score, as test-retest scores over four weeks displayed good coefficient alphas in the range of $r$ = .86-.92. (273). There was a moderate correlation between parent-teacher agreement scoring the scale ($r$ = .53) (270); which would have been expected for this type of scale (273).

### iii.    1.3.3.3.3. Social Communication Questionnaire (SCQ)

The SCQ had undergone substantial development prior to its use in CATS II. The Autism Diagnostic Interview (ADI) was developed by Le Couteur et al. (274), was revised (ADI-R) by Lord et al. (275), and was a common tool for screening for ASCs (276). The Autism Screening Questionnaire (ASQ) (276) included 40-questions that were based on the ADI-R and provided a score across three areas of functioning: reciprocal social interaction (e.g. interest in other children), language and communication (e.g. use of conventional gestures) and repetitive and stereotyped patterns of behaviour (e.g. unusual preoccupations). The ASQ was developed further into the SCQ (233), and its criteria used for screening for a diagnosis of an ASC is comparable to the DSM-IV (125).

The SCQ has been reported to be one of the most widely adopted screening tools for ASCs (277) and can screen from 4-40 years of age. Currently, there are two versions of the SCQ available; 'SCQ lifetime' (used in CATS II) which specifies an individual's entire developmental history and secondly, 'SCQ current' which only specifies an individual's behaviour in the past three months (278). The SCQ comprised of 40 yes-no questions with scores of ≥ 15 being an indication of the individual possibly having an ASC (for verbal and nonverbal children) (258). It has been found that by perhaps lowering the cut off to scores if ≥ 12, there would be an increased sensitivity (279), but this could lead to incorrectly identifying children as having an ASC when they could be typically developing (258).

Correlations between the ASQ and the ADI were highly significant across all domains (276). A problem with Berument et al.'s research into the ASQ was that all of the parents in the sample had previously completed the ADI-R, which may have affected the validity rates as the parents would have been used to similar questioning. However, the revised SCQ version was shown to have good external validity (278). Caution was still warranted, as Eaves et al. (258) identified that higher functioning children with ASCs were less likely to be identified by screening from the SCQ.

### iv.    1.3.3.3.4. General CATS Questionnaire

The general CATS questionnaire was used to gather demographic information about the participants (see Appendix 3: CATS II general questionnaire). The questionnaire was divided in two sections; the first to gather information on the mother and the second on the child around pregnancy and birth. The former section gathered demographic information and some paternal information; occupation and height. There was also a brief section on the mothers' medical history and any current drug therapy. The child section of the questionnaire gathered information on medical complications during pregnancy, breastfeeding, gestational age and also language spoken at home, school and the child's handedness. This questionnaire collated key information when controlling for certain covariates in the data. Most of the data generated was descriptive (i.e. medical history), and thus was not used in the main analysis here. Quantitative data such as breastfeeding, handedness of the child and language at home and school was used in the current research.

### *1.3.4.    Recruitment and options for participation*

An initial contact pack inviting the CATS I mothers to participate in the research was mailed to all in the SGTF groups; including those who did not participate in CATS I (See Appendix 4: Initial contact pack). Re-involvement in CATS II was centred on the age of the child; those born earliest during CATS I were contacted first in a rolling recruitment process over the 49 months of the project. When recruitment began in 2011 there was a very low response rate, consequently ethical amendments and approval was sought to make use of the Welsh Demographics Service; this enabled telephone calls to prompt responses from potential participants. The Patient Data Register was also used to ensure up-to-date addresses for individuals from the SGTF groups.

A slightly different protocol was adopted when contacting individuals from the normal GTF group who consented to participate in CATS I but did not participate in the cognitive assessments at age 3. There were 15,744 potential participants that could have been

contacted from the UK CATS I cohort; 240 participants were required from this group. As participants were contacted by year of registration into the study a random selection were mailed from each of the four years CATS I recruited for, totalling 5,000 packs being sent to mothers from the normal GTF group. Many participants mailed their response forms back from the initial contact packs indicating their willingness to take part in the study. See Hales et al. (280) for a flow chart of the recruitment process and overview of participation options including the data collected.

There were three options for re-involvement in the CATS study:

1. Post Packs

These packs contained a cover letter, specific consent, mother and child information sheets, a questionnaire pack, information on how to provide a saliva sample using the spit tubes (Oragene•DNA (OG-500) manufactured by DNA Genotek), two spit tubes and a freepost label (see Appendix 5: Information sheets from the post packs). From this information, the project could obtain samples of the mother and child's DNA, an indication of the child's behaviour, child pregnancy and delivery information and also a snapshot of the mother's medical history.

2. Remote/Home Visit

On successful booking of a remote/home visit, a confirmation letter was mailed out which included a copy of the specific consent and information letters for that visit (See Appendix 6: Appointment letters); and a text reminder was sent the day before.

Once at the visit, participants firstly had the opportunity to discuss the research and raise any initial questions if they wished. The mothers were reminded not to reveal which study group they belonged to so that assessments could be conducted blindly. The order of the visit was as follows, consent was taken from the mother, and then she was handed the behavioural questionnaires for completion. As the mothers were sent the appointment letters in advance, they would already have been made aware about an optimal test environment and whether or not they should stay present for the assessment. If any mothers did stay in the testing room, they were prompted that they were free to leave the child to the assessment. The WISC-IV and NEPSY-II items were subsequently administered to the child and the visit would finish upon completion of the spit tube DNA samples.

Following the visit, parents were mailed a report of the WISC-IV results (as generated by PsychCorp WISC Scorer software). If they had any questions, they could telephone or email to make an enquiry.

3.  Research Centre Visit

A 'complete dataset' was obtained from the participants if they attended the research centre for a morning visit. Similar to the remote/home visits, a confirmation letter was mailed out which included a copy of the specific consent and information letters for that visit (See Appendix 6: Appointment letters), and a text reminder was also sent the day before.

The visits were either at 9am or 10.45am and lasted for around 2-2.5 hours (see Table 8 below for a detailed breakdown of how the morning visits were constructed, note the additional data collected at these appointments for the CATS II project which are not used in the current thesis). As can be seen in the table, the research centre visits were under strict time constraints. Appointments could not have been offered later to participants as children were fasted for the blood sample collection; to enable lipid and blood sugar measurements. As can be seen in Table 8, there was a cross over between the two appointments which often led to minor practical difficulties when, for example, participants arrived late for their 9am appointment. Two appointments were offered per day to try and get as many people re-involved to CATS as possible. Therefore, up to three cognitive assessments could be conducted in a day, as most of the home visits commenced from 3.30pm onwards.

Table 8

*Research Centre Visit Appointment Breakdown*

| Time slot one | | Time slot two | |
|---|---|---|---|
| 9.00<br>PARTICIPANTS ARRIVE | | 10.45<br>PARTICIPANTS ARRIVE | |
| 9.00-9.20<br>Consents, blood/saliva collection and pregnancy testing in Clinical Research Facility | | 10.45-11.05<br>Consents, blood/saliva collection and pregnancy testing in Clinical Research Facility | |
| 9.20-9.40<br>Breakfast | | 11.05-11.25<br>Breakfast | |
| 9.40-10.10<br>Medical Physics Dept.: height, weight, blood pressure, arterial stiffness and bone density scans. Child measurements taken in current timeslot. | | 11.25-11.55<br>Medical Physics Dept.: height, weight, blood pressure, arterial stiffness and bone density scans. Child measurements taken in current timeslot. | |
| 10.10-10.40<br>Medical Physics Dept.: mother's measurements taken. | 10.10-11.40<br>Child cognitive testing | 11.55-12.25<br>Medical Physics Dept.: mother's measurements taken. | 11.55-13.25<br>Child cognitive testing |
| 10.40-11.00<br>Mother completion of questionnaires | | 12.25-12.45<br>Mother completion of questionnaires | |
| 11.40 FINISH | | 13.25 FINISH | |

### 1.3.5. General Analysis

As stated in the current chapter overview, this thesis used the same data that I collected for the CATS II research project. Details of how the CATS II data was analysed (for publication), can be found on the study protocol (280). Different analysis techniques were adopted for CATS II and were similar to the exploration design found in the CATS I paper: i.e. looking at IQs below a threshold (200). All of the data was cleaned then kept for the current analysis, and a copy also passed to the designated statistician for the CATS II project.

Data was analysed in IBM SPSS Statistics, version 20. The initial analysis included re-running the CATS I data for the UK cohort only (see previous chapter, 1.2., Re-analysis of the intelligence at age 3; UK CATS I cohort). This served as the pilot study for the subsequent analyses. The cognitive measures were scrutinised for dataset accuracy by using histograms, z-scores and cross tabulations to identify any outliers and errors in the data set. Descriptive statistics were presented as means and SDs.

### 1.3.5.1. WISC-IV

The data collected from the WISC-IV (intelligence measure) attempted to answer two (in bold) of the four research aims:

i. **By reassessing the children at the older age of 9, would there be a continued non-significant difference identified for intelligence.**

ii. Would there also be non-significant findings at age 9 in other potential areas of cognition.

iii. **As there were no differences between the treated and untreated SGTF groups at age 3, would any differences to the normal GTF group be measurable; as there is a wealth of studies displaying that an underactive thyroid during pregnancy does not affect a child's cognition (53, 56, 91-93, 95, 96).**

iv. Would there also be non-significant differences between the groups that extend beyond cognition, i.e. behaviour.

Based on the literature reviewed in General introduction (chapter 1.1.), it was hypothesised that there would be a significant difference between the normal GTF and untreated SGTF groups (34, 49-57, 90, 91), but there would not be a significant difference between the treated and untreated SGTF groups (based on the CATS I findings of offspring intelligence measured at age 3 (200) and chapter 1.2., Re-analysis of intelligence at age 3; UK CATS I cohort).

The primary analysis for CATS II and this thesis (based on the main deficits reported in the literature), was assessing the five IQs between all three groups of participants by a MANCOVA. This multivariate analysis was followed by subsequent univariate analysis of variances (ANOVA) dependent upon statistically significant results. The multivariate analysis of variances (MANOVA) was chosen as the data required a model to fit all three groups (between-subjects factor) and multiple dependent variables into one analysis. The MANOVA also allowed manipulation for adjustments of several covariates and included four models of analysis, analysed in a step-by-step manner controlling for a total of six covariates. Thus the final output for the WISC-IV was analysed by a MANCOVA (see Figure 6 below for the four models of analysis used in the current research from the CATS II cognitive study protocol (280)). These models of analysis were adopted so that the work here would be comparable to the CATS II analysis. As mentioned, most of the covariate information was obtained from the General CATS questionnaire; of note, mother age at time of consent into CATS I was quartiled but mean ages were calculated for the reader's information. Furthermore, the social deprivation score was calculated from postcode scores from StatsWales (281) and OpenDataCommunities (282) (England). The Welsh and English ranked scores were changed into quintiles to make the data comparable. See Tables 9 and 10 for further information on the scoring; the higher the score, the less socially deprived the ranking was.

Table 9
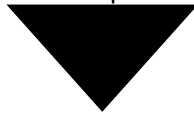
*StatsWales Social Deprivation Scores Converted into Quintiles*

| Social Deprivation | StatsWales Score | CATS II Ranking |
|---|---|---|
| Most deprived | 1-382 | 1 |
| | 383-764 | 2 |
| | 765-1,146 | 3 |
| | 1,147-1,528 | 4 |
| Least deprived | 1,529-1,909 | 5 |

Note. CATS=Controlled Antenatal Thyroid Screening.

Table 10

*OpenDataCommunities (England) Social Deprivation Scores Converted into Quintiles*

| Social Deprivation | OpenDataCommunities Score | CATS II Ranking |
|---|---|---|
| Most deprived | 1-6,497 | 1 |
| | 6,498-12,994 | 2 |
| | 12,995-19,489 | 3 |
| | 19,490-25,986 | 4 |
| Least deprived | 25,987-32,482 | 5 |

Note. CATS=Controlled Antenatal Thyroid Screening.

The WISC-IV and NEPSY-II results were not compiled into one main MANOVA for two reasons. Firstly, the MANOVA would only include those participants who had completed all testing (as discussed in chapter 1.5., Additional cognitive assessments at age 9; CATS II data, there were multiple reasons why there was a decreased number of participants who completed the NESPY-II). Secondly, as this thesis' findings aimed to be comparable to the CATS II findings, the WISC-IV was analysed separately as IQ was the primary outcome of CATS II and the NEPSY-II served as *additional* cognitive assessments.
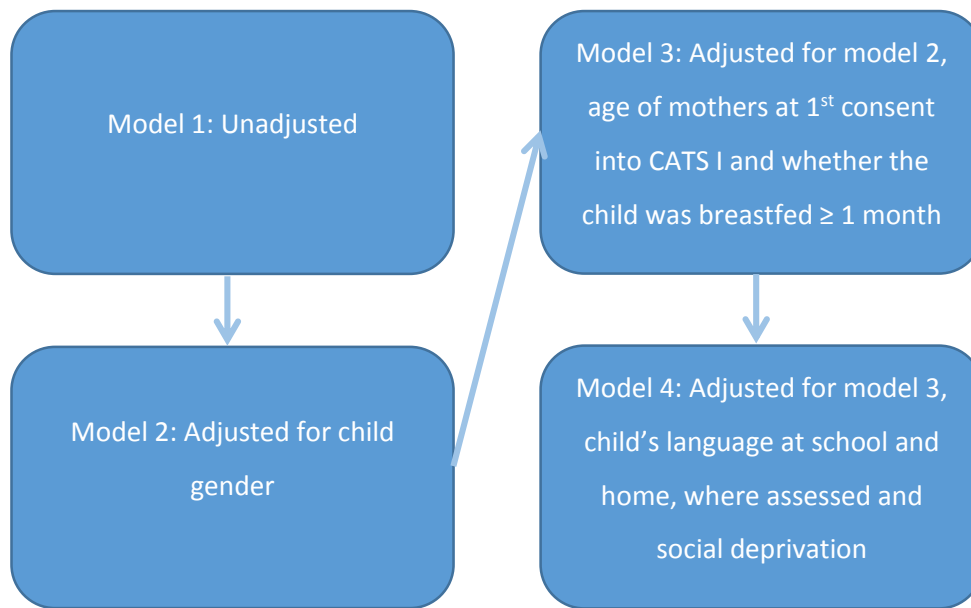
Figure 6: Models of Analysis

CATS=controlled antenatal thyroid screening.

### 1.3.5.2. NEPSY-II

The data collected from the NEPSY-II (additional cognitive domains) aimed to cover two (bold) of the four research aims:

i. By reassessing the children at the older age of 9, would there be a continued non-significant difference identified for intelligence.

**ii. Would there also be non-significant findings at age 9 in other potential areas of cognition.**

**iii. As there were no differences between the treated and untreated SGTF groups at age 3, would any differences to the normal GTF group be measurable; as there is a wealth of studies displaying that an underactive thyroid during pregnancy does not affect a child's cognition (53, 56, 91-93, 95, 96).**

iv. Would there also be non-significant differences between the groups that extend beyond cognition, i.e. behaviour.

Based on the literature reviewed in the General introduction (chapter 1.1.) and similar to the IQ hypotheses, it was predicted that there would be a cognitive deficit for the untreated SGTF group compared to the normal GTF group (as supported by (10, 32, 49, 54-58)). Similarly, that the treated SGTF and untreated SGTF groups would have non-significant differences between their scores (alike to the CATS I findings (200) and chapter 1.2., Re-analysis of intelligence at age 3; UK CATS I cohort).

This secondary analysis was to explore the additional cognitive assessments administered to the children, i.e. the NEPSY-II tests. Similar to the IQ analysis, this took the form of a MANOVA before the relative adjustments were made for a MANCOVA. The MANOVA also included all three groups of participants and investigated the differences between the scaled scores for the MD, MDD, FTDH, FTNDH and narrative memory (NM) tests. The subtest of LM was investigated separately (by ANOVA) as the test was introduced latterly into the CATS II study, thus it was administered to fewer participants. Therefore, if it *had* been included in the main NEPSY-II MANCOVA, the group numbers would have dropped as the MANCOVA would only include those participants that had a complete set of data.

### 1.3.5.3.    Questionnaires

The data collected from the questionnaires (behaviour) attempted to answer one of the four thesis research questions:

i.    By reassessing the children at the older age of 9, would there be a continued non-significant difference identified for intelligence.

ii.    Would there also be non-significant findings at age 9 in other potential areas of cognition.

iii.    As there were no differences between the treated and untreated SGTF groups at age 3, would any differences to the normal GTF group be measurable; as there is a wealth of studies displaying that an underactive thyroid during pregnancy does not affect a child's cognition (53, 56, 91-93, 95, 96).

iv.    **Would there also be non-significant differences between the groups that extend beyond cognition, i.e. behaviour.**

As the null hypothesis was adopted for the WISC-IV and NEPSY-II that there would be no differences between the treated and untreated SGTF groups, similarly it was proposed that there would be no differences between behaviour either (supported by (19, 45, 144-146, 148, 171, 172, 174)). It was predicted that there would be a behaviour deficit for the untreated SGTF group compared to the normal GTF group; as has been suggested in the literature (19, 44, 45, 143, 148-151, 170, 175-180).

Therefore, a third main analysis was conducted to explore the behavioural questionnaires. This was undertaken by using total scores of domains by a MANOVA (model one), and MANCOVA by models two and three of the analysis (here, model three also included controlling for social deprivation and child age). School and home language and where the child was assessed were not taken into consideration as these were proposed not have affected the questionnaire outcomes.

### 1.3.5.4. Exploratory Analysis (section 2)

In the second section of this thesis (from page 125), multiple exploratory analyses were undertaken. Briefly, this section comprises of the following:

i) Investigations of the effects of the covariates on the dependent variables of CATS II; explored by t-tests and MANOVAs.

ii) IQs of CATS I and CATS II were compared by correlations and a repeated measures MANCOVA.

For further details of the specific analyses completed, please see the relevant results chapters.

### 1.3.6. Chapter Summary

This chapter described the methodology behind the assessment measures used for this thesis and attempted to disaggregate the separable work for this thesis and the CATS II research publication. This chapter described the population of CATS II, eligibility for re-participation and the desired group sizes that were worked towards. The development, reliability and validity of the cognitive assessments I conducted for the study and also the questionnaires, were discussed. Further detail of what each tool consisted of can be found in the following three chapters, as well as a brief overview of how they were scored. This chapter concluded with how the participants took part in the study, and highlighted how the cognitive assessments only took place for those who attended the research centre or opted for a remote/home visit. As the cognitive assessment data collection period lasted for around 28 months with a maximum of three assessments being conducted in a day, every effort was made to maximise recruitment into CATS II. The following chapter details the WISC-IV data analysis.

## 1.4. Intelligence measured at age 9; CATS II data

### 1.4.1. Chapter Overview

This chapter contains the work relating to the IQ test assessments and statistical analysis between the three groups of participants. Within this chapter, there is a brief overview of the topic as the main literature identified is described in detail in the General introduction (chapter 1.1.). The methods section describes the process of the assessment administration, there is a statistical analysis plan and explanation of how the data was prepared. The results section follows, then a discussion of the results including the limitations and main conclusions.

### 1.4.2. Introduction

The aim of the work presented in this chapter was to investigate whether there were differences in IQ measurements between the three groups of participants (treated SGTF, untreated SGTF and normal GTF). Based on previous research, it was hypothesised that there would be a significant difference between the normal GTF and untreated SGTF groups (34, 49-57, 90, 91), but no difference between the treated and untreated groups based upon the CATS I findings of IQs at age 3 ((200) and chapter 1.2., Re-analysis of intelligence at age 3; UK CATS I cohort).

### 1.4.3. Method

The anticipated sample size for CATS II was 480 participants to address the null hypothesis that there would be no difference of IQs between all three participant groups. The participants were assessed for IQ by the WISC-IV (78), UK version. The WISC-IV was administered in a standardised format and I was the sole examiner on the study which provided consistency. I was blinded to the participant groups as not to introduce any bias to the testing (or scoring) and 10% of packs were double scored by an educational psychologist on the project team to ensure reliability.

The WISC-IV can be used to assess children between the ages of 6 years 0 months and 16 years 11 months. No hints or help for the examinee were allowed during testing as this could have incurred a 'spoilt' answer. Attempting to build rapport before the assessment was important to help maximise effort from the participant. Administration time of the core ten subtests took between 65-80 minutes. The more intelligent a child was, the longer the assessment was; as more items were administered per test. The order of the tests were, blocks design, similarities, digit span, picture concepts, coding, vocabulary, letter-number

sequencing, matrix reasoning, comprehension and symbol search. See Table 11 for further information of each subtest, as organised by their respective domains.

Table 11
*Descriptions of the Wechsler Intelligence Scale for Children- Fourth Edition, UK Subtests*

| IQ Domain | Subtest | Description |
|---|---|---|
| VCIQ: A measure of an individual's verbal capabilities | Similarities | The child was presented with two words that had a common object or concept and was requested to describe orally how they were similar. |
| | Vocabulary | The child was requested to give verbal definitions of words. |
| | Comprehension | This item required the child to answer questions based on their understanding of social situations and general principles. |
| PRIQ: An individual's non-verbal and fluid reasoning capabilities | Block Design | All items required the child to view a constructed model or picture from the stimulus book and then use red-and-white blocks to re-create the pattern within a specified time limit. |
| | Picture Concepts | The child was presented with two or three rows of pictures and was requested to choose one picture from each row to form a group with a common characteristic. |
| | Matrix Reasoning | The child was presented with an incomplete matrix and was requested to select the missing portion from five options. |
| WMIQ: Identified an individual's short term memory capacity | Digit Span | Digit span was comprised of two components: digit span forwards and digit span backwards. Digit span forwards required the child to repeat numbers in the same order as they were read aloud to them. Digit span backwards required the child to repeat the numbers in reverse order to how they were said to them. |
| | Letter-number Sequencing | The child was read a sequence of numbers and letters and was required to recall the numbers in ascending order and then the letters in alphabetical order. |
| PSIQ: Measured the individual's processing speed. | Coding | The child was required to copy symbols that were paired with numbers. |
| | Symbol Search | The child was required to scan a search group and indicate whether the target symbol(s) matched any symbols in the search group within a specified time limit of two minutes. |

Note. VCIQ=verbal comprehension intelligence quotient, PRIQ=perceptual reasoning intelligence quotient, WMIQ=working memory intelligence quotient, PSIQ=processing speed intelligence quotient.

The average for a scaled score was 10 and SDs were ± 3 points, whereas IQs had a mean of 100 and SDs were ± 15 points. The design of the scale is such that about 68% of children obtain IQs between 1 SD above and below the mean, 96% in the 2 SD range and 99.8% in the 3 SD range (IQs: 55-145) (78). Percentile ranks were also calculated, the ranks would range from 0.1-99.9 with 50 being the mean and median.

The WISC-IV raw scores were calculated by hand and then inputted into an-automated WISC-IV scorer to generate age-equivalent scaled scores. The automatic scorer calculated the sum of scaled scores for a domain to produce the IQ. The FSIQ would be calculated by the software adding the total sum of ten scaled scores. The WISC-IV scorer also generated a parent-report which detailed the IQ scores, percentile ranks and classifications (e.g. average, high average etc.). The percentile ranks were useful to parents as they quantified how their child theoretically performed against a group of 100 'typically developing' children. If the participants had any questions regarding their child's results, they were able to directly contact me.

### 1.4.4.    Statistical Analysis

Data entry and data cleaning was completed on a blinded dataset (i.e. the participants were still all grouped as one complete cohort, as they had been during the data collection phase of the study. This method of 'blinding' meant a reduction in any potential research bias between groups, for example when removing outliers from the dataset.

Checks were executed to investigate whether the correct child was re-recruited into CATS. This was completed by comparing child date of birth against the estimated delivery date. This was also completed during the data collection phase and subsequently six participants were removed prior to data cleaning. As stated in the protocol paper for CATS II (280), the age ranges were ≥ 7 years 0 months to ≤ 10 years 11 months. Based on this, one set of data was removed for the child being 6.5 years and five datasets were removed for the children being ≥ 11 years to 11 years and 5 months. Further to this, in the original CATS I study (200), participants were excluded if they were carrying twins. A set of twins were assessed in CATS II and thus, were also removed.

The following nominal data was checked for miss-inputting by range checks: how participants took part in the study (i.e. attending research centre or home/remotely), child gender, dates of participation and assessment, age at testing, and school and home language for the child. WISC-IV scaled scores, followed by the IQs and percentile ranks, were checked for input errors by frequency analysis to ensure none of the values were outside the plausible ranges. Following this, the distributions (floor/ceiling effects were explored) and outliers were examined. From this method, nine IQs were flagged as input errors, and corrected by pulling hardcopy data.

Histograms and z-scores were computed for all IQs. The histograms (Figures 7 and 8) indicated how there may have been a possible outlier in the dataset. Z-scores were

computed, which confirmed this individual as having four out of five IQ z-scores < -3; none were > 3. After this individual had completed cognitive testing, a note had been added to the dataset:

"NB: concept of PS tasks for WISC not grasped. No effort on matrix reasoning."

Based on this, this participant was removed from all subsequent CATS II analyses. There were four other individuals that could have been classified as outliers based on their z-scores for VCIQ and WMIQ. These individuals were not removed as they only presented an abnormal z-score on one IQ score. Further to this, for our study population to be representative of the general population, we would still require some individuals to fall within the 'tail-ends' on the bell-curves in order to represent our sample pragmatically and to allow variability. Moreover, if these four individuals had been removed, then it could be argued that you could reanalyse the z-scores to search for *more* possible outliers and also analyse z-scores for the ten subtests; in which participant numbers would decrease and decrease. Therefore the total CATS II sample was 452 with complete IQ data.

With the identified removed participants, the final 452 participants' histograms were recalculated and were as follows (Figure 9 and 10): all ten histograms calculated show the data fitting well into a normal distribution curve with little skewness or kurtosis effects.
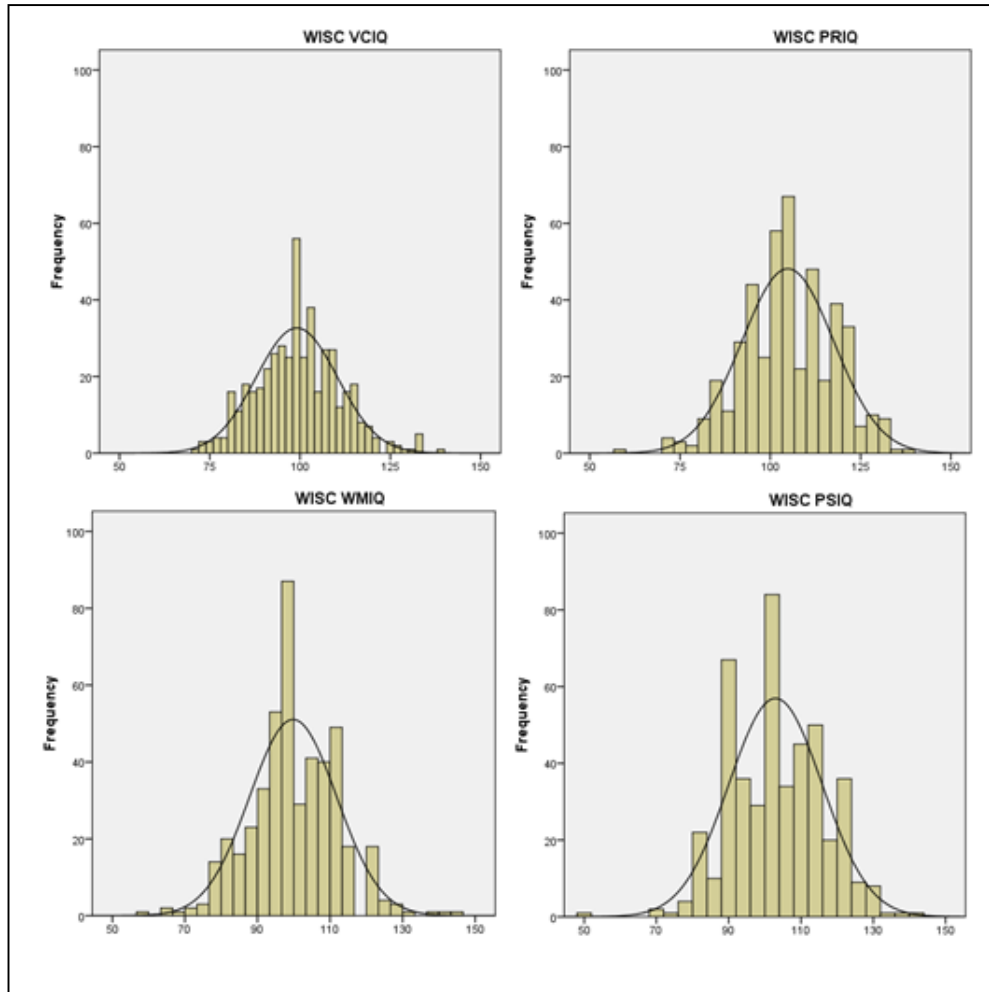
Figure 7: Histograms of Verbal Comprehension Intelligent Quotient (IQ), Perceptual Reasoning IQ, Working Memory IQ and Processing Speed IQ (n=461) in the Controlled Antenatal Thyroid Screening study II with removed participants

WISC=Wechsler intelligence scale for children-fourth edition, UK. X axis=IQ scores, Y axis=frequency.
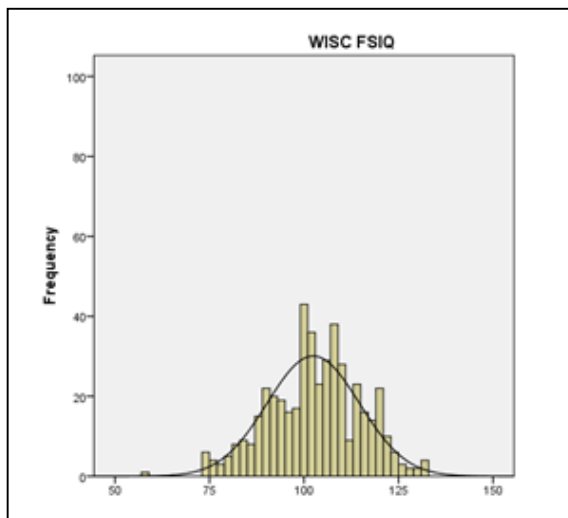
Figure 8: Histogram of Full Scale Intelligent Quotient (IQ) (n=461: complete dataset) in the Controlled Antenatal Thyroid Screening study II

WISC=Wechsler intelligence scale for children-fourth edition, UK. X axis=IQ scores, Y axis=frequency.
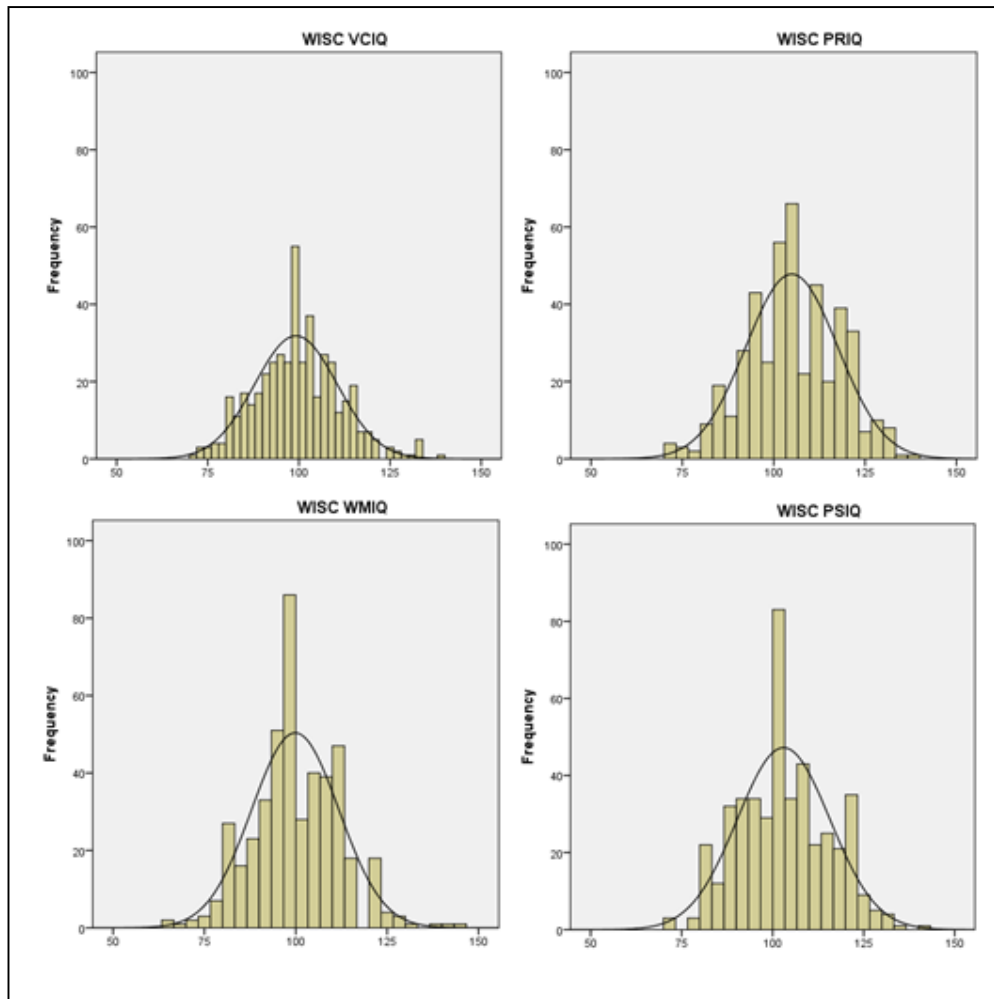
Figure 9: Histograms of Verbal Comprehension Intelligent Quotient (IQ), Perceptual Reasoning IQ, Working Memory IQ and Processing Speed IQ (n=452) in the Controlled Antenatal Thyroid Screening study II with removed participants

WISC=Wechsler intelligence scale for children-fourth edition, UK. X axis=IQ scores, Y axis=frequency.
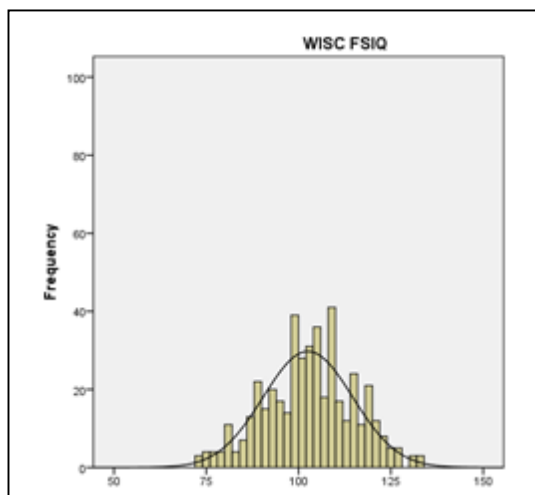


Figure 10: Histogram of Full Scale Intelligent Quotient (IQ) (n=452) in the Controlled Antenatal Thyroid Screening study II with removed participants

WISC=Wechsler intelligence scale for children-fourth edition, UK. X axis=IQ scores, Y axis=frequency.

All Kolmogorov-Smirnov (and Shapiro-Wilk) normality tests were generated with all $p$'s < .044 for the IQ measures. However, means and medians appeared close and skewness and kurtosis ranges were all within the -1 - +1 range and thus were normal (see Table 12 below). The WISC-IV data was analysed by parametric tests as the variables were continuous, and based on the means, medians, skewness, kurtosis and the histograms (see figures 9 and 10), the data was accepted as being normally distributed.

Table 12
*Means, Medians, Skewness and Kurtosis for all 5 Intelligent Quotients (IQs)*

| IQ Domain | Mean | Median | Skewness | Kurtosis |
|---|---|---|---|---|
| WISC-IV VCIQ | 99.19 (11.32) | 99.00 | .303 | .376 |
| WISC-IV PRIQ | 104.94 (12.57) | 104.00 | -.025 | -.299 |
| WISC-IV WMIQ | 99.86 (11.93) | 99.00 | .233 | .599 |
| WISC-IV PSIQ | 103.00 (12.73) | 103.00 | .133 | -.507 |
| WISC-IV FSIQ | 102.54 (12.12) | 102.00 | -.103 | -.317 |

Note. Standard deviations appear in parentheses below means. WISC-IV=Wechsler Intelligence Scale for Children- fourth edition, UK, VC=verbal comprehension, PR=perceptual reasoning, WM=working memory, PS=processing speed.

.

.

Table 13

*Wechsler Intelligence Scale for Children- Fourth Edition, UK (WISC-IV) Participant Group Demographics*

| | | Group | | |
|---|---|---|---|---|
| | | Normal GTF (n=233) | Treated SGTF (n=118) | Untreated SGTF (n=101) |
| Participated (n) | Research Centre (n=assessed at home) | 191 (79) | 78 (51) | 47 (29) |
| | Remote/Home | 42 (18%) | 40 (33.9%) | 54 (53.5%) |
| Total home visits (n) | | 121 (52%) | 91 (77.1%) | 83 (82.2%) |
| Mean Age (yrs) | | 9.66 | 9.40 | 9.46 |
| Gender (n) | Male | 117 | 65 | 50 |
| | Female | 116 | 53 | 51 |
| Child's language at school and home (n) | English School and English at Home | 181 (77.7%) | 94 (79.7%) | 88 (87.1%) |
| | Welsh School and English at Home | 42 (18%) | 20 (16.9%) | 11 (10.9%) |
| | Welsh School and Welsh at Home | 7 (3%) | 3 (2.5%) | 1 (1%) |
| | English School and Other Language at Home (not Welsh or English) | 2 (.98%) | 1 (.8%) | 1 (1%) |
| | Welsh School and Other Language at Home (not Welsh or English) | 1 (.4%) | 0 | 0 |
| Whether the mother breastfed over 1 month | Yes | 150 (64.6%) | 72 (61.5%) | 56 (56%) |
| | No | 82 (35.3%) | 45 (38.5%) | 44 (44%) |
| | Missing | 1 | 1 | 1 |
| Mother age at time of consent into CATS I Mean of quartile (mean, years) | | 2.5 (31.82) | 2.2 (30.26) | 2.3 (30.99) |
| Social deprivation score (Mean of quintile) | | 3.7 | 3.8 | 3.3 |

Note. Further information appears in parentheses following n's were applicable. SGTF=suboptimal gestational thyroid function, CATS=controlled antenatal thyroid screening.

Table 13 shows the group demographics of each participant group. How the participants opted to take part in CATS II varied between the groups. More than 80% of the normal GTF group opted to attend the research centre for a visit, compared to around 60% of the treated and 45% of the untreated SGTF groups. 'Participant group' and 'where the child was assessed' were compared by a Chi-square and confirmed to be significantly different; $X^2$ (2, n = 452) = 36.765, $p < .001$. This decreasing attendance between the groups could be a product of individuals from the SGTF groups having participated previously and perhaps, not

wanting to sacrifice as much time to the project again. This was also reflected in the number of home visits that were a consequence of shortened research centre visits. The ages of the children were compared by an ANOVA, and significance was identified: $F$ (2, 449) = 6.182, $p$ = .002, $\eta_p^2$ = .027. The treated SGTF were significantly younger than the normal GTF group: $p$ = .004, 95% CI [-.4623, -.0664]. Conversely, this should not have affected the IQ testing as the WISC-IV takes account of child age (described further in the discussion). Gender of the offspring and child's language at school and home appeared to be equal between groups, as did breastfeeding, mother age at time of consent into CATS I and social deprivation score. By having any of these differences between the groups, it could indicate compromised randomisation. I attempted to overcome these effects by controlling for six covariates; child gender, age of mother at time of consent into CATS I, whether the mother breastfed over one month, where the child was assessed, child's language at school and home and social deprivation score (see Figure 6 page 59. chapter 1.3., Methods for the cognitive and behavioural data collection for the CATS II study, for further information).

The data cleaning process resulted in 15 children being removed and leaving the final sample at 452. The WISC-IV data was analysed using a MANCOVA in IBM SPSS Statistics version 20. I decided to replicate the proposed four models of analysis from the CATS II protocol (280) so that the findings would be comparable (see chapter 1.3., Methods for the cognitive and behavioural data collection for the CATS II study, for further information). The participant group information was added to the database after data cleaning was complete. A blinded analysis could not have been undertaken as, from the start of the data collection phase, those responsible for recruitment were aware of the target numbers for each group and thus it was deducible which was which group based on the frequency of participants in each. Any significant differences identified at the multivariate level were followed up with post hoc tests (Bonferroni was selected as one of the most 'strict' post hoc corrections to control the overall type one error rate when multiple testing is carried out (210)), to investigate where the significant differences were.

As shown in chapter 1.3. (Methods for the cognitive and behavioural data collection for the CATS II study), the analysis included four models to adjust for all covariates (see Figure 6). Information for models two to four was mainly generated by the 'General CATS Questionnaire' (see Appendix 3: CATS II general questionnaire). Age of mothers at time of first consent (recruitment) into CATS I, i.e. age during their pregnancy, was collated from CATS I. Social deprivation was assessed by postcode scores from StatsWales (281) and OpenDataCommunities (282). The Welsh and English ranked country scores were separated

71

into quintiles to allow all participants to have their social deprivation controlled as a covariate for the analysis.

Similar to the CATS I UK only cohort analysis (chapter 1.2., Re-analysis of intelligence at age 3; UK CATS I cohort), percentages of participants scoring IQs of ≤ 85 (1 SD from the mean) were computed, to enable comparisons to be drawn from the main CATS II findings to the work completed for this thesis. The frequency of participants scoring ≤ 85 were tested by a Chi-Square and any significant values were followed up and adjusted for covariates by a multinomial logistic regression. This was explored in CATS I as there was an attempt to replicate the findings from Haddow et al.'s (54) benchmark study which identified significantly more children with IQ scores below 85 if they were born to mothers with high TSH compared to those with normal thyroid function. Lazarus et al. (200) proposed that the 15% of the untreated SGTF group would score ≤ 85 (Haddow et al.) and that there would be 5% from the treated SGTF; comparable to Haddow et al.'s normal thyroid group. In line with the WISC-IV (78), it was anticipated that 16% of the normal GTF would score ≤ 85.

### 1.4.5. Results

#### 1.4.5.1. General attendance information

As can be seen in the bar chart below (Figure 11), those recruited back into CATS II were very close to the targets of 240 for the normal GTF group, and 120 into each treated and untreated SGTF group. The smallest group recruited back into CATS II were those from the untreated SGTF group. A reason for this differential attrition may be that mothers from this group did not receive treatment from us during their pregnancy and subsequently might view the study team as 'not helping/helpful'. As we did not 'help them' during their pregnancies to correct for their SGTF, why would they now help us with the continued study? Further recruitment frequencies and demographics are displayed by participant group in Table 13.

Figure 11: Graph to show participants with completed cognitive assessments in the Controlled Antenatal Thyroid Screening study II (n = 452)

SGTF=suboptimal gestational thyroid function. X axis= study groups, Y axis = count of participants.

### 1.4.5.2. Analysis

Table 14 shows the means and SDs for all study IQs as per group and adjusted for by model four (see Figure 6) (NB total n has decreased due to missing data for breastfeeding (n = 3)). The graph (Figure 12) displays the means achieved by the groups pictorially (unadjusted model); error bars have also been included.

Table 14

*Intelligent Quotient (IQ) means by study group and model 4*

|  | Participant Group | Mean | N |
|---|---|---|---|
| WISC-IV VCIQ | normal GTF | 99.81 (11.256) | 232 |
|  | treated SGTF | 97.62 (10.002) | 117 |
|  | untreated SGTF | 99.74 (12.841) | 100 |
|  | Total | 99.22 (11.339) | 449 |
| WISC-IV PRIQ | normal GTF | 105.37 (12.298) | 232 |
|  | treated SGTF | 104.32 (12.219) | 117 |
|  | untreated SGTF | 104.83 (13.653) | 100 |
|  | Total | 104.98 (12.571) | 449 |
| WISC-IV WMIQ | normal GTF | 99.91 (11.236) | 232 |
|  | treated SGTF | 99.61 (13.128) | 117 |
|  | untreated SGTF | 100.04 (12.294) | 100 |
|  | Total | 99.86 (11.964) | 449 |
| WISC-IV PSIQ | normal GTF | 103.66 (12.748) | 232 |
|  | treated SGTF | 102.99 (12.688) | 117 |
|  | untreated SGTF | 101.68 (12.806) | 100 |
|  | Total | 103.04 (12.741) | 449 |
| WISC-IV FSIQ | normal GTF | 103.10 (11.683) | 232 |
|  | treated SGTF | 101.66 (11.978) | 117 |
|  | untreated SGTF | 102.42 (13.315) | 100 |
|  | Total | 102.57 (12.127) | 449 |
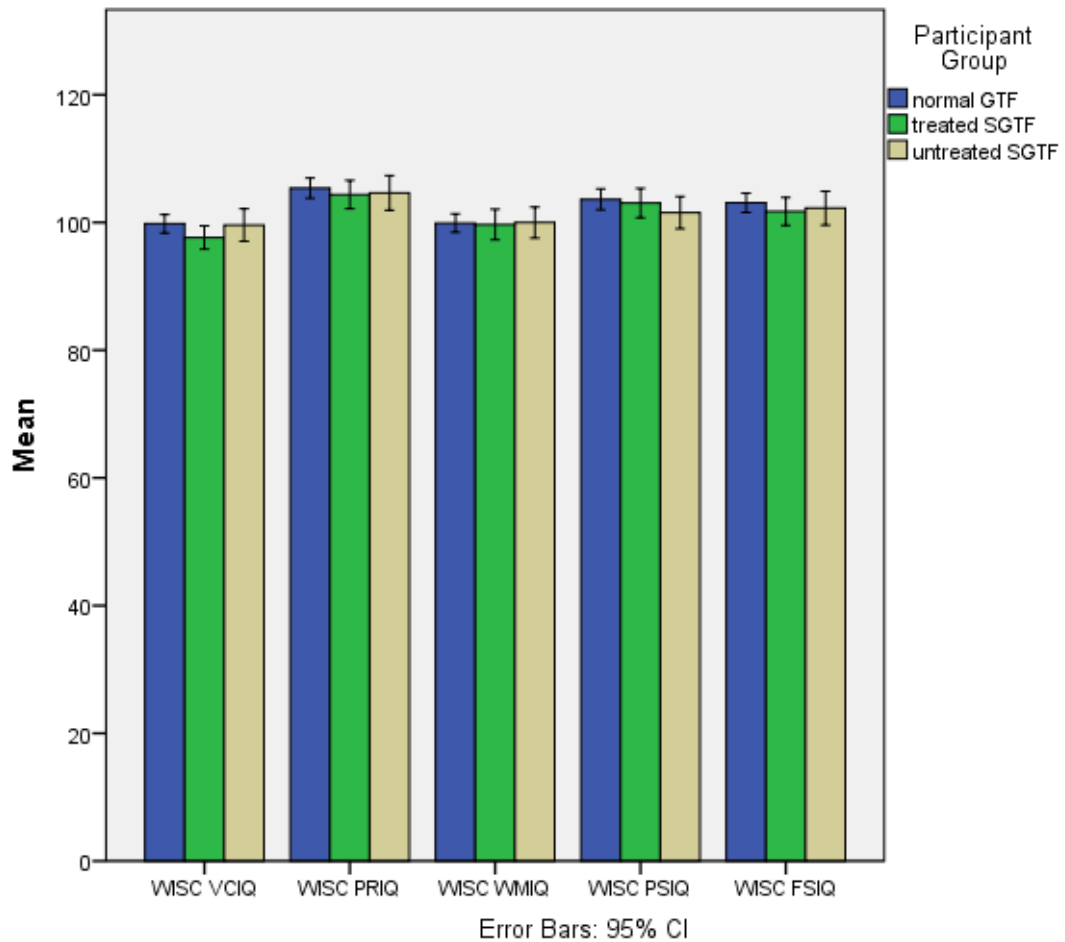
Figure 12: Means per group from the Intelligent Quotient (IQ) testing

WISC-IV=Wechsler Intelligence Scale for Children- fourth edition, UK, VC=verbal
comprehension, PR=perceptual reasoning, WM=working memory, PS=processing speed,
FS=full scale, SGTF=suboptimal gestational thyroid function, CI=confidence interval. X axis=
IQ domain, Y axis= mean of score.

As stated, the WISC-IV data was investigated by four models of analysis (see chapter 1.3.,
Methods for the cognitive and behavioural data collection for the CATS II study, and the CATS
II protocol (280)). The unadjusted model of analysis by MANOVA found no significant
differences between the normal GTF, treated SGTF and untreated SGTF groups on IQ
measures using Roy's largest root (most powerful multivariate statistic (210)), $\Lambda_{ROY}$ = .012, $F$
(5, 446) = 1.087, $p$ = .367, $\eta_p^2$ = .012. The second model of analysis controlled for child gender
and also identified a non-significant effect between groups: $\Lambda_{ROY}$ = .013, $F$ (5, 445) = 1.201, $p$

= .308, $\eta_p^2$ = .013. Model three adjusted for model two plus age of mothers at first consent into CATS I (i.e. age at which they were pregnant) and whether their child was breastfed for more than one month. Similarly, by Roy's largest root, no significance was reached: $\Lambda_{ROY}$ = .012, $F$ (5, 440) = 1.088, $p$ = .366, $\eta_p^2$ = .012. The final model of analysis controlled for models two, three, school and home language of the child, where the cognitive assessments took place, and also the social deprivation of the mother and child (based on their postcode). There was a non-significant effect between groups for IQ measures, $\Lambda_{ROY}$ = .013, $F$ (5, 437) = 1.112, $p$ = .353, $\eta_p^2$ = .013. There were no main effects or interactions found from the data. As the multivariate analysis was non-significant, no further investigations of univariate effects were completed. Partial Eta Square results indicated that the effect size in CATS II were small (283), which was in keeping with the small sample sizes in the study. Thus the study may have lacked power to detect a significant effect, if one existed.

For the secondary analysis, IQs were analysed for the percentage per group scoring ≤ 85 and tested for significance by use of a Chi-Square. Results can be found in Table 15 below. This was computed as it was initially investigated in CATS I. This was a further exploratory analysis as, if there was no significant difference for the continuous outcome of IQ, there may have been a binary outcome measurable by lower IQ scores. As can be seen in Table 15, the PRIQ was almost significantly affected, it appeared that the untreated SGTF children had many more scoring ≤ 85. This was investigated further with a multinomial logistic regression to enable the adjustments of covariates to the dependent variable.

Table 15
*Intelligent Quotients (IQs) ≤ 85 (%)*

|  | IQs ≤ 85 (%) | | | | |
|---|---|---|---|---|---|
|  | VCIQ | PRIQ | WMIQ | PSIQ | FSIQ |
| Normal GTF | 28 | 11 | 18 | 22 | 15 |
| (n = 233) | (12%) | (5%) | (8%) | (9%) | (6%) |
| Treated | 19 | 7 | 14 | 8 | 10 |
| SGTF | (16%) | (6%) | (12%) | (7%) | (8%) |
| (n = 118) |  |  |  |  |  |
| Untreated | 12 | 12 | 10 | 10 | 12 |
| SGTF | (12%) | (12%) | (10%) | (10%) | (12%) |
| (n = 101) |  |  |  |  |  |
| Pearson Chi-Square | $p$=.520 | $p$=.051 | $p$=.438 | $p$=.648 | $p$=.247 |

*Note.* Percentages of scores per group are appear in parentheses below totals. VC=verbal comprehension, PR=perceptual reasoning, WM=working memory, PS=processing speed, FS=full scale, SGTF=suboptimal gestational thyroid function.

A multinomial logistic regression was only executed for the PRIQ, as this was the only dependent variable that was close to significance ($p$ = .051) as calculated by the Chi-square. The regression controlled for all six covariates that were also adjusted for in the MANCOVA, and revealed that, following adjustments, non-significance was sustained and there were no differences between the groups for PRIQ ≤ 85 between the normal GTF, treated and untreated SGTF groups. See Tables 16 and 17 for further details of beta values, significance and -2 log likelihood comparing (comparison of normal GTF to treated SGTF was also non-significant, $p$ = .884).

Table 16
*Table Displaying the Regression Model's Fit for the Data*

| Model | Model Fitting Criteria | Likelihood Ratio Tests | | |
|---|---|---|---|---|
| | -2 Log Likelihood | Chi-Square | *df* | Sig. |
| Intercept Only | 187.870 | | | |
| Final | 156.443 | 31.428 | 8 | .000 |

Note. See improved figure for -2 Log Likelihood. Df=degrees of freedom.

Table 17

Main Output from the Multinomial Logistic Regression, Perceptual Reasoning Intelligent Quotient (PRIQ) ≤ 85

| PRIQ ≤ 85 | B | Std. Error | Wald | df | Sig. | Exp(B) | 95% Confidence Interval for Exp(B) | |
|---|---|---|---|---|---|---|---|---|
| | | | | | | | Lower | Upper |
| Gender | -.029 | .408 | .005 | 1 | .944 | .972 | .436 | 2.164 |
| Where Assessed | -.607 | .503 | 1.452 | 1 | .228 | .545 | .203 | 1.462 |
| Language | -.576 | .500 | 1.328 | 1 | .249 | .562 | .211 | 1.497 |
| Mother Age | -.422 | .227 | 3.450 | 1 | .063 | .656 | .420 | 1.024 |
| Breast Fed 1mns | -.397 | .424 | .877 | 1 | .349 | .672 | .293 | 1.544 |
| Social Deprivation | -.464 | .143 | 10.538 | 1 | .001* | .629 | .475 | .832 |
| [Normal GTF] | -.532 | .484 | 1.209 | 1 | .271 | .587 | .227 | 1.516 |
| [Treated SGTF] | -.504 | .535 | .889 | 1 | .346 | .604 | .212 | 1.723 |
| [Untreated SGTF] | 0[b] | . | . | 0 | . | . | . | . |

Note. The reference category was PRIQ ≥ 85. *Significance < .05. SGTF=suboptimal gestational thyroid function, B=beta, df=degrees of freedom.

## 1.4.6. Discussion

The aim of this chapter was to test the null hypothesis that there would be no difference between the three groups of participants (those who were treated during their pregnancies for SGTF, those who were not treated, and those who had normal GTF; total n = 452) for childhood IQ measures at a mean age of 9 years 7 months. Based on $p = .353$, we failed to reject this null hypothesis. Therefore, no specific differences in the IQs were found between the normal GTF and untreated SGTF group, as well as the treated to untreated SGTF groups.

Percentages of participants with IQs ≤ 85 were also computed as a secondary, exploratory analysis. It was found that the untreated SGTF group scored ≤ 85 more frequently than the normal GTF group, and it can be inferred that with much larger participant groups significance may have been found. The most notable chi-square statistic was on the PRIQ domain. The results were extremely close to reaching significance, with the untreated SGTF group achieving a much greater number of IQ scores ≤ 85 compared to the treated SGTF group and the normal GTF group. These percentages for the PRIQ are the closest to those predicated by Haddow et al. (54). The PRIQ scores ≤ 85 were investigated further by a multinomial logistic regression, once adjusted for covariates, the effect was lost.

There were no significant differences found between the three groups of participants for the four models of specified analysis. One of the reasons for this may be that CATS II was underpowered. This was indicated by the small Partial Eta Square results. Furthermore, by looking for differences between the mean IQs as a continuous outcome, a six IQ point difference would have been needed to reach significance (as based on original CATS II study calculations to identify sample sizes for the project (280)). There could be many explanations why a non-significant result was identified here in contradiction to many other studies that have found an underactive thyroid to have a detrimental effect to a child's intelligence (34, 49-57, 90, 91). The following discussion is around the reliability of intelligence testing and the fluidity of the concept of intelligence itself and explored the mean IQs identified in the current dataset.

### 1.4.6.1.     IQ testing challenges, in the context of CATS II

IQ testing has been widely criticised in the literature for not being a good representation of an individual's capabilities for a number of reasons, these all tend to focus around the fact that an IQ test, *is* a test, and thus occurs at a single time point. The WISC-IV, like numerous other assessments, only measures how a child performs on a particular day at a particular time (a 'snapshot' of a child's development (78)). The child's motivation to complete the IQ test is also an important factor to consider as those less motivated to complete the assessment would not have performed as well compared to those who were (227-229). Intelligence tests tend to rely on the notion that the assessment was the most important thing the individual had to do at that particular point in time (284). 'Stage-fright' for the child was also important to consider; when beginning the WISC-IV, it was stated how establishing and maintaining a rapport with the child was key to elicit cooperation and effort during testing (78). Rapport was always sought before all of the assessments were conducted. This was easier with participants who were seen in the research centre as there was more time for the child to see me, the examiner, whilst other data collection aspects for the project were underway, e.g. the bone density scans. When the child was assessed in at their home, a discussion about the child's day and/or upcoming festivities was used to help build an initial examiner-examinee relationship. Those children who kept their concentration levels high, inevitably had quicker assessments as less time was spent attempting to keep them on-task and motivated. This point was eloquently illustrated by Young (285) [pp. 105],

"…intelligent children were those who were obedient…"

At all times, use of the words 'intelligence' and 'intelligent' were avoided as it could have left the child feeling anxious (286). Trying to ease any anxious tendencies for the individuals was important, as Moutafi (287) found that under anxiety-provoking instructions, individuals performed significantly worse than participants in a non-stressful condition. Having an additional adult viewing a cognitive assessment would not induce optimal testing (216). This was pertinent to the home assessments as frequently parents would opt to 'watch' the testing, therefore 'where the child was assessed' was viewed as a potential covariate and controlled for.

As well as the problems of the child sitting to take the assessment, there are also general problems of the IQ tests themselves. Intelligence tests evaluate an individual's ability to do or say something meaningful during the assessments (288), thus it could be argued that these tests are verbally dependent. Furthermore, an individual could score different ranges of IQs dependent on which IQ test was administered (289). Kranzler and Floyd (290) agree with Kaufamn (289) and discussed how different intelligence tests can produce a range of results. As an individual could display discrepancies between IQ tests; revisiting the definition of 'intelligence' would be beneficial. However, there is no widely accepted definition of intelligence. It can be summarised then, that it would be difficult to design a test to assess a, somewhat, fluid concept; let alone being able to examine an assessment's statistical validity (291). Fiedman et al. (292) argued and evidenced that some executive functions (processes that control and regulate thought and action) were not measured by traditional IQ tests. Intelligence testing is (and has been) closely correlated with academic performance, but conscientiousness and openness do play a role also (293, 294) and are not accounted for in IQ testing. To try and overcome these problems, the WISC-IV was selected as it is one of the most commonly used IQ measures (234-238), however as with any cognitive measure, it carries limitations. In intelligence tests, an individual's 'IQ' is observed, recorded and scored; whereas intelligence is one's performance and has been argued to be unobservable (227). Young (285) discusses how using record forms during the testing measures the outcome result rather than the process of the assessment, i.e. how the individual is performing; that it quantifies an individual. Whilst this is beneficial for research, more in-depth assessments may be required if investigating an individual's needs.

As discussed, during the blinded cleaning of the WISC-IV dataset, an outlier was removed. This individual had four IQ z-scores < -3 and was clearly observable on histograms of the CATS II cohort (see Figures 7 and 8). The WISC-IV had a floor effect in so far as it does not allow IQs below 40 (78). If an individual's raw score was zero for a subtest, the scaled score became

one (295) which could have masked a floor effect; this occurred in the removed participant. During the Symbol Search subtest (for PSIQ), the examinee was required to select 'yes' or 'no' depending on whether a target symbol was located in a fixed number of presented symbols. If a child scored an incorrect 'yes' or 'no' on an item, the score was taken away from the total correct answers (e.g. if an individual marked 15 correct and 1 incorrectly, the total raw score would have been 14). The removed participant answered extremely quickly and selected 'yes' or 'no' without looking at any symbols- this was evident by the speed in which the task was completed. Consequently, this individual ended up with a minus score (as more selections were incorrectly marked than correctly marked). However, the raw score had to be reported as zero as the WISC-IV was not equipped for taking minus scores. Orsini et al. (296) found that this hidden floor effect could artificially increase an individual's IQ, they summarised that those with IQs around the score of 40 may in fact be overestimates of that persons true ability.

### 1.4.6.2.    CATS II mean IQs

As can be seen in Table 14, the mean IQs for the groups were all close to the desired normative mean of 100. As the WISC-IV was UK standardised over a decade ago, it could be argued that the means in CATS II were subject to the Flynn effect (see chapter 2.2., IQ comparisons between ages 3 and 9; children from the CATS sample, for further discussion). The IQ that appeared to be 'pulling away' from the others was the PRIQ (normal GTF 105, treated SGTF 104, untreated SGTF 105 and total group 105). Higher IQs in non-verbal (perceptual reasoning) compared to verbal were first identified in 1942 (297). It was recognised that over the course of a decade, there would be a six IQ point increase for non-verbal, and only 2.6 point increase for verbal; this had been confirmed in many subsequent studies (222, 298-300). By looking at multiple studies, it was concluded that in the UK between 1979 to 2008, fluid IQ (nonverbal and reasoning) increased, but crystallised IQ (verbal and educational abilities) stayed the same (301). A proposed reason for this was the 'nutrition theory' (71, 299, 302-306). This theory posits that the advancements of nutrition impacts on the foetus and also on the younger child when their brain is developing and growing. This explains why fluid intelligence has increased and verbal has not, and can help to explain why a higher PRIQ was found in the CATS II study as compared to the VCIQ.

On an individual administration level, the PRIQ to VCIQ differences were difficult to explain and quantify as the tests that were administered to measure an individual's PRIQ are somewhat "black and white" in nature. For example, it was the same preliminary instructions given before each test and scoring was either zero or one. VCIQ would be open to more

scrutiny however as the scoring was based on the examiner's judgement of answers given. The slightly low verbal scores may be due to a portion of the children in CATS II attending welsh schools (16.2%) and some speaking welsh at home (2.4%).

Even though no significant differences were found for any of the IQs between the groups, it was visible in Table 14 that the means were consistently higher for the normal GTF group (except for WMIQ). These differences were slight, and may have reached significance with a much larger cohort. Looking closer at the differences between the SGTF groups, we can see that there was slight variance; the most notable was for PSIQ. WMIQ and FSIQ had virtually no difference between SGTF groups; for VCIQ and PRIQ treated SGTF group had slightly lower results compared to the untreated SGTF group. These results support CATS I showing that treatment had no effect on childhood cognition. One of the reasons for this may have been, as discussed, that the IQ tests were in some way flawed. However, if this was the case how have previous studies identified a difference in intelligence measured? It appears more likely, that treatment may not have been given at the optimal timing and perhaps initiated too late (200), or it may not have been complied with in some cases.

### 1.4.6.3. Limitations

One of the limitations of this data was that the findings were from small groups and thus were not generalisable to the general population. From the outset, the CATS II study design had attempted to avoid bias by having the cognitive assessments blinded by participant group, I continued this by also blinding myself during the data cleaning phase. Bias was introduced in other ways however, namely age of child when participating in CATS II ($p$ = .002: see Table 13 above and chapter 2.1., Significant effects from the covariates; CATS II data, for further information). A possible reason for this may have been because of the way in which the children from the normal GTF group were recruited into CATS II. There were 15,744 possible participants to approach for the 240 required for the normal GTF group. Participants were contacted by year of registration into CATS I and invited to take part in the research (280). Participants from both SGTF were recruited in smaller groups as and when they were needed, thus not based on their age like the normal GTF children. This method of recruitment may explain the age differences, but does not help with understanding why the difference was only significant between the treated SGTF group and the normal GTF group ($p$ = .004). However, this would not define the IQ results in CATS II as the WISC-IV will have taken account of child age, by year and four month intervals.

The WISC-IV used in CATS II was the most recent version available at the time, but it was still a decade since standardisation and therefore could be regarded as out of date. On an assessment level, this was apparent on a few images in picture concepts (e.g. black board, stamp, type writer) and one of the questions from the comprehension test (referred to stamps on letters). For picture concepts, this was overcome by prompting the child that they could ask what an image was if they were not sure; this will not have affected the validity of the test as it examines the relationships between the objects presented, rather than what the objects were. Furthermore, it was stipulated in the administration manual that the examiner could have named the pictures if requested (78).

### 1.4.6.4.    Conclusions

This results chapter has answered two (bold) of the four research aims:

i. **By reassessing the children at the older age of 9, would there be a continued non-significant difference identified for intelligence.**

ii. Would there also be non-significant findings at age 9 in other potential areas of cognition.

iii. **As there were no differences between the treated and untreated SGTF groups at age 3, would any differences to the normal GTF group be measurable; as there is a wealth of studies displaying that an underactive thyroid during pregnancy does not affect a child's cognition (53, 56, 91-93, 95, 96).**

iv. Would there also be non-significant differences between the groups that extend beyond cognition, i.e. behaviour.

The results reported identified that thyroid function during pregnancy did not have a negative impact on childhood IQ; there were no significant differences between intelligence scores between offspring born to the normal GTF, treated or untreated SGTF groups. This was contradictory to some literature available, but was in keeping with CATS I that also found no significant differences between children for IQ measured at an earlier time point. As there was no difference found between results of CATS I and CATS II, this added validity to the cognitive assessments conducted (CATS I incorporated 2 examiners, and I was the only examiner for CATS II). A secondary analysis exploring the IQs ≤ 85 revealed no differences between the three groups. A repeated study with a larger population might have yielded different results, however CATS I was not underpowered to see effects.

### 1.4.7.    Chapter Summary

This chapter has discussed the main findings from the WISC-IV IQ testing on the three groups of participants; children born to mothers that were treated for their SGTF, untreated for their SGTF and those born to mothers who had a normal GTF. No significant differences were

identified between any of the three groups on any of the five IQ scores. Significance was almost achieved for the PRIQ domain between the groups for a score ≤ 85, but caution was advised interpreting this result as no adjustments for the covariates were taken for this calculation. This chapter covered the hypotheses for the data, details of the WISC-IV and how the children were assessed, how the data was cleaned, statistical analysis, the results and the discussion of the results in the context of the IQ test. The discussion also included a section on the criticisms noted in the literature of IQ testing as a whole, as it is important to consider these when interpreting the main findings. The following chapter details the additional cognitive assessments from the NEPSY-II; long term memory, working memory and fine motor coordination.

## 1.5. Additional cognitive assessments at age 9; CATS II data

### 1.5.1. Chapter Overview

This chapter contains the data arising from the additional cognitive assessments administered in CATS II. There is a brief introduction and methods overview (further information on the development of the assessment tool can be found in chapter 1.1. (General introduction) and chapter 1.3. (Methods for the cognitive and behavioural data collection for the CATS II study). Information about the participants, demographics for the three groups and reasons for the differing numbers of completed subtests, are explained. The four model statistical analysis is highlighted in the results section. Finally, the chapter closes with a discussion including limitations and main conclusions. The data presented in this chapter was also used in the main CATS II project; however, the statistical package and analysis selected were different to the project to emphasise the separate work undertaken for this thesis.

### 1.5.2. Introduction

As well as effects on intelligence for offspring born to mothers that had an underactive thyroid during their pregnancy, there was literature available that also investigated additional possible deficits on specific cognitive domains; memory and motor-coordination (27, 34, 49-51, 54, 92, 97, 103-106, 110-113, 307-312).

The aim of the work presented in this chapter was to investigate whether being born to a mother who had normal GTF, was treated for SGTF or was untreated for SGTF, would have had an effect on specific cognitive domains. The specific cognitive domains were working memory, long term memory and fine-motor coordination. Further deficits in reading ability (49, 53) and hearing loss (91, 124) were not investigated as part of CATS II because of time constraints of the assessment structure.

Similar to the IQ hypotheses, it was predicted that there would be a cognitive deficit for the untreated SGTF group compared to the normal GTF group (as supported by (10, 32, 49, 54-58)). Also, the treated SGTF group and untreated SGTF would have non-significant differences between their scores; similar to the CATS I IQ findings (200) and chapter 1.2., Re-analysis of intelligence at age 3; UK CATS I cohort.

### 1.5.3. Method

Similar to the WISC-IV, a total of 480 participants were required to address the null hypothesis that there would be no difference of mean scores between the additional cognitive measures and all three participant groups. The NEPSY-II (230) was adopted to

explore any possible differences between long term memory, fine motor coordination and also further working memory tests amongst the participants. I was blinded to the participant group for the NEPSY-II assessments as the NEPSY-II followed the WISC-IV testing. Subtests were administered in the same order for participants and as the NEPSY-II was placed second it was as a whole (or specific tests were) subject to being 'dropped' dependent on the child's fatigue, testing environment issues etc. If a subtest (or subtest*s*) were removed, this did not affect the order of the remaining subtests.

The NEPSY-II was standardised for children between the ages of ≥ 3 years 0 months to ≤ 16 years 11 months. The complete NEPSY-II included 32 tests and assessed children across six domains: attention and executive functioning, language, memory and learning, sensorimotor, social perception and visuoperceptual processing. From the literature review, it was decided to assess across the memory and language, and sensorimotor domains. Similar to the WISC-IV administration, hints or helps were prohibited whilst assessing.

Administration time of the selected five subtests (four memory, one sensorimotor) was about 30 minutes on average; in total the cognitive assessments would take around an hour and a half for participants. The order of the tests was list memory, memory for designs, fingertip tapping, narrative memory, memory for designs delayed, and finished with the delayed recall from list memory. See Table 18 for further information of each subtest, as organised by their respective domains.

Table 18:

*Descriptions of the Developmental Neuropsychological Assessment Second Edition (NEPSY) Subtests Used in the Controlled Antenatal Thyroid Screening study II*

| NEPSY-II Domain | Subtest | Description |
|---|---|---|
| Memory and Language | List memory and list memory delayed (LM) | The child was presented with a list of 15 words and asked to immediately recall the words. After a delay of 25-35 minutes, the child was asked to recall the 15 words from memory |
| | Memory for designs (MD) | The child was shown a grid with four to ten designs on a page, and then it was removed. The child would then be required to select designs from a set of cards and place the cards in a grid in the same location as shown to them. |
| | Memory for designs delayed (MDD) | 15-25 minutes after MD was administered, the child would attempt the final grid again. |
| | Narrative memory (NM) | The child was required to listen to a story and then repeat the story back. Prompt questions were administered for any missing information. The subtest would finish with recognition questions. |
| Sensorimotor | Fingertip tapping dominant hand & non-dominant hand (FTDH & FTNDH) | The first trial was in the dominant hand and required the individual to tap the tip of their index finger on their thumb as quickly as they could 20 times. This was then repeated for the non-dominant hand. Next, the child would have to tap their index fingertip against their thumb, middle, ring and little fingertip in the dominant hand followed by their non-dominant hand. |

Note. LM=list memory, MD=memory for designs, MDD=memory for designs delayed, NM=narrative memory, FTDH=fingertip tapping dominant hand, FTNDH=fingertip tapping non-dominant hand.

The NEPSY-II generated six scaled scores for comparison. The raw scores were calculated first by hand, then using age-derived scales, scaled scores were finally calculated. Similar to the WISC-IV scaled scores, a mean was 10 within the range of 1-19, and SDs were ± 3 points from the mean. No feedback of the NEPSY-II was made available to parents. As there was no feedback, no percentile ranks were calculated for the NEPSY-II as the scores were for the project purposes only.

### 1.5.4. Statistical Analysis

The NEPSY-II was cleaned blind similar to the WISC-IV. As the NEPSY-II was administered secondary to the WISC-IV, if there were any administration issues (such as environment problems, child fatigue issues, child requesting to stop etc.) the NEPSY-II or some of its tests were omitted to ease testing. For example, if the child showed signs of great fatigue following the WISC-IV, the NEPSY-II results would not have been a true reflection of that child's ability.

Furthermore, if the testing environment contained multiple distractions, this would have also affected the results. Therefore, only 416 children completed the WISC-IV and aspects of the NEPSY-II (see Table 20 for breakdown).

In total, there were 36 participants who did not complete *any* of the NEPSY-II subtests. I made qualitative notes for all of these participants on the data set, and reasons for non-completion of the NEPSY-II fell into eight broad themes:

1. Long WISC assessment (any over 1 hr and 20)                n=14
2. Mother requested quick visit                               n=7
3. Unsuitable test environment                               n=4
4. Child ill                                                 n=3
5. Child was difficult to assess                             n=3
6. Two families attended the research centre the same day    n=2
7. Child refused NEPSY and requested to finish cognitive testing   n=2
8. Child distressed                                          n=1

Along with the 14 participants who were noted as having long WISC-IV assessments, nine of these also included the child appearing visibly tired (n = 5) or the child wanting to finish and not being able to sit and concentrate (n = 4). In particular, had these 14 completed a NEPSY-II assessment it may have included a large measurement error (the difference between an individual's true score and the individual's obtained score (230)). There were two further participants who generated small amounts of data from the NEPSY-II: one who only completed FTDH and FTNDH and one other who completed both fine motor coordination tasks and NM only. In both cases the mothers requested to finish the assessments early.

Ranges of scores were initially checked and searches for missing values were executed. As covered by the WISC-IV cleaning, the following demographic information had already been checked and thus was ready for the secondary cognitive analysis: how participants took part in CATS II, child gender, where the child was assessed, age of child at cognitive assessment, child's language at school and home, age quartiles for the mother, whether the child was breastfed over one month and also the social deprivation score quintiles (as calculated by StatsWales (281) and OpenDataCommunities (282)).

As explained in chapter 1.4. (Intelligence measured at age 9; CATS II data), nine individuals were removed from the complete dataset; one as an outlier, two as they were twins, one child for being aged ≤ 7 years old and finally, five children who were ≥ 11 years old. Similar to the WISC-IV data cleaning procedures (see chapter 1.4., Intelligence measured at age 9; CATS II data), z-scores were computed for target items for the NEPSY-II. It was identified that

all participants for LM, MD and MDD achieved z-scores between -3 to +3. On the FTD test, three individuals had z-scores < -3. Likewise, on FTND there was one individual < -3 and NM observed three participants with scores < -3. None of these individuals presented low z-scores on more than one test measure, had outside normal z-score ranges on any of the WISC-IV items, or had z-scores > 3. The individual from the WISC-IV who presented extremely low scores on four out of the five IQs was the only removed participant based on z-scores, and the decision was made *not* to remove others who were outside normal ranges on only one IQ measure. Likewise, with the NEPSY-II, it was argued not to remove these seven individuals as that would mean looking at domains separately rather than a whole, i.e. an individual scoring a < -3 z-score on VCIQ was not removed, so then, why should a different individual be removed for low scoring on FTD/FTND tasks?

All Kolmogorov-Smirnov (and Shapiro-Wilk) normality tests were returned as all *p*'s < .001 for all additional NEPSY-II tests. However, means and medians appeared close and skewness and kurtosis ranges were all within the -1 - +1 range and thus were normal (see Table 19 below). The NEPSY-II data was analysed by parametric tests as the variables were continuous, and based on the means, medians, skewness and kurtosis the data was accepted as being normally distributed.

Table 19
*Means, Medians, Skewness and Kurtosis for the Developmental Neuropsychological Assessment (NEPSY)-II Tests*

| NEPSY-II Subtest | Mean | Median | Skewness | Kurtosis |
|---|---|---|---|---|
| NEPSY-II List Memory and List Memory Delayed Scaled Score | 10.79 (2.85) | 11.00 | -.134 | -.338 |
| NEPSY-II Memory for Designs Total Scaled Score | 10.01 (3.05) | 10.00 | -.382 | -.512 |
| NEPSY-II Memory for Designs Delayed Total Scaled Score | 10.09 (2.74) | 10.00 | .216 | -.120 |
| NEPSY-II Fingertip Tapping Dominant Combined Scaled Score | 12.11 (1.54) | 12.00 | -.356 | -.032 |
| NEPSY-II Fingertip Tapping Non-Dominant Combined Scaled Score | 12.39 (1.40) | 12.00 | -.211 | -.675 |
| NEPSY-II Narrative Memory Free and Cued Recall Scaled Score | 11.32 (2.76) | 12.00 | -.367 | .239 |

Note. Standard deviations appear in parentheses below means.

Table 20

*Developmental neuropsychological assessment (NEPSY)-II participant group demographics*

| | | Normal GTF (n=219) | Treated SGTF (n=109) | Untreated SGTF (n=88) |
|---|---|---|---|---|
| Participated (n) | Research centre (n=assessed at home) | 180 (79) | 74 (51) | 41 (26) |
| | Remote/Home | 39 (17.8%) | 35 (32%) | 47 (53%) |
| Total home visits (n) | | 118 (53.9%) | 86 (78.9%) | 73 (82.9%) |
| Mean Age (yrs) | | 9.67 | 9.41 | 9.44 |
| Gender (n) | Male | 109 | 59 | 41 |
| | Female | 110 | 50 | 47 |
| Child's language at school and home (n & %) | English School and English at Home | 169 (77.2%) | 85 (78%) | 76 (86.4%) |
| | Welsh School and English at Home | 41 (18.7%) | 20 (18.3%) | 11 (12.5%) |
| | Welsh School and Welsh at Home | 7 (3.2%) | 3 (2.8%) | 0 |
| | English School and Other Language at Home (not Welsh or English) | 1 (0.5%) | 1 (0.9%) | 1 (1.1%) |
| | Welsh School and Other Language at Home (not Welsh or English) | 1 (0.5%) | 0 | 0 |
| Whether the mother breastfed for over 1 month | Yes | 139 (63.8%) | 67 (62%) | 48 (54.5%) |
| | No | 79 | 41 | 40 |
| | Missing | 1 | 1 | 0 |
| Mother age at time of consent into CATS I Mean of quartile (mean, years) | | 2.5 (31.72) | 2.2 (30.40) | 2.3 (30.74) |
| Social deprivation score (mean of quintile) | | 3.7 | 3.8 | 3.3 |
| NEPSY-II Subtests (n & missing n) | LM | 170 (49) | 77 (32) | 70 (18) |
| | MD | 218 (1) | 109 | 87 (1) |
| | MDD | 218 (1) | 109 | 87 (1) |
| | FTDH | 219 | 109 | 88 |
| | FTNDH | 218 (1) | 109 | 88 |
| | NM | 218 (1) | 109 | 87 (1) |

Note. Further information appears in parentheses following n's were applicable. SGTF=suboptimal gestational thyroid function. CATS=controlled antenatal thyroid screening, LM=list memory, MD=memory for designs, MDD=memory for designs delayed, NM=narrative memory, FTDH=fingertip tapping dominant hand, FTNDH=fingertip tapping non-dominant hand.

Table 20 shows the group demographics of each participant group. As mentioned in chapter 1.4. (Intelligence measured at age 9, CATS II data), the untreated SGTF group consisted of far more home visits and assessments at home compared to the treated SGTF and normal GTF groups. 'Participant group' and 'where the child was assessed' were compared by a Chi-square and confirmed to be significantly different; $X^2$ (2, n = 416) = 33.913, $p$ < .001. The ages of the children were compared by ANOVA, and alike to the findings in the previous chapter, a significant difference persisted: $F$ (2, 413) = 5.998, $p$ = .003, $\eta_p^2$ = .028. However, the normal GTF group were significantly older than both the treated ($p$ = .007, 95% CI [.0565, .4595]) and untreated ($p$ = .043, 95% CI [.0051, .4390]) SGTF groups. This will not have affected the NEPSY-II scores, as like to the WISC-IV this assessment battery also takes account of age. The other covariates were similar to what was discussed in the chapter 1.4. (Intelligence measured at age 9; CATS II data). By controlling for six covariates, I have attempted to overcome potential biased effects of recruitment.

The NEPSY-II data was analysed in IBM SPSS version 20 by the four models of analysis proposed in the published CATS II protocol (280) and chapter 1.3. (Methods for the cognitive and behavioural data collection for the CATS II study). As can be seen in Table 20, there was a shortfall in LM tests across all three groups of participants. One of the reasons for this deficit was that the LM test was introduced later in the data collection phase. Furthermore, it was placed at the beginning (working memory aspect) and at the end of the NEPSY-II assessments (the long term memory aspect), and thus was easy to exclude if the child was not coping well with the long battery of assessments. If the data was analysed in a grouped MANCOVA, this would have only included those individuals who participated in every subtest of the NEPSY-II. As there was a large amount of data missing for the LM subtest, this would have yielded a small cohort; and even smaller participant groups. Therefore, the decision was made to analyse the LM test by a separate Analysis of Covariance (ANCOVA) and the remaining five subtests by a larger MANCOVA. Any significant differences at the multivariate level were followed up with post hoc tests to investigate where the significant differences were, and to control for multiple testing; a Bonferroni correction was chosen. The different models for the analysis were the same as the WISC-IV (see Figure 6).

Similar to the CATS I UK only cohort analysis (chapter 1.2., Re-analysis of intelligence at age 3; UK CATS I cohort) and WISC-IV analysis (chapter 1.4., Intelligence measured at age 9; CATS II data), percentages of participants scoring ≤ 1 SD from the mean (≤ 7) were computed and tested by Chi-Square for significance as a secondary, exploratory analysis. This was executed to allow comparisons to be drawn from the main CATS II findings to the work completed for

this thesis. Any significant values were followed up by a multinomial logistic regression to allow for adjustments of the six covariates: child gender, mother age at time of consent into CATS I, whether the mother breastfed over one month, where the child was assessed, language at school and home, and social deprivation. This was explored in CATS I as there was an attempt to replicate the findings from Haddow et al.'s (54) benchmark study which identified significantly more children with IQ scores below 1 SD if they were born to mothers with high TSH compared to those with normal thyroid function. It was predicted that individuals from the normal GTF groups would have 16% with scaled scores ≤ 7 (as anticipated by the NEPSY-II (230)), whereas those in the treated SGTF group would contain 5% of scaled scores ≤ 7 compared to the untreated SGTF group achieving 15% (this predication was based on the work by Haddow et al. (54)).

### 1.5.5. Results

#### 1.5.5.1. General attendance information

As stated in the chapter 1.3. (Methods for the cognitive and behavioural data collection for the CATS II study), CATS II aimed to recruit a total of 480 participants back into the study. This consisted of 120 from the treated and untreated SGTF groups, and 240 from the normal GTF group. The WISC-IV analysis contained a final cohort of 452 individuals. As previously stated, the NEPSY-II was dropped in 8% of cases and thus a reduced dataset was generated. The treated SGTF group fell short by 9.2% of target (n = 109), the untreated group were 26.7% behind the recruitment target (n = 88) and the normal GTF were 8.8% below target (n = 219). See Table 20 for group demographic information.

#### 1.5.5.2. Analysis

Table 21 shows the means and SDs for all study NEPSY-II scores and was adjusted for by the fourth model of analysis (see Figure 6). The graph (Figure 13) displays the means achieved by the groups (unadjusted model); error bars have also been included.

Table 21

*Developmental Neuropsychological Assessment (NEPSY)-II Means by Study Group and Model Four of Analysis*

|  | Participant Group | Mean | N |
|---|---|---|---|
| NEPSY-II List Memory and List Memory Delayed Scaled Score | Normal GTF | 10.94 (2.837) | 170 |
|  | Treated SGTF | 10.53 (3.079) | 76 |
|  | Untreated SGTF | 10.70 (2.628) | 70 |
|  | Total | 10.78 (2.849) | 316 |
| NEPSY-II Memory for Designs Total Scaled Score | Normal GTF | 10.36 (2.916) | 215 |
|  | Treated SGTF | 9.56 (3.259) | 108 |
|  | Untreated SGTF | 9.85 (2.979) | 87 |
|  | Total | 10.04 (3.036) | 410 |
| NEPSY-II Memory for Designs Delayed Total Scaled Score | Normal GTF | 10.34 (2.646) | 215 |
|  | Treated SGTF | 9.74 (2.793) | 108 |
|  | Untreated SGTF | 10.00 (2.889) | 87 |
|  | Total | 10.11 (2.743) | 410 |
| NEPSY-II Fingertip Tapping Dominant Combined Scaled Score | Normal GTF | 12.24 (1.596) | 215 |
|  | Treated SGTF | 11.88 (1.406) | 108 |
|  | Untreated SGTF | 12.02 (1.532) | 87 |
|  | Total | 12.10 (1.539) | 410 |
| NEPSY-II Fingertip Tapping Non-Dominant Combined Scaled Score | Normal GTF | 12.51 (1.370) | 215 |
|  | Treated SGTF | 12.19 (1.391) | 108 |
|  | Untreated SGTF | 12.31 (1.465) | 87 |
|  | Total | 12.38 (1.399) | 410 |
| NEPSY-II Narrative Memory Free and Cued Recall Scaled Score | Normal GTF | 11.56 (2.764) | 215 |
|  | Treated SGTF | 10.94 (2.741) | 108 |
|  | Untreated SGTF | 11.21 (2.783) | 87 |

|        | Total | 11.32 | 410 |
|        |       | (2.768) |   |

Note. Standard deviations appear in parentheses below means. SGTF=suboptimal gestational thyroid function.
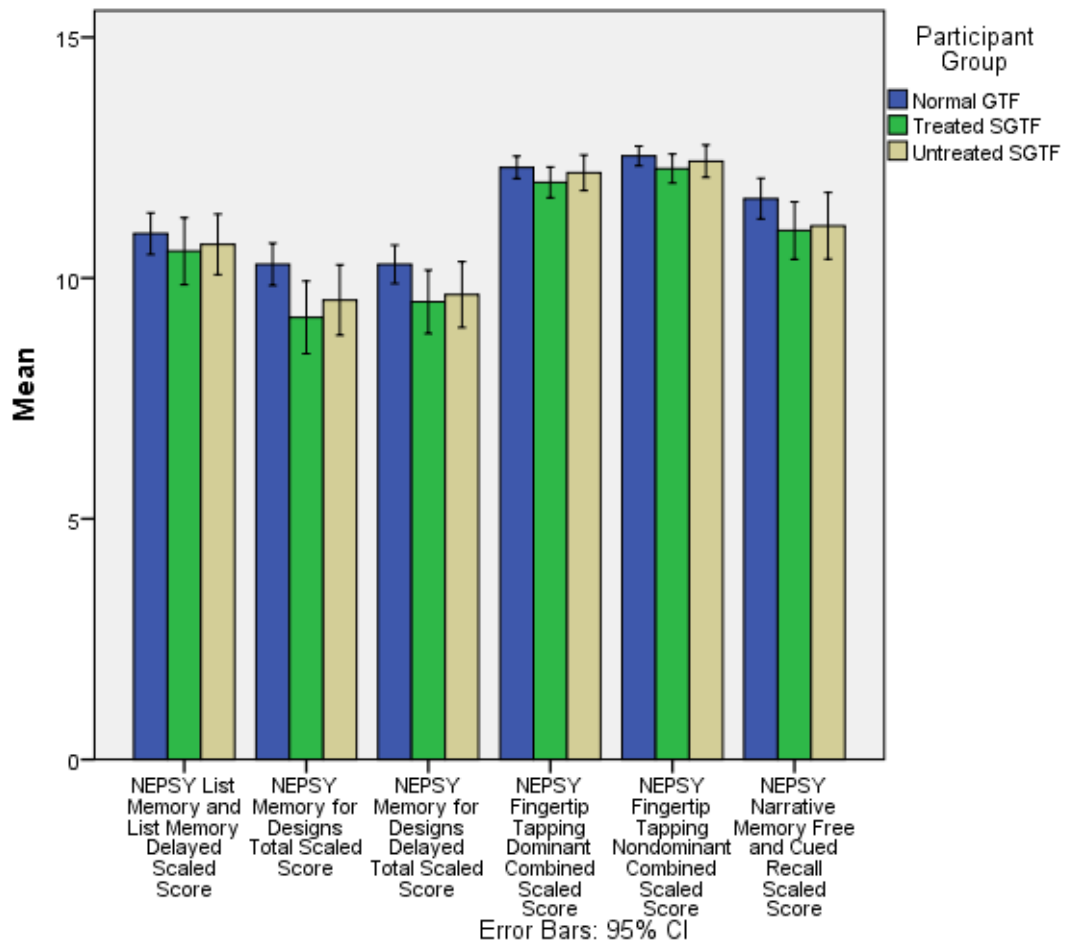


Figure 13: Means per group for scores obtained on the Developmental neuropsychological assessment (NEPSY)-II

CI=confidence interval. X axis=NEPSY-II subtest, Y axis=mean of scores per participant group.

The LM test was analysed by an ANOVA for the unadjusted, first model of analysis (see Figure 6). It was found that there were no significant differences between the three participant groups, $F (2, 314) = .482$, $p = .618$, $\eta_p^2 = .003$. By adjusting for child gender (model two), the ANCOVA was still non-significant, $F (2, 313) = .498$, $p = .608$, $\eta_p^2 = .003$. Model three of the analysis and thus controlling for model two and age of mother at recruitment into CATS I and whether the child was breastfed, it was also found that there was no significant differences between the groups, $F (2, 310) = .354$, $p = .702$, $\eta_p^2 = .002$. The final model of analysis controlled for model two, three and the child's language at school and home, where the child

was assessed and the social deprivation of the family, significance was still not achieved: $F$ (2, 307) = .489, $p$ = .613, $\eta_p^2$ = .003. Therefore, by all models of analysis, there were no significant differences between the groups on the joint working memory and long term memory task: LM; the covariates made little differences to the probability values.

There were no significant differences between the normal GTF, treated SGFT and untreated SGTF groups on the other five NEPSY-II measures, all $p$'s > .078. For the unadjusted, first model of analysis, MANOVA, using Roy's largest root, there was a non-significant effect between groups for NEPSY-II measures, $\Lambda_{ROY}$ = .025, $F$ (5, 406) = 1.999, $p$ = .078, $\eta_p^2$ = .024. Model two yielded a similar non-significant MANCOVA result: $\Lambda_{ROY}$ = .024, $F$ (5, 405) = 1.973, $p$ = .082, $\eta_p^2$ = .024. Model three also generated a non-significant result $\Lambda_{ROY}$ = .021, $F$ (5, 401) = 1.660, $p$ = .143, $\eta_p^2$ = .020. Finally, model four which controlled for all covariates also produced a non-significant result between the three groups of participants: $\Lambda_{ROY}$ = .018, $F$ (5, 398) = 1.431, $p$ = .212, $\eta_p^2$ = .018. As no significant differences were revealed by ANOVA/ANCOVA and MANOVA/MANCOVA, no discriminate analyses were undertaken.

The secondary analysis involved analysing scaled scores per group scoring ≤ 7 and tested for significance by use of a Chi-Square. Results can be found in Table 22 below. The only result that achieved significance was the MD subtest. This was explored further by a multinomial logistic regression.

Table 22
*Scaled Scores ≤ 7 (%) from the Developmental Neuropsychological Assessment (NEPSY)-II*

|  | Scaled Scores ≤ 7 (%) | | | | | |
|---|---|---|---|---|---|---|
|  | LM | MD | MDD | FTDH | FTNDH | NM |
| Normal GTF | 18/170 (11%) | 42/218 (19%) | 32/218 (15%) | 2/219 (1%) | 1/218 (0.5%) | 16/218 (7%) |
| Treated SGTF | 16/77 (21%) | 36/109 (33%) | 25/109 (23%) | 0/109 (0%) | 0/109 (0%) | 10/109 (9%) |
| Untreated SGTF | 9/70 (13%) | 17/87 (12%) | 16/87 (18%) | 1/88 (1%) | 0/88 (0%) | 9/87 (10%) |
| Pearson Chi-Square | $p$=.094 | $p$=.014* | $p$=.178 | $p$=.572 | $p$=.636 | $p$=.662 |

Note. *Significance < .05. Percentages of scores per group are appear in parentheses below totals. LM=list memory, MD=memory for designs, MDD=memory for designs delayed, NM=narrative memory, FTDH=fingertip tapping dominant hand, FTNDH=fingertip tapping non-dominant hand, SGTF=suboptimal gestational thyroid function.

A multinomial logistic regression was executed to explore the differences between the groups for the subtest MD. The regression, controlled for all six covariates that were adjusted for in the MANCOVA, revealed that those offspring from the treated SGTF group were 2.257 times more likely to achieve a scaled score for the MD subtest ≤ 7 compared to those from the untreated SGTF group. It was also apparent that breastfeeding had a significant effect, suggesting that if the child was breastfed there was a decreasing chance of scoring ≤ 7 MD subtest. See Tables 23 and 24 for further details of beta values, significance and -2 log likelihood.

Table 23

*Table Displaying the Regression Model's Fit for the Data*

| Model | Model Fitting Criteria | Likelihood Ratio Tests | | |
|---|---|---|---|---|
| | -2 Log Likelihood | Chi-Square | *df* | Sig. |
| Intercept Only | 361.038 | | | |
| Final | 341.965 | 19.073 | 8 | .014 |

Note. See improved figure for the -2 Log Likelihood.

Df=degrees of freedom.

Table 24

*Main Output from Multinomial Logistic Regression, Memory for Designs (MD) ≤ 7*

| MD ≤ 7 | B | Std. Error | Wald | df | Sig. | Exp(B) | 95% Confidence Interval for Exp(B) | |
|---|---|---|---|---|---|---|---|---|
| | | | | | | | Lower | Upper |
| Gender | -.060 | .242 | .061 | 1 | .805 | .942 | .586 | 1.514 |
| Where Assessed | -.467 | .280 | 2.773 | 1 | .096 | .627 | .362 | 1.086 |
| Language | -.204 | .240 | .726 | 1 | .394 | .815 | .510 | 1.304 |
| Mother Age | -.113 | .119 | .908 | 1 | .341 | .893 | .708 | 1.127 |
| Breast Fed 1mns | -.517 | .249 | 4.314 | 1 | .038* | .596 | .366 | .971 |
| Social Deprivation | -.007 | .089 | .006 | 1 | .938 | .993 | .835 | 1.181 |
| [normal GTF] | .181 | .335 | .293 | 1 | .589 | 1.199 | .621 | 2.314 |
| [treated SGTF] | .814 | .349 | 5.448 | 1 | .020* | 2.257 | 1.139 | 4.472 |
| [untreated SGTF] | 0[b] | . | . | 0 | . | . | . | . |

Note. The reference category was MD ≥ 7. *Significance < .05. SGTF=suboptimal gestational thyroid function, B=beta, df=degrees of freedom.

To enable to account for differences between the treated SGTF and the normal GTF groups, a further regression was executed. It was found that the normal GTF group were 1.899 times less likely to score ≤ 7 for the MD subtest compared to the treated SGTF group (p = .024). Further information for the regression model can be found in Tables 25 and 26.

Table 25

*Table Displaying the Regression Model's Fit for the Data*

| Model | Model Fitting Criteria | Likelihood Ratio Tests | | |
|---|---|---|---|---|
| | -2 Log Likelihood | Chi-Square | *df* | Sig. |
| Intercept Only | 343.982 | | | |
| Final | 327.911 | 16.070 | 7 | .024 |

Note. See improved figure for the -2 Log Likelihood.
Df=degrees of freedom.

Table 26

*Main Output from Multinomial Logistic Regression, Memory for Designs (MD) ≥ 7*

| MD ≥ 7 | *B* | Std. Error | Wald | *df* | Sig. | Exp*(B)* | 95% Confidence Interval for Exp(B) | |
|---|---|---|---|---|---|---|---|---|
| | | | | | | | Lower | Upper |
| Gender | .106 | .275 | .149 | 1 | .699 | 1.112 | .649 | 1.906 |
| Where Assessed | .542 | .308 | 3.091 | 1 | .079 | 1.720 | .940 | 3.147 |
| Language | .099 | .250 | .155 | 1 | .693 | 1.104 | .676 | 1.801 |
| Mother Age | .114 | .133 | .739 | 1 | .390 | 1.121 | .864 | 1.454 |
| Breast Fed 1mns | .474 | .282 | 2.832 | 1 | .092 | 1.606 | .925 | 2.790 |
| Social deprivation | .000 | .000 | .069 | 1 | .793 | 1.000 | .999 | 1.000 |
| [normal GTF] | .641 | .283 | 5.119 | 1 | .024* | 1.899 | 1.090 | 3.310 |
| [treated SGTF] | 0[b] | . | . | 0 | . | . | . | . |

Note. The reference category was MD ≤ 7. *Significance < .05. SGTF=suboptimal gestational thyroid function, B=beta, df=degrees of freedom.

### 1.5.6. Discussion

The aim of this chapter was to test the null hypothesis that there would be no difference between the three groups of participants (those who were treated during their pregnancies for SGTF, those who were not treated, and those who had normal GTF; total n = 416) for working memory, long term memory and fine motor coordination measures at a mean age of 9 years 7 months. We failed to reject the null hypothesis on all subtests from the NEPSY-II. Therefore, no specific differences were found between the normal GTF and untreated SGTF group, as well as the treated to untreated SGTF groups.

The secondary analysis of percentages of participants with scaled scores ≤ 7 was also computed. The normal GTF group scored ≤ the untreated SGTF group for LM, MDD, FTDH and NM. We could infer from this trend, that having SGTF during pregnancy could potentially

have a slight negative impact on the offspring across these domains. Furthermore, it could be inferred that with much larger participant groups, significance may have been found. The most interesting result was on the MD domain. The Pearson chi-square statistic reached significance with a large number of participants (33%) from the treated SGTF group scoring ≤ 7 for this working memory subtest on the NEPSY-II. The mean score for MD from the treated SGTF group was not too dissimilar to the untreated SGTF group. This was explored further by a multinomial logistic regression which confirmed that those from the treated SGTF were 2.257 times more likely to score ≤ 7 compared to the untreated SGTF group. Furthermore, the normal GTF group were 1.899 times less likely to score ≤ 7 for the MD subtest compared to the treated SGTF group ($p$ = .024). However, WMIQ scores ≤ 85 were non-significant (i.e. at a domain level) as well as the other WM subtests from the NEPSY-II, so this result for the MD subtest may be yielding a type one error; this has been reported as being commonplace when converting continuous variables to binary outcomes (213).

The findings of the NEPSY-II tests were contradictory to some research that has found an underactive thyroid having an effect on specific neuropsychological processes (long term and working memory (51, 97, 103-106, 307-312) and motor coordination (27, 49, 50, 54, 110, 112)), but also add to the validity of findings from other studies that found thyroid function did not affect these various aspects of cognition (34, 51, 92, 93, 111, 113, 200, 215).

The results identified from the NEPSY-II were comparable to the results from the WISC-IV in the previous chapter. As the analysis from the WISC-IV revealed no WMIQ differences between groups, it was not surprising then that no differences were found on the working memory tasks of the NEPSY-II (LM, MD and NM), as these two cognitive scales appear to support each other. Before the NEPSY-II was available, concurrent validity of its intellectual functioning was assessed by the WISC-IV. The correlation of the WISC-IV to other intellectual functioning measures suggested that the NEPSY-II was sufficiently predictive of cognitive performance for verbal and nonverbal domains (230).

The 'population' that comprises the current results chapter was not as well-ordered as the WISC-IV results. Testing was very much completed on an individual basis, i.e. the NEPSY-II items were *in addition* to the complete WISC-IV; thus aspects of, or the complete assessment were dropped in some instances to either ease the experience for the child and keep rapport high, the test environment conditions were insufficient or based on parental requests. No participants were removed based on variable z-scores (± 3). No participants had z-scores ≥ 3 for any subtests and no ≤ -3 scores were recorded for LM, MD or MDD tests. Three ≤ -3 z-

scores were recorded for FTDH, one ≤ -3 for FTNDH and a further three ≤ -3 for NM. The seven ≤ -3 scores were obtained by seven different participants and none of these seven had outlier z-scores on the WISC-IV. The decision was reached not to remove these seven participants as this was a pragmatic study to include a varying population. Further to this, the NEPSY-II was reported to have developed its floors and ceilings to encompass those of all abilities (255) (NB: floor effects were reported to arise when a test was too difficult and a large portion of the normative sample would then perform poorly, whereas ceiling effects occurred when a test was too easy for the normative sample (254)). Some subtests of the NEPSY-II were specifically developed to address the lower limits of ability at the youngest ages (230) which was an additional reason for keeping the seven participants in the main analysis.

The NEPSY-II assessment included multiple drawbacks. Some of the criticisms were similar to the WISC-IV assessment, such as being reliant on motivation, rapport, the individual's obedience and the effects of a 'snapshot' view, (see chapter 1.4., Intelligence measured at age 9; CATS II data, for further information). A further issue with the NEPSY-II was that it was not UK standardised. Kline (313) explored the importance of standardisation, even though the NEPSY-II was standardised, it may have benefitted from the UK cohort standardisation similar to the WISC-IV used in the current research. Further differences between these two cognitive measures was that unlike the WISC-IV with its' ten subtest order, the NEPSY-II was left free to the examiner to decide on order and test selection; this was aided by the NEPSY-II tests being alphabetically ordered in the scoring manual (254). The order that was decided for CATS II was guided by the time required for delayed tasks (LM and MDD).

### 1.5.6.1.    Limitations

With the NEPSY-II being placed after the WISC-IV, the children were often fatigued when the testing occurred. This led to some decisions to drop certain tests and therefore has led to a somewhat, fragmented dataset. With the NEPSY-II being 'tagged' on to the WISC-IV, this led to long assessment times for the participants: a couple reached just over two hours. As the NEPSY-II was always viewed as secondary to the WISC-IV for main outcomes, the WISC-IV was always viewed as the first priority. The incomplete datasets lead to the LM subtest being analysed in a separate ANCOVA and has left the research open to scrutiny because of multiple testing. A MANCOVA would have been more appropriate to use for all of the NEPSY-II subtests, however this would have meant losing statistical power from the dwindling participant groups. With separate testing, the small group numbers still made the research difficult to generalise to the wider population.

As mentioned, the NEPSY-II was not UK standardised and certain subtests within it were not re-normed from the original assessment in 1998. Even though the NEPSY-II used in the current research was developed in 2007, it was out of date and this was particularly visible in the FTDH and FTNDH. The mean scores for the groups, and cohort, were all well above the expected result of 10. With the means for both tests being at 12, this put the groups at a 75th percentile rank, instead of being placed around the norm of 50. The reasoning for such a high mean could be that the children were affected by their technology use, i.e. a high number of children now have access to mobile phones, computers, laptops, iPad/tablets etc., and thus must have had a proficient level of finger dexterity to enable them to adequately use these products.

### 1.5.6.2.    Conclusions

This results chapter has answered two (bold) of the four research aims:

i. By reassessing the children at the older age of 9, would there be a continued non-significant difference identified for intelligence.
ii. **Would there also be non-significant findings at age 9 in other potential areas of cognition.**
iii. **As there were no differences between the treated and untreated SGTF groups at age 3, would any differences to the normal GTF group be measurable; as there is a wealth of studies displaying that an underactive thyroid during pregnancy does not affect a child's cognition (53, 56, 91-93, 95, 96).**
iv. Would there also be non-significant differences between the groups that extend beyond cognition, i.e. behaviour.

Whilst the project tried to seek further answers from the IQ only CATS I study, investigations into long term memory, additional working memory and fine motor coordination appeared to have no mean significant differences between those from the normal GTF, treated SGTF or untreated SGTF groups. Larger studies would be needed to investigate the theory that CATS II was underpowered and that levothyroxine therapy was started too late for the treated SGTF group to be able to see any discrepancies to the untreated SGTF group.

### 1.5.7.    Chapter Summary

The current chapter has attempted to explore the possibility that there may have been further deficits from children exposed to SGTF. By investigating the three groups of participants, no significant differences were identified by use of the NEPSY-II subtests. There was, however, a significant difference between the treated and untreated SGTF groups on the MD subtest for scores below 1 SD from the mean; the treated group contained more children in this category. This was concluded as a type one error as no such differences were

identified on the other memory subtests of the NEPSY-II, or on the domain level in the WISC-IV. This chapter covered the hypotheses for the data, details of the NEPSY-II and why there were differing group sizes compared to the WISC-IV, data cleaning, statistical analysis, the results and the discussion of the results. The following chapter contains the final data that I collected for the CATS II project, and my analysis of it; the child behavioural questionnaires.

## 1.6. Behavioural questionnaires at age 9; CATS II data

### 1.6.1. Chapter Overview

The current chapter describes and discusses findings from the child behavioural questionnaires completed by the mothers who took part in CATS II. There is a brief introduction and methods overview (further information can be found chapters 1.1., General introduction and 1.3., Methods for the cognitive and behavioural data collection for the CATS II study), which lead into the analysis plan for the data. The results are discussed in respect of the primary and secondary analyses.

### 1.6.2. Introduction

Behavioural differences between individuals that have been affected by gestational underactive thyroid function are emerging in the literature; specifically, in respect to ADHD (44, 45, 143, 147-151) and ASCs (19, 170-172, 174-180). The aim of the analysis presented in this chapter was to test the null hypothesis that there would be no difference between the normal GTF, treated SGTF and untreated SGTF groups on any of the behavioural questionnaires administered to the mothers in respect of their offspring in CATS II. The null hypothesis was adopted as no cognitive differences were displayed between the treated and untreated SGTF at age 3 and 9 and thus, similarly, it was proposed that no differences between groups in regard to behaviour would be identified either. In the alternate hypothesis, it was predicted that there would be a behaviour deficit for the untreated SGTF group compared to the normal GTF group; as had been suggested in the literature.

This chapter details the participants and demographics for the three groups. It also discusses replacing missing items on the questionnaires as well as the three-models of statistical analysis, results and discussion section.

### 1.6.3. Method

A total of 480 participants were needed to address the null hypothesis that treatment for an underactive thyroid during pregnancy would have no effect on the offspring's behaviour as compared to those who were born to mothers who were untreated; with the normal GTF group acting as a baseline comparison. This recruitment aim was the same as the cognitive assessment goal as the questionnaires were completed at the same time by the mothers. Four questionnaires were administered to the mothers, of which three relate to the behavioural analysis of the offspring; SDQ, Child ADHD Questionnaire and SCQ. As with the cognitive assessments, I scored the questionnaires blind to avoid any group bias.

The SDQ questionnaire generated scores for emotional symptoms, conduct problems, hyperactivity, peer problems, a total difficulties score and also a prosocial rating. Higher scores on a subscale indicated a problem behaviour, except for the prosocial scale where higher scores indicated positive behaviour (proposed 80% scoring 'close to average': 8-10). The total difficulties ranged from 0-40 (proposed 80% scoring 'close to average': 0-13) and were obtained by summing the scores from hyperactivity (proposed 80% scoring 'close to average': 0-5), emotional symptoms (proposed 80% scoring 'close to average': 0-3), conduct problems (proposed 80% scoring 'close to average': 0-2) and peer problems (proposed 80% scoring 'close to average': 0-2). From the SDQ, 6 scores were used in the current analysis.

The Child ADHD Questionnaire generated four scores for analysis: inattention, overactivity, impulsivity and a total ADHD score. Subscale scores were generated by adding all items in a subscale together; inattention 0-27, overactivity 0-12, impulsivity 0-15. Based on the modified DuPaul rating scale (232), the following means and SDs (in brackets) were developed based on a general population: inattention 6.05 (6.21), overactivity 2.17 (2.74) and impulsivity 3.59 (3.44).

The final behavioural questionnaire was the SCQ investigating symptoms synonymous with ASCs. The SCQ generated a total score, if it was ≥ 15, this would indicate a possible ASC. The first item (of 40 questions) enquired about the level of language of the child and thus was not included in the total score. The SCQ only generated a total score for the current analysis.

The questionnaires were scored soon after completion; all notes added by the mothers were also stored electronically. Participants were mailed requesting further completion if large sections or sides of a questionnaire were missing. No feedback to the parents was generated for the questionnaires, unless requested.

### 1.6.4. Statistical Analysis
A total of 483 participants completed the questionnaires and a total of 18 participants were removed from this analysis. As a result of the WISC-IV cleaning (see chapter 1.4., Intelligence measured at age 9; CATS II data) nine were excluded; one was identified as an outlier, six children for ages outside the specified target (see study protocol (280)) and two participants were a set of twins. On the questionnaire dataset, one participant was the sibling of the child who was in CATS II and therefore was removed, and three participants did not complete the behavioural questionnaires; this left 471 participants who had completed the questionnaires. A final six participants were excluded from the analysis due to not having answered one of the three questionnaires (four participants), and two others due to missing

information for adjustments; no breastfeeding data, and no social deprivation score was available for a participant who now lived in France. Therefore, 465 participants' data were used in the final analysis.

Data were cleaned without the participants coded to their groups to help minimise any researcher bias. As CATS II was a pragmatic study, we included a wide range of IQs and SENs as we wanted to achieve variability in our study so that it would be more generalisable to the wider population. This meant, not computing z-scores, as those likely to achieve ± 3 score would be those individuals who were recognised as, for example, having some autistic traits or ADHD related difficulties. This in-turn could have skewed the data and thus not have been a true representation of behaviour difficulties expressed within our three groups.

Nominal data was investigated for miss-input by range checks; this included how participants took part in the study, child gender, child date of birth. The ordinal scores from the SDQ questionnaire were first checked for correct ranges (0-2), then the total and rankings for each domain. This was also completed for the child ADHD questionnaire domains (desired range 0-3) and the SCQ (0 or 1).

Missing values were commonplace throughout a number of questionnaires. Replacing missing values by 'series mean' in SPSS was adopted to overcome the problem (seen as one of the better methods for dealing with this problem (314)). For the SDQ, there was only ever a maximum of three missing scores for the questionnaire, and for all of the difficulty subscales, if participants did omit an item it was only ever the one item. One individual did not answer any items on the SDQ. The replacing of missing values was not conducted across the entire SDQ, but on the domain level (i.e. the new score for the missing item on the emotional scale, was averaged by scores the individual received on the emotional questions only and thus there was no 'item interference' from how they performed in a different domain). One participant omitted answering the child ADHD questionnaire and also the SCQ. The missing data for the child ADHD questionnaire was completed in much the same way as the SDQ; i.e. by domain level. Eight participants missed scoring the back page of the questionnaire (items 19-26: 23-26 were needed for the analysis). For these eight individuals, it meant that 25% of the overactivity items were missing (one question) and 33% of the inattention items were missing (three questions). As the amount of missing data was low, mean items to generate a 'missing value' were still used. Finally for SCQ, two individuals had omitted this questionnaire, one participant only completed 50%, and another had around

40% missing. Based on these high levels of omissions, all four participants were removed from the SCQ analysis.

All Kolmogorov-Smirnov (and Shapiro-Wilk) normality tests were returned as $p < .001$ 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 (see Table 27 below). The prosocial SDQ received a Kurtosis value of 4.535, this was because there was a high peak around participants receiving the top score for the scale (i.e. they were rated by their mothers to have a high prosocial demeanour). The questionnaire data was analysed by parametric tests as the total scores were continuous, and based on the means, medians, skewness and kurtosis, the data was accepted as being normally distributed with caution.

Table 27

*Means, Medians, Skewness and Kurtosis of Behavioural Questionnaires*

| Questionnaire | Mean | Median | Skewness | Kurtosis |
|---|---|---|---|---|
| SDQ Emotion Total | 2.30 (2.14) | 2.00 | .896 | .229 |
| SDQ Conduct Total | 1.37 (1.60) | 1.00 | 1.345 | 1.705 |
| SDQ Hyperactivity Total | 3.38 (2.71) | 3.00 | .590 | -.456 |
| SDQ Peer Problems Total | 1.38 (1.71) | 1.00 | 1.485 | 1.972 |
| SDQ Total Difficulties Total | 8.43 (6.02) | 7.00 | .854 | .158 |
| SDQ Prosocial Total | 8.79 (1.75) | 9.03 | -1.962 | 4.535 |
| ADHD Inattention Mean | 6.24 (5.70) | 5.00 | 1.228 | 1.210 |
| ADHD Overactivity Mean | 2.39 (2.64) | 1.43 | 1.364 | 1.403 |
| ADHD Impulsivity Mean | 3.38 (3.01) | 3.00 | 1.140 | .986 |
| ADHD Total Mean | 12.01 (10.17) | 9.00 | 1.254 | 1.384 |
| SCQ Total Mean | 4.51 (3.86) | 4.00 | 1.349 | 2.326 |

Note. Standard deviations appear in parentheses below means. SDQ=strengths and difficulties questionnaire, ADHD=attentional deficit hyperactivity disorder, SCQ=social communication questionnaire.

Table 28
*Questionnaire Participant Group Demographics*

| | | Normal GTF (n=245) | Treated SGTF (n=120) | Untreated SGTF (n=106) |
|---|---|---|---|---|
| Mean Age (yrs) | | 9.66 | 9.43 | 9.43 |
| Gender (n) | Male | 123 (50.2%) | 64 (53.3%) | 50 (47.2%) |
| | Female | 122 | 56 | 56 |
| Whether the mother breastfed for over 1 month | Yes | 160 (65.3%) | 71 (59.2%) | 58 (54.7%) |
| | No | 85 | 49 | 47 |
| | Missing | 0 | 0 | 1 |
| Mother age at time of consent into CATS I Mean of quartile (mean, years) | | 2.51 (31.74) | 2.22 (30.32) | 2.31 (30.89) |
| Social deprivation score (mean of quintile) | | 3.68 | 3.83 (missing 1) | 3.38 |
| Questionnaires (n & missing n) | SDQ | 244 (1) | 120 | 106 |
| | Child ADHD Questionnaire | 244 (1) | 120 | 106 |
| | SCQ | 242 (3) | 120 | 105 (1) |

Note. Further information appears in parentheses following n's were applicable. SGTF=suboptimal gestational thyroid function, CATS=controlled antenatal thyroid screening, SDQ=strengths and difficulties questionnaire, ADHD=attentional deficit hyperactivity disorder, SCQ=social communication questionnaire.

Table 28 shows the group demographics of each participant group for the analysis of the questionnaires. The ages of the children were compared by an ANOVA, and as with to the cognitive findings, there was a significant difference identified: $F$ (2, 468) = 5.426, $p$ = .005, $\eta_p^2$ = .023. Again, it was the normal group who were older than the treated ($p$ = .019, 95% CI [.0283, .4349]) and the untreated ($p$ = .029, 95% CI [.0177, .4420]) SGTF groups. This should not have affected the questionnaire results as the inventories were designed to be completed by a wide age range (231-233). The characteristics (gender, whether the mother breastfed over one month, age of mother at time of consent into CATS I and social deprivation rating), appeared to have equal spread across the groups. However, as bias could have been introduced by these subtle differences, these covariates were controlled for.

The questionnaire data was analysed in IBM SPSS Statistics version 20 by a MANCOVA. Models one and two of the analysis were adopted (see the published CATS II protocol (280), and chapter 1.3., Methods for the cognitive and behavioural data collection for the CATS II study) and also model three, but adjusted to include social deprivation. Child age was also taken into consideration in model three, as the questionnaires could assess across a wide age range and were not as age stringent in scoring compared to the WPPSI-III, WISC-IV and NEPSY-II. Data was analysed by use of total scores for domains only, and it did not include

any ordinal data or classifications. Univariate analysis commenced dependent on significant multivariate results.

The second exploratory analysis investigated questionnaire outcomes by a binary cut-off. For SDQ, the scores for each domain were classified as either 'close to average', 'high average', 'high' or 'very high' (apart from the prosocial domain which were 'close to average', 'low average', 'low' or 'very low') (231). Scores resulting in a 'high' (/'low' for the prosocial domain) classification or above were selected for the SDQ cut-off. Thapar et al. (232) had established SDs for the Child ADHD Questionnaire, thus individuals scoring ≥ 1 SD were included; similar to the binary analyses of the WISC-IV and NEPSY-II with scores 1 SD from the mean. SCQ had a cut-off of 15 points to warrant further investigation for a possible ASC (233), therefore scores ≥ 15 were included here. Unadjusted models were firstly compared by Pearson Chi-square, with adjusted multinomial logistic regressions completed secondly (as with the MANCOVA, these controlled for child gender, whether the mother breastfed over one month, age of mother at time of consent into CATS I, social deprivation score and also child age).

Exploratory analyses of maternal T4 and TSH levels in respect of offspring behaviour questionnaire results can be found in appendix 10. As discussed in chapter 1.2., those women randomised to the treatment branch of CATS I who had SGTF, began levothyroxine therapy at 150 μg. As a consequence, mean T4 values at 6 weeks post consent and at 30 weeks gestation (when women were visited to check whether levothyroxine adjustments were required) were reasonably high, and some women were classified as being 'over-treated' (T4 > 17.7 pmol/L). The offspring questionnaire results of such women were compared against others in the study. Prevalence rates of undesirable behaviour were also explored in this appendix.

### 1.6.5. Results

### 1.6.5.1. General attendance information

As stated in the chapter 1.3. (Methods for the cognitive and behavioural data collection for the CATS II study), CATS II aimed to recruit a total of 480 participants back into the study. This consisted of 120 from the treated and untreated SGTF groups, and 240 from the normal GTF group. The questionnaires were completed by a total of 471 participants. The treated SGTF group achieved the target (n = 120), the untreated group were 11.7% behind the recruitment target (n = 106) and the normal GTF above the desired target (n = 245). See Table 28 for group demographics.

*1.6.5.2. Analysis*

Table 29 shows the means and SDs for all study behavioural questionnaires as per group and adjusted for by model three (see Figure 6) (NB total n decreased due to missing data for questionnaires, breastfeeding > 1 month and a participant living abroad). The graphs (Figures 14, 15 and 16) display the means achieved by the groups pictorially (unadjusted model), error bars have also been included.

Table 29

*Behavioural Questionnaires Descriptive Statistics by Study Group and Model Four of Analysis*

| Questionnaire Domain | PARTICIPANT GROUP CODE | Mean | N |
|---|---|---|---|
| SDQ Emotion Total | Normal GTF | 2.30 (2.23) | 242 |
| | Treated SGTF | 2.40 (2.05) | 119 |
| | Untreated SGTF | 2.22 (2.00) | 104 |
| | Total | 2.31 (2.13) | 465 |
| SDQ Conduct Total | Normal GTF | 1.23 (1.50) | 242 |
| | Treated SGTF | 1.65 (1.84) | 119 |
| | Untreated SGTF | 1.28 (1.42) | 104 |
| | Total | 1.35 (1.58) | 465 |
| SDQ Hyperactivity Total | Normal GTF | 3.12 (2.51) | 242 |
| | Treated SGTF | 3.79 (2.94) | 119 |
| | Untreated SGTF | 3.43 (2.73) | 104 |
| | Total | 3.36 (2.69) | 465 |
| SDQ Peer Problem Total | Normal GTF | 1.55 (1.87) | 242 |
| | Treated SGTF | 1.20 (1.61) | 119 |
| | Untreated SGTF | 1.19 (1.34) | 104 |
| | Total | 1.38 (1.70) | 465 |
| SDQ Total Difficulties Total | Normal GTF | 8.20 (5.95) | 242 |
| | Treated SGTF | 9.04 (6.46) | 119 |
| | Untreated SGTF | 8.12 (5.44) | 104 |
| | Total | 8.40 (5.98) | 465 |
| SDQ Prosocial Total | Normal GTF | 8.77 (1.77) | 242 |
| | Treated SGTF | 8.81 (1.80) | 119 |
| | Untreated SGTF | 8.77 (1.70) | 104 |

| | | | |
|---|---|---|---|
| | Total | 8.78<br>(1.76) | 465 |
| ADHD inattention<br>Mean | Normal GTF | 6.10<br>(5.64) | 242 |
| | Treated SGTF | 6.91<br>(6.12) | 119 |
| | Untreated SGTF | 5.74<br>(5.22) | 104 |
| | Total | 6.23<br>(5.68) | 465 |
| ADHD overactivity<br>Mean | Normal GTF | 2.08<br>(2.29) | 242 |
| | Treated SGTF | 2.90<br>(3.06) | 119 |
| | Untreated SGTF | 2.47<br>(2.70) | 104 |
| | Total | 2.38<br>(2.62) | 465 |
| ADHD impulsivity<br>Mean | Normal GTF | 3.17<br>(2.81) | 242 |
| | Treated SGTF | 3.85<br>(3.43) | 119 |
| | Untreated SGTF | 3.28<br>(2.83) | 104 |
| | Total | 3.37<br>(2.99) | 465 |
| ADHD Total Mean | Normal GTF | 11.36<br>(9.49) | 242 |
| | Treated SGTF | 13.66<br>(11.66) | 119 |
| | Untreated SGTF | 11.48<br>(9.44) | 104 |
| | Total | 11.97<br>(10.11) | 465 |
| SCQ Total Mean | Normal GTF | 4.31<br>(3.68) | 242 |
| | Treated SGTF | 5.13<br>(4.68) | 119 |
| | Untreated SGTF | 4.07<br>(2.93) | 104 |
| | Total | 4.47<br>(3.83) | 465 |

Note. Standard deviations appear in parentheses below means. SDQ=strengths and difficulties questionnaire, ADHD=attentional deficit hyperactivity disorder SCQ=social communication questionnaire, SGTF=suboptimal gestational thyroid function.
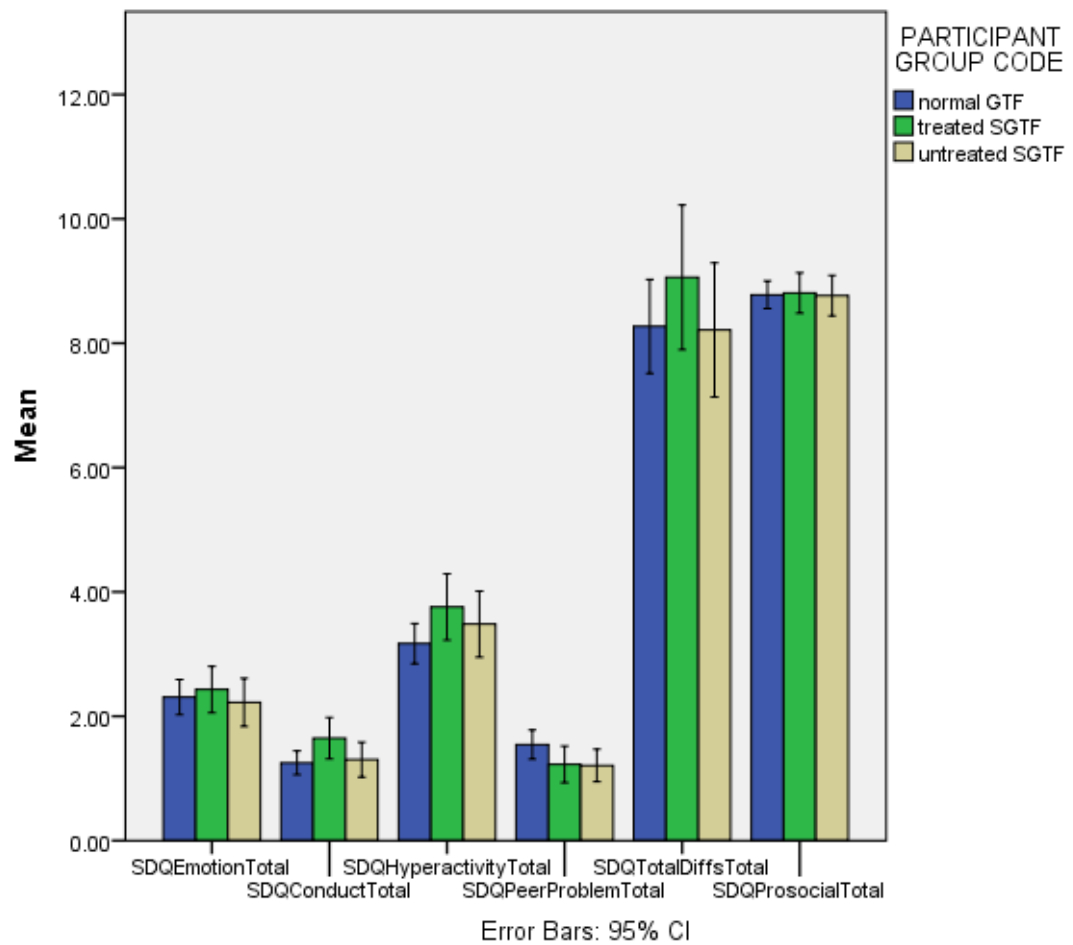
Figure 14: Means of scores achieved per group for the Strengths and Difficulties Questionnaire (SDQ)

SGTF=suboptimal gestational thyroid function, CI=confidence interval. X axis=domains from the SDQ, Y axis=mean of participant groups.
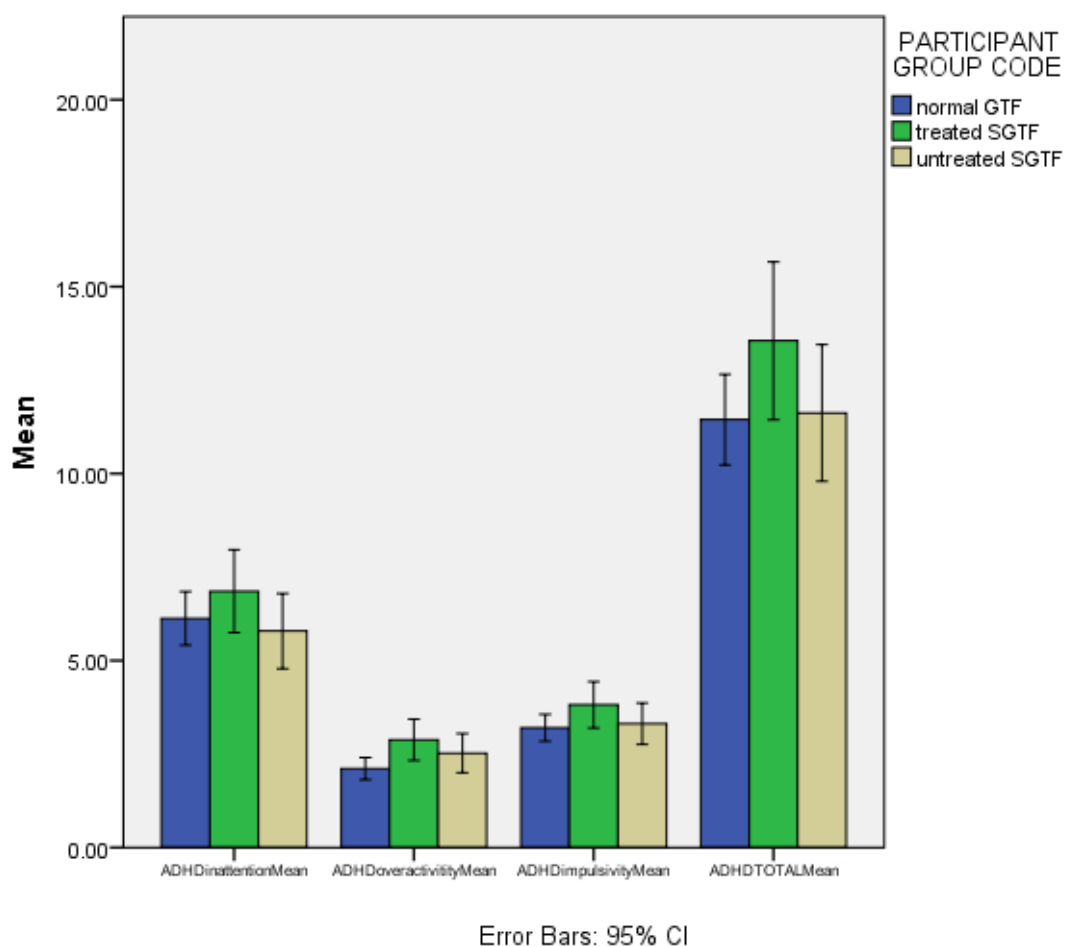
Figure 15: Means of scores achieved per group for the Child Attention Deficit Hyperactivity Disorder (ADHD) Questionnaire

SGTF=suboptimal gestational thyroid function, CI= confidence interval. X axis=domains from the ADHD questionnaire, Y axis=mean of participant groups.
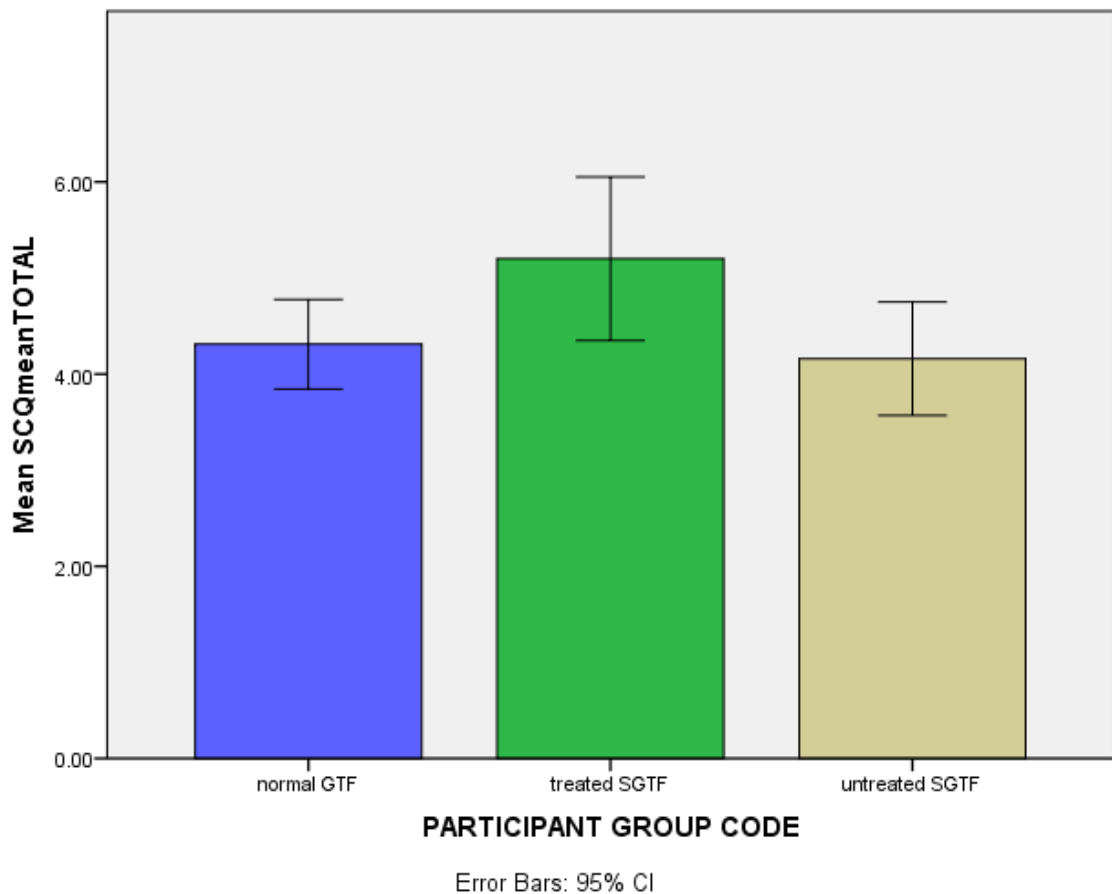
Figure 16: Means of scores achieved per group for the Social Communication Questionnaire (SCQ)

SGTF=suboptimal gestational thyroid function, CI= confidence interval. X axis=domains from the SCQ, Y axis=mean of participant groups.

The first model of analysis identified a significant effect of group on the behavioural questionnaires: $\Lambda_{ROY}$ = .057, $F$ (9, 457) = 2.902, $p$ = .002, $\eta_p^2$ = .054. When adjusted for gender (model two), the significant result persisted: $\Lambda_{ROY}$ = .057, $F$ (9, 456) = 2.886, $p$ = .003, $\eta_p^2$ = .054, and by the final analysis of model three there was also a significant difference: $\Lambda_{ROY}$ = .052, $F$ (9, 450) = 2.587, $p$ = .006, $\eta_p^2$ = .049. The univariate tests revealed that the following questionnaires (and domains within a specified questionnaire) were significant; SDQ Peer Problems ($F$ (2, 458) = 3.623, $p$ = .027, $\eta_p^2$ = .016) and SCQ Total ($F$ (2, 458) = 3.099, $p$ = .046, $\eta_p^2$ = .013). The rest of the questionnaires yielded non-significant results when compared between the three groups, all $p's$ > .062; SDQ Emotion, SDQ Conduct, SDQ Hyperactivity, SDA Total Difficulties, SDQ Prosocial, ADHD Inattention, ADHD Overactivity, ADHD Impulsivity and ADHD Total.

Post hoc analysis was executed by Bonferroni testing. Bonferroni was a much more conservative test so that it can take account of multiple testing, and has been described as the most robust of the univariate techniques (210). With this in mind, upon post hoc analysis, SDQ Peer Problems had lost significance between the groups. For SCQ Total, those children from the treated SGTF group had significantly higher scores than those from the untreated SGTF group ($p$ = .047, 95% CI [.012, 2.412]).

The second exploratory analysis investigated questionnaire outcomes by a binary cut-off. For SDQ, scores were included if above the 'high' classification ('low' for prosocial), the ADHD questionnaire included those ≥ 1 SD and the SCQ included those ≥ 15 (258). Results for the unadjusted chi-square can be found in Table 30 (significant results were highlighted **). Multinomial logistic regressions were executed to explore any differences between the groups for the significant questionnaire results from the preliminary unadjusted chi-square; SDQ emotion, ADHD Overactivity, ADHD Impulsivity and SCQ. There was a non-significant regression for SDQ emotion ($p$'s > .189).

For ADHD Overactivity, the normal GTF group were 2.027 times less likely to score ≥ 1SD compared to the treated SGTF group ($p$ = .017), all other interactions were non-significant. For ADHD Impulsivity, the normal GTF group were 2.060 times less likely to score above 1 SD compared to the treated SGTF group ($p$ = .038), all other interactions were non-significant. Finally, for SCQ, the normal GTF group were 4.132 times less likely to score above the threshold compared to the treated SGTF group ($p$ = .031), all other interactions were non-significant. Further information about the goodness of fit and parameter estimates for the multinomial logistic regressions can be found in Appendix 7: Behavioural questionnaires, additional regression models.

Table 30

*Chi-Square (Unadjusted) Statistics for Participants Scoring Above the Specified Threshold for the Questionnaires*

| | SDQ emotion | SDQ Conduct | SDQ Hyperactivity | SDQ Peer problems | SDQ Total | SDQ Prosocial | ADHD Inattention | ADHD Overactivity | ADHD Impulsivity | SCQ |
|---|---|---|---|---|---|---|---|---|---|---|
| | Higher scores indicating worse behaviour* | | | | | | | | | |
| Normal GTF | 33/244 (13%) | 21/244 (7%) | 17/244 (7%) | 40/244 (16%) | 29/244 (12%) | 28/244 (11%) | 30/244 (12%) | 33/244 (13%) | 20/244 (8%) | 4/242 (2%) |
| Treated SGTF | 31/120 (26%) | 19/120 (16%) | 17/120 (14%) | 14/120 (12%) | 17/120 (14%) | 14/120 (12%) | 22/120 (18%) | 31/120 (26%) | 21/120 (17%) | 7/120 (6%) |
| Untreated SGTF | 26/106 (24%) | 8/106 (7%) | 12/106 (11%) | 9/106 (8%) | 9/106 (8%) | 10/106 (9%) | 11/106 (10%) | 26/106 (24%) | 12/106 (11%) | 1/105 (1%) |
| Pearson Chi-square | *p*=.005** | *p*=.060 | *p*=.079 | *p*=.111 | *p*=.412 | *p*=.830 | *p*=.165 | *p*=.005** | *p*=.031** | *p*=.030** |

*Note.* Percentages of scores per group are appear in parentheses below totals. 'High' (*Strengths and Difficulties Questionnaire (SDQ) scores ≥ 'High' classification, Child attention deficit hyperactivity disorder questionnaire (ADHD) scores ≥ 1 SD, Social Communication Questionnaire (SCQ) scores ≥ 15 point threshold for possible autism spectrum conditions). **Significance < .05. SGTF=suboptimal gestational thyroid function.

∴

.

The aim of this chapter was to explore the null hypothesis of there being no difference between the treated SGTF, untreated SGTF and normal GTF groups on any of the behavioural questionnaires at the sub-domain, total-scored level. Based on the $p$ = .006 from the multivariate analysis, the null hypothesis was rejected. Post hoc analysis indicated that the treated SGTF group had significantly higher mean scores for the SCQ questionnaire compared to the untreated SGTF group ($p$ = .047). The mean scores of the groups do not infer a clinical significance, and as can be seen in the Table 29 and on page 111, the scores were still within the average range and were below clinical significance (for example, the treated SGTF group mean for SCQ was not above the threshold of 15 to warrant further investigation for ASCs).

The secondary exploratory analysis of binary outcomes from the behavioural questionnaires identified that for SCQ, the normal GTF were 4.132 times less likely to score above the threshold (≥ 15) compared to the treated SGTF group ($p$ = .031). It was interesting that by investigating above the threshold for binary outcomes rather than continuous, this changed the significant difference between the treated SGTF group to the normal GTF group, rather than the untreated SGTF group (seen in the continuous analysis); still inferring that the treated SGTF group presented more ASC symptoms however. On the Child ADHD Questionnaire, the normal GTF group were less likely to score ≥ 1 SD compared to the treated SGTF group for Overactivity and Impulsivity (2.027 OR $p$ = .017 and 2.060 OR $p$ = .031, respectively). However, as can be seen in Appendix 7: Behavioural questionnaires, additional regression models, the significant logistic regressions mentioned here do contain wide confidence intervals suggesting a larger amount of variability. Furthermore, the adjusted regressions all suggest that the treated SGTF group have more ADHD and ASC traits compared to the normal GTF group. Although this analysis was not the primary outcome but was executed for comparative purposes to the main CATS II analysis, similar to the MANCOVA, it has demonstrated the treated SGTF group to be significantly higher (indicating more behaviour problems) on the behavioural questionnaires. Caution has been advised when interpreting binary outcomes from continuous data however (211, 212), as there would be an increased risk of a type one error (213).

As CATS I was the first study to look at treatment for SGTF, there was little literature available that investigates those whom were treated during their pregnancies and the behavioural outcomes of their offspring compared to those who were not treated. Treatment for congenital hypothyroidism has been evidenced to have a negative effect on the child's

behaviour if it was inadequate (315). Also, the authors identified that behaviour problems such as aggression and poor attention were more frequent in children with congenital hypothyroidism compared to controls (rated by their teachers, not parents). This was contradictory to the findings here as no (specific) deficits in attention were identified and aggression was not investigated.

CATS I used a definition of 'sub-optimal' thyroid function for recruitment of an underactive thyroid (200, 280): this was T4 in the lowest 2.5$^{th}$ percentile, TSH in the highest 2.5$^{th}$ percentile, or both. In essence, CATS has a mixture of mothers who were either hypothyroxinaemic during their pregnancies (low T4, normal TSH), or who had GSH (high TSH, normal T4). Ghassabian et al. (316) found that GSH resulted in higher externalising scores in children aged 1 and a half and 3 years old. It was also found that hypothyroxinaemia during pregnancy was not associated with internalizing or externalising scores for the children. However, in iodine insufficient areas, hypothyroxinaemic mothers during their pregnancies reported children with an abnormally high frequency of ADHD (151). Children exposed to hypothyroxinaemia in early pregnancy (measured at a mean of 13.6 weeks gestation) were more likely to express ADHD symptoms at age 8; independent of confounders (45). Furthermore, it was identified in Modesto et al.'s study that those born to mothers with GSH did not express any ADHD symptoms, also a small portion of the maternal hypothyroxinaemic mothers did receive a low levothyroxine treatment, but this had no effect to the high ADHD outcomes measured in the offspring. We could infer from this that by subdividing our SGTF groups into low T4 or high TSH could indicate differing results to what had been identified. There may be more mothers who had GSH in the untreated group compared to the treated which would affect the results. However, this would generate small study groups and would not be representative of the wider population.

Ghassabian et al. (149) also investigated TPO-Ab+ during early pregnancy and found that, if present, there was an increased risk of externalising problems in preschool children: in particular, ADHD problems. Brown et al. (175) found that the prevalence of maternal TPO-Ab+ increased the odds of ASC in offspring by 80% ($p$ = .009) compared to mothers negative for this antibody. These maternal TPO studies are interesting, as akin to CATS II, deficits for ASCs and ADHD were identified. Unfortunately TPO was not measured during the CATS I pregnancies, making comparisons and comments difficult. Some studies have found no link between gestational thyroid disorders and ASCs in children (171, 172, 174), whilst others have found a link (19, 170, 175-180) which supports the current finding that one of the SGTF groups had significantly higher scores on the SCQ compared to the normal GTF group.

119

As noted in chapter 1.1. (General introduction), Stein and Weiss (147) did identify a link between thyroid function and ADHD. The study investigated the thyroid function in the child however, so caution is advised in comparing to the CATS II cohort; although CATS II will eventually have thyroid function status for some of the children which may be a useful analysis. Furthermore, deficits were only found with low FT4 for inattentive ADHD types and no link to hyperactivity (/overactivity) or impulsivity were identified. This contradicts the current binary findings as no links were identified between any of the three groups for inattention. There was literature that suggested there is no link at all between thyroid hormone function and ADHD (19, 45, 144-146, 148), which the MANCOVA analysis in the current chapter also confirmed.

The question as to why the treated SGTF performed worse on the behavioural questions compared to the other two groups is one that is difficult to answer. As discussed, it does indicate that an underactive thyroid during pregnancy does have some effect on the offspring; and that levothyroxine treatment in some way may be affecting the offspring's behaviour. Regarding the cognitive testing for CATS II, it was identified that there were no differences between the three groups for IQ, long term memory or fine motor coordination. The treated SGTF mothers were followed-up during their pregnancies to check for the correct dosage of levothyroxine (200) to ensure none of them had been overprescribed the hormone. It could be suggested that this regular checking, assessment and prescription of drugs may have induced a certain level of stress in the mothers. It has been evidenced in the literature that stress during pregnancy could lead to emotional and behavioural problems for the child (317). The behaviour focus appears to be around ADHD; maternal stress was found to be linked to hyperactivity and inattention in boys and total behavioural and emotional problems in both boys and girls (318), as well as a modest contribution being identified between maternal stress and ADHD (319). However, further investigation is needed as stress for the mothers should have been kept to a minimum by study team members, i.e. we have identified that you have sub-optimal thyroid function, but we are correcting this for you.

Out of the 11 behavioural questionnaire domains inputted to the MANCOVA, only five outputs yielded means where the normal GTF had the lowest problem behavioural score (see Table 29 for group means). It can be inferred from this, that an underactive thyroid during pregnancy has some effect on the offspring's behaviour; with treatment making the behaviour more pronounced.

The supplementary analyses in appendix 10 identified that, within the CATS II cohort, treatment for underactive thyroid function during pregnancy resulted in offspring having more ADHD Overactivity and autism-type behaviours, though not clinically significant. One of the reasons could be that around a third of the mothers in CATS II were over-treated with levothyroxine during their pregnancies, which identified a need for clinicians to closely monitor dosage levels during the gestation period.

Although treatment for SGTF has not been previously investigated (by an RCT), there has been recent research that has suggested higher levels of T4 being detrimental to the offspring in respect of intelligence and brain morphology (37). CATS II was the first study to show that treatment could have a detrimental effect to the offspring behaviour at age 9.

Adjustment of levothyroxine dosage in CATS II was centred on TSH levels, rather than T4. TSH levels were within the normal range for the treated SGTF group during pregnancy (200). The mean of T4 decreased between 6 weeks post consent and at 30 weeks gestation, indicating that there was an attempt to correct the over-treatment.

One of the explanations of the treated SGTF group expressing more behavioural difficulties may be due to the placental transfer of thyroid hormones. During the first trimester, the foetus is completely surrounded by the placenta, and it regulates the uptake of all types of thyroid hormone for the foetus (320). Fetal T4 is correlated to maternal circulating T4, but is lower compared to maternal serum values, which indicates the placental barrier system (59). D2 and Deiodinase 3 (D3, inactivating thyroid-hormone enzyme), are expressed in the placenta. D3 metabolises the majority of the maternal T4 from the placenta; placental D2 decreases as pregnancy continues (321), whereas placental D3 activity increases (322). Placental D3 may therefore have a more substantial role in the placental barrier system of protecting the foetus from excessive maternal T4 levels (323). However, if there is excessive circulating maternal T4, the placental barrier may become 'saturated' and unable to fully regulate the maternal T4 transfer sufficiently (324). This could explain how excess levels of maternal T4 are able to reach the foetus.

The implications of these findings are that maternal T4 levels have more of an influence on offspring behaviour then maternal TSH levels. The research also suggested that higher levels of maternal T4 were more detrimental than lower levels of maternal T4 to offspring. Furthermore, treatment of pregnant women with levothyroxine may need to be monitored more closely during pregnancy.

Small participant groups make the research difficult to generalise to the wider public. To ascertain a broader picture of the children's functioning, teacher questionnaires could have been gathered. However, the time frame for this to be completed may have been challenging. The NEPSY-II and WISC-IV both allowed behaviour observations throughout the assessments which were not adopted as part of the CATS II protocol. If the study was being repeated, it may be beneficial to gather this information from the examiner and see if it could confirm findings from the behavioural questionnaires completed by the mothers.

A problem with the data and with questionnaires in general, was ensuring a 100% completion rate. As mentioned, the team attempted to overcome this by 'spot-checking' the questionnaires filled out by the mothers; for remote/home visits and post packs, this was inevitably more difficult. By accepting a low percentage of missing data for calculating mean scores, I ensured accuracy in the missing items as the means generated were from the vast majority of completed items.

*1.6.6.2. Conclusions*

This results chapter has answered one (bold) of the four research aims for this section:

i. By reassessing the children at the older age of 9, would there be a continued non-significant difference identified for intelligence.
ii. Would there also be non-significant findings at age 9 in other potential areas of cognition.
iii. As there were no differences between the treated and untreated SGTF groups at age 3, would any differences to the normal GTF group be measurable; as there is a wealth of studies displaying that an underactive thyroid during pregnancy does not affect a child's cognition (53, 56, 91-93, 95, 96).
iv. **Would there also be non-significant differences between the groups that extend beyond cognition, i.e. behaviour.**

The hypothesis that there would be differences between the normal GTF and untreated SGTF group was not identified in either the primary or secondary analyses. The second hypothesis was that the treated SGTF group would display less behavioural problems than the untreated SGTF groups. This was unfounded as on the SCQ, the treated group were identified as scoring significantly higher. Furthermore, the treated SGTF group displayed more behaviour issues ≥ specified cut-offs compared to the normal GTF group for the SCQ and ADHD Overactivity and Impulsivity in the secondary analysis. Investigating different types of maternal thyroid dysfunction (i.e. hypothyroxinaemia and subclinical hypothyroidism), with treatment beginning earlier in the pregnancy, in larger groups, may in the future yield different results.

Within this study, treatment for SGTF had no effect on cognition in the group as a whole, but it did appear to affect the child's behaviour.

### 1.6.7. Chapter Summary

The current chapter outlined the results from the data collection of the child behavioural questionnaires completed by the mothers from the normal GTF and treated and untreated SGTF groups. The results indicated that the treated SGTF group offspring had higher scores (indicating more behaviour difficulties) on the SCQ, ADHD Overactivity and ADHD Impulsivity (compared to the normal GTF and for SCQ they also had higher scores than the untreated SGTF group). The results suggested that treatment for an underactive thyroid during pregnancy may increase the chances of ADHD and ASC type behaviours for the offspring; although it is important to note that the overall means do not suggest a clinical significance of a SEN as means were all within the 'average' range for the behavioural questionnaires. This work concludes the first section of this thesis, the following includes more exploratory investigations from both the CATS I and II data. The next chapter investigated the significant effects the covariates had on the dependent variables from the WISC-IV (chapter 1.4., Intelligence measured at age 9; CATS II data), NEPSY-II (chapter 1.5., Additional cognitive assessments at age 9; CATS II data) and behavioural questionnaires (current chapter).

2. Exploratory Analysis of the Controlled Antenatal Thyroid Screening Study II; cognitive and behavioural data

## 2.1. Significant effects from the covariates; CATS II data

### 2.1.1. Chapter Overview

There were a total of six covariates that were adjusted in the analyses of section one of this thesis; these were selected as they may have had an effect on the dependent variables measured. This chapter reports on the covariates from the final model MANCOVAs that were shown to have a significant effect on the WISC-IV, NEPSY-II and the behavioural questionnaire outcomes. Child age was also adjusted in the behavioural questionnaire analysis (presented in chapter 1.6., Behavioural quesitonnaires at age 9; CATS II data), but was not investigated here. Undesirable behaviour is not restricted to a specifc age range, and the final MANCOVA from the quesionnaire chapter revealed a non-significant effect of child age on mean questionnaire scores ($p$ = .221).

The introduction section is organised by the six covariates accounted for by order of the four-model statistical analysis (see Figure 6). Model one was the unadjusted analysis and model two adjusted for the first covariate, child gender. Model three adjusted for model two plus the age of the mother when she first consented into CATS I during her pregnancy and whether she breastfed for longer than one month. The final model of analysis, model four, adjusted for model three plus where the child was assessed (i.e. home or at the research centre), the language spoken at the child's school and home, and social deprivation. The hypotheses and results are then organised by the data collection tools.

### 2.1.2. Introduction

#### 2.1.2.1. Gender (model 2)

Gender differences in cognitive domains have captured the interest of many researchers over the decades. With the WISC-IV being so widely adopted for intelligence testing, it has attracted scrutiny of possible gender differences. There appears to be a trend for gender-specific abilities on specific cognitive domains (325). Females have been reported to obtain higher mean scores in the verbal domain, whilst males perform best in the domains of spatial and mathematical reasoning abilities (326). For example, males performed significantly better on the Mazes subtest from the first edition WPPSI compared to females (327, 328).

Males have been evidenced to show advantages in verbal domains, quantitative reasoning and visual-spatial ability (240, 329-331). Camarata and Woodcock (332) analysed gender differences across three cognitive assessment batteries in a total of 5,602 females and 4,863 males from America, they identified that males significantly outperformed females for verbal abilities. Goldbeck et al. (333) analysed males and females on the WISC-IV (German version)

and found that males outperformed females in the VC and PR domains. Conversely, Halpern et al. (326) and Weiss et al. (334) found that females had a superior VC domain, whereas Chen and Zhu (335) found no differences between genders for this domain. Camarata and Woodcock (332) but there were no gender differences for long-term retrieval, visual-spatial abilities, fluid reasoning (latter two are similar to the concepts behind the PRIQ domain) and working memory (332); this latter finding has been supported elsewhere (333).

Goldbeck et al. (333) discussed how tasks involving PS, memory skills or the comparison of symbols (similar to the concept behind the symbol search WISC-IV subtest), favours females and the reason may be akin to an advantage in phonological coding (336). Goldbeck et al. found that females outperformed males in PSIQ ($p <$ .001); with sex differences being significant in both subtests (coding and symbol search). Camarata and Woodcock (332) found that PS differences were significant ($p <$ .001) with females scoring more than eight IQ points higher than males, however this difference was found in adolescence and appeared to diminish in young adulthood. It appears well documented that females outperform males in PS abilities, (240, 329-331); and the superiority of female performance on PS tasks compared to males has been recognised since the mid-80s (328). However, recently Chen and Zhu (335) identified no gender differences for the PS domain from the large norming sample for the American WISC-IV.

Studies have also found no differences between males and females for IQ. Chen and Zhu (335) analysed the WISC-IV US standardisation responses from the 2,200 sample (1,100 males, 1,100 females). No significant differences were found and it was concluded that WISC-IV scores for males and females should be interpreted in the same way; no differences for FSIQ has also been supported by Goldbeck et al. (333). It was also identified in the research that females performed less well on letter-number sequencing and symbol search as males appeared better at spatial perception, mental rotation and spatial visualisation, and Chen and Zhu concluded that females would have required extra cognitive capacity for these two subtests. Mulenga et al. (337) also found evidence of males outperforming females in the visual-spatial domain; in the NEPSY assessment rather than the WISC-IV. However, caution is advised in generalisation as this data was from the first edition NEPSY (252) and was from a small sample in Zambia. As mentioned in chapter 1.3. (Methods for the cognitive and behavioural data collection for the CATS II study), there is now a newer version of the WISC available- the fifth edition (247). Chen et al. (248) again assessed the factor invariance between genders for this scale and found no differences between male and females outcomes (based on the 2,200 US standardisation sample, 1,009 males, 1,101 females). Chen

identified, however, that male children performed slightly better on the visual spatial domain, and females on PS domain: though not significant.

Gender differences have also been examined on the SDQ. It was found in a large Norwegian sample (n = 29,631) on the self-report version of the SDQ, that females exhibited more emotional problems compared to males, and that males were reported to have higher means for the conduct and peer problem domains (338). Conversely, more recently in a smaller sample from the Netherlands (n = 2,185), it was found that males expressed more difficulties on the emotion domain compared to females, as well as scoring worse for peer problems, hyperactivity/inattention and on the prosocial domain (all $p$'s < .001) (339). Becker and Rothenberger (340) found gender also had a significant influence on the scales hyperactivity-inattention and emotional symptoms. Males had a significantly higher risk for hyperactivity-inattention and females had a significantly higher risk for emotional problems compared to males.

Any findings of gender differences from the hyperactivity/inattention domain from the SDQ would be comparable to any found on the Child ADHD Questionnaire. It has been widely researched that males may present more symptoms of ADHD compared to females (341-345) with the prevalence being as much as 3-4:1 male: female ratio (128). It has also been identified that males may express ADHD-typical behaviour differently to females; males may squirm and fidget, whilst females may talk excessively or act angry/resentful (346). As well as ADHD, ASCs are also viewed as being more prevalent amongst males compared to females (347-350). ASCs are four to five times more common in males than females with prevalence rates of one in 42 males compared to one in 189 females (351, 352).

### 2.1.2.1.1.    Gender summary for hypotheses formulation

The literature mentioned offers conflicting findings as to whether males or females would perform better on the VCIQ domain. Even though Chen (335) found no differences between the sexes for the VCIQ domain in the large US normative sample for the WISC-IV, the current hypothesis follows Goldbeck et al. (333) who found that males performed better than females for the VCIQ domain. Goldbeck et al.'s research was selected as favourable as it too was based on a European sample using a WISC-IV that had been re-normed for the population. For PRIQ and WMIQ it was hypothesised that there would be no significant differences (332) and for PSIQ, it was predicted that females would outperform the males (328, 332). As the literature appeared in equipoise for which sex was outperforming which, it was predicted that gender differences would not reject the null hypothesis for the FSIQ.

Similarly, for the NEPSY-II subtests it was predicted there would be no significant differences between the genders as the majority of the testing was related to the WM domain. For the behavioural questionnaires, males were predicted to score higher on the Child ADHD questionnaire and SCQ compared to females (128, 341-345, 347-350). For the SDQ it was hypothesized that there would be no gender differences for domains except for conduct and peer problems with males scoring higher (338, 339).

### 2.1.2.2. Mother age (model 3)

The average age of mothers has increased over the years from 26.4 years in 1974 to 29.3 years in 2002 (353). A reason for the increase in age of the first time mother may be a societal trend to delay beginning a family for a career or financial reasons (354). Increased maternal age can lead to increased risks for the foetus (355-357), and although the offspring's development has not been widely researched (358, 359), literature is starting to appear. It has been identified that increased maternal age has a direct linear association with a superior performance on intelligence tests (360, 361). A reason for this linear trend could be the social advantage of these mothers, i.e. economic security and increased education (362).

Edwards and Roff (363) had a large US sample (n = 23,717) and FSIQ WISC-III measurement data of 7 year old children and data on their mothers' age at birth. The graph on page five of the article shows that as maternal age increases, so does the offspring's IQ. With respect to more general cognitive abilities, it has been identified in the literature that young maternal age was negatively associated to children's mathematics and reading abilities at age 10 (364), however the statistically significant results diminished once maternal background covariates were controlled for.

Edwards and Roff (363) also plotted findings from the Bayley Motor scale for a slightly inflated sample, including the previous WISC-III, of 26,529 children at 8 months old. For maternal age on motor outcomes, there appears little effect after the age of 20 until the mother was > 40 years old and then there was a dramatic drop in the mean motor results (graph can also be found on page five of the article). A possible reason for this could have been that older mothers may not be as involved in upcoming technology and consequently not invest as much time improving finger dexterity.

As well as maternal age having an effect on the child's cognition, it was also reported that it can affect the child's behaviour. There was varying evidence for mothers who gave birth at younger ages to be more likely to have children with disruptive behaviours (361, 365-367). Furthermore, children born second or thirdly to young mothers were reported to have much

greater disruptive behaviours compared to the first born, and still worse when compared to older mothers (368); adjusted for covariates including social background. Saha et al. (369) studied a large cohort of 55,908 pregnancies in the early 60s in the US, and identified that with every five year increase in maternal age, there was a 12% decreased risk for externalising behaviours in the offspring; however, it was difficult to generalise from this study as only one culture was examined and the children were all aged 7 years (361). Externalising behaviours of aggression, opposition and overactivity have been reported to decrease as mother age increases (367), which is linked to ADHD like symptoms. A study that included the SDQ, found that male adolescent children born to mothers < 20 years of age had a higher prevalence of poor social functioning and that a 'U-curve' of performance was identified that meant that mothers ≥ 40 years of age were also more likely to have children who would exhibit poor social functioning, compared to those offspring born to mothers between aged 25-29 years old (370). Autism has also been found to be more prevalent amongst offspring born to older mothers compared to younger ones (371-376), but some studies have found contradictory findings (377-379).

### 2.1.2.2.1. Mother age summary for hypotheses formulation

It was predicted that the older the mothers were at time of consent into CATS I (i.e. during their pregnancy), the higher the scores would be for the WISC-IV (based on the above mentioned literature, (360, 361, 363, 364)). This hypothesis was also adopted for the NEPSY-II WM and long-term memory tasks. Conversely, for fine-motor coordination scores on the NEPSY-II, it was predicted that there would be no maternal age effect (based on the work by Edwards and Roff's (363)). There was reported literature that 'disruptive behaviours' were more common in children born to younger mothers; it was hypothesised for the SDQ and child ADHD questionnaire results that the older the mothers were, the less disruptive behaviours would have been reported (361, 365-367, 370). However, for the SCQ, the older the mothers were, the higher the questionnaire scores would be (371-376).

### 2.1.2.3. *Breastfeeding (model 3)*

There is a reported link between breastfeeding and cognitive development for offspring in early to middle childhood (380-383). It has been found that if the child was breastfed, they would perform better on scales of nonverbal ability, mathematics and reading ability compared to those who were not breastfed (384). It has also been evidenced that the length of breastfeeding is crucial (385-389). Mortensen et al. (386) found a non-linear relationship between length of time being breastfed and later adult intelligence (as measured on a

Wechsler scale). It was identified that the longer an infant was breastfed (by month), the higher the intelligence was, but any breastfeeding for longer than nine months incurred a slight drop in verbal, performance and FSIQ. Furthermore, it was identified in their sample that the verbal and performance IQ difference for those breastfed for less than one month (which also included those who were not breastfed at all), compared to those in the seven to nine month category, were 6 IQ points lower. Ruiz et al. (390) used the WISC-IV in their modest (n = 103) Spanish sample and found that for the VC, PR, WM and PS IQs all were significantly higher ($p < .001$) for those children that were breastfed for six months compared to those who were not. There does appear to be a reported trend of longer breastfeeding to a higher IQ, but it has been found to be non-significant in some cases (391-393); all three studies adjusted for multiple covariates, including social deprivation.

Controlling for covariates for maternal breastfeeding has been stated to be of importance, as after controlling, results may be non-significant (380). Der et al. (394) found that the positive association between breastfeeding and offspring IQ was not significant when controlling for maternal IQ; this has been supported elsewhere in the literature (393, 395, 396). Jacobson et al. (397) also found similar results that when controlling for the mothers' intelligence, previous significant results for breastfeeding were no longer significant on specific vocabulary domains. However, a very recent publication by Kanazawa (398) has identified that in a British sample, differences in IQ dependent on breastfeeding were still significant when controlling for maternal IQ. Maternal education is also important to consider when exploring the effects of breastfeeding. A study by Bertini et al. (399) identified there was an association between a lack of breastfeeding and low level maternal education (n = 900 mothers). Maternal education has also been positively associated to performance on cognitive tests within the context of breastfeeding (400). Skafida (401) identified that maternal education was a more robust predictor of breastfeeding compared to social deprivation, and that mothers less socially deprived with more educational qualifications were more likely to breastfeed (n = 5012 babies). This identified literature highlights how breastfeeding is somewhat mediated by maternal education

There is some evidence that motor coordination is affected by breastfeeding. Leventakou (402) recently identified that in a sample of Greek children assessed at 18 months of age, by Bayley scales, there was a positive linearly associated relationship between breastfeeding and all Bayley scales, except for gross motor; with fine motor coordination significantly related. Similarly, Dee (403) found that mothers who had not initiated breastfeeding were

more likely to report concerns about their child's fine-motor coordination than those who did initiate breastfeeding.

As well as the cognitive performance of children being affected by breastfeeding, child behaviour has also been reported. Infant feeding could influence behaviour by the nutrients within the milk (404), for example it has been shown that the increased fatty acid intake from breast milk has led to improved neurological development and fewer behavioural problems for the child (405). Data from the UK Millennium Cohort Study, > 10,000 mothers completed the SDQ about their child in the study (404), found that abnormal SDQ scores were less common in those children that were breastfed; the most accentuated effects were around those that were breastfed for > 4 months. In term babies who breastfed for > 4 months, this was associated with lower odds of emotional and conduct scores.

As well as the reported effects on cognition, breastfeeding has also been shown to have an impact on white matter development in the brain (406, 407). In 8 year old children, the breastfeeding factor has also been shown to have a significant effect on specific white matter tracts (408) and these tracts are similar to those altered in 8 year olds with ADHD (409). Breastfeeding has been identified to not be associated with ADHD, but there appears an association when a shortened duration of breastfeeding occurred (410).

Hong et al. (411) report on the conflicting evidence for whether breastfed or bottle fed infants may exhibit more ASC characteristics. Children who were not breastfed have been shown to be more likely to develop ASCs (412) and that even down to late initiation (> one hour) after birth has shown increased risk of developing ASCs (413). Further links to brain development can be identified in Steinman and Mankuta's review (414), it was discussed how ASCs can be affected by dysmyelination which can be brought on by an inadequate supply of insulin-like growth factors in newborns. They concluded that breastfeeding was a method for increasing insulin like growth factor and thus may ease symptoms of ASCs.

In 2000, it was reported that 69% of women began breastfeeding (415), currently 81% of mothers initiate breastfeeding (416). Although the CATS II prevalence was around 61-62% breastfeeding over one month during the period of 2003-2007, it has been identified that Welsh rates of breastfeeding are generally lower to that found in the UK (417). In 2005, 48% of mothers were breastfeeding at 6 weeks, and 25% continued to six months (418), therefore the CATS II breastfeeding rates being lower than the 2000 prevalence, could be due to CATS II focusing on those breastfeeding over one month.

Breastfeeding summary for hypotheses formulation

The literature appears united that breastfeeding would be best when considering the child's intelligence and cognition (380-384, 386-390, 402, 403). As the CATS II study did not collect maternal IQ data or maternal education, this important covariate could not be controlled for. It was hypothesised that children who were breastfed for over one month would have higher IQs and NEPSY-II memory and fine-motor coordination scaled scores, compared to those who were breastfed less than one month (as these also include those who were bottle fed). A similar hypothesis was adopted for the results from the behavioural questionnaires, that those whom were breastfed over one month would have lower SDQ, ADHD and ASC scores from the three questionnaire tools (based on the above literature (404, 410, 412-414)).

### 2.1.2.4.     Where the child was assessed (model 4)

Where the child was assessed was not adjusted in the behavioural questionnaire analysis as this was proposed not to affect the scoring by the mothers. Therefore, no literature search or hypothesis generation was conducted to investigate any possible effects on the behaviour of the children.

There was little research available detailing any possible effects of the testing environment on cognitive assessments. Both the WISC-IV (250) and NEPSY-II (230) manuals denote the importance of a distraction, noise-free environment to ensure optimal testing. To my knowledge (by extensive literature searching), no articles had explored the possibility of a significant effect of the testing environment on assessment outcome.

In a book by Kamphaus (216), he noted what extraneous factors psychologists should attempt to control for during cognitive testing. Kamphaus provided a list of the ideals of a testing room (pg. 96):

"The testing room should;

- Free from interruptions
- Be pleasantly, but minimally decorated so as not to distract the child
- Be well lit (…)
- Have adequate ventilation
- Be quiet (…)
- Be a few degrees cooler than a room meant for adults (…)"

It was further noted in Kamphaus' book (216) that the optimal time for testing may be during the morning period. This was almost impossible for most home visits in CATS II, as the children were in school. A problem encountered was children being fatigued from their

school day by the time I arrived to conduct the assessment battery; this was sometimes mentioned by the mothers. Other activities preceding the assessment may have had an effect on the test outcome, for example whether they were preceding the test with calm story time, or high energetic play (419).

All of the research centre tests occurred over the morning period, and only had the child and myself in the room, bar one assessment. The home visits contained a greater number where the mother was present during testing for part of or the whole of the assessment (unfortunately this was not noted after the assessments as this may have been an interesting covariate to investigate). The WISC-IV and NEPSY-II administration manuals both discussed the negative impact the presence of a parent or additional adult may have on the test outcome. If a child was ≤ 3 years of age, a parent could be present to aid the assessment (216). With the older child, and what I instinctively aimed to achieve if a parent was present during testing, was to have the adult sat away from the examinee's line of sight to minimise the distraction (419).

### 2.1.2.4.1.    Place of assessment summary for hypotheses formulation

As differences between scores on cognitive tests were not well reported in the literature, it was difficult to develop a hypothesis for this covariate. As the research centre cognitive assessments (clinical environment) were conducted in a controlled setting, where aspects such as distractions in the room could be more easily controlled for, it was predicted that those assessed at their home would perform worse that those assessed at the research centre.

### 2.1.2.5.    *Child's language at school and home (model 4)*

As previously mentioned, the covariate of child's language spoken at school and home was controlled for. This was because the large majority of assessments occurred in Wales and many children attended Welsh-speaking schools. There were several combinations of school and home languages in the study;

a. English school and English at home
b. Welsh school and English at home
c. Welsh school and Welsh at home
d. English school and other language (not Welsh or English) spoken at home
e. Welsh school and other language (not English or Welsh) spoken at home

Further information can be found in Tables 13 and 20 about how many participants fell into which combination of languages. As those attending Welsh schools will have done-so in a

one-language format, i.e. solely in Welsh, for the purposes of this literature search, were seen as bilingual. Children from the categories d. and e., will have been trilingual; but few fell into these classifications, for example, only one child in CATS II attended a Welsh school and spoke an additional language at home, not Welsh or English. As the language of the child's school or home-life were not controlled during the behavioural questionnaire data analysis, no literature search or hypothesis generation were conducted for differences between mono-and-bilingual children's behaviour.

The literature appeared conflicting for studies that had investigated specific cognitive differences between monolingual and bilingual children. Bilingual children had been reported to perform better on cognitive measures compared to monolingual children (420). However, bilingual children appeared to consistently perform worse for their VC domain when compared to monolingual children (420-422). There was research to suggest however, that on the PR domain of functioning (specifically investigating the demands of executive functioning tasks), bilingual children outperformed their monolingual peers (423-427). Specific to the WISC-IV, differences in favour of bilingual compared to monolingual individuals had been found on the Block Designs and Vocabulary subtests (428, 429). The possible reason for this 'over performance' in nonverbal abilities was that the bilingual mind would have required excellent executive resources to be able to select the correct language at the given time, and not make use of the other language (430, 431). However, others believed that both languages could be active in a bilinguals mind, when the intention was to only use one of them (432).

Such studies looking at the different cognitive aspects of bilingual children have been criticised for not controlling for social deprivation factors whilst also having small sample sizes (430, 431, 433). Conversely when social deprivation factors were controlled for, consistent results of a lower VC to higher PR domain for bilingual compared to monolingual children has been identified (433) (also supported by (420, 434)).

Few studies have investigated any bilingual effects from the NEPSY-II. Korkman et al. (435) administered specific NEPSY-II subtests to bilingual and monolingual children and found no differences (including the NM subtest). This was replicated in a study administering all 14 NEPSY-II subtests and found no differences on ten subtests (including FTDH, FTNDH and NM), but bilingual children outperformed on 'Imitating Hand Positions' and 'Design Copy', and performed worse on two of the verbal subtests (436). Recently, a study was conducted using the ten core subtests of the WISC-IV and nine selected NEPSY-II subtests (same versions as

current thesis), on a sample of Finnish bilingual children (n = 100) (437). The only significant difference that Karlsson et al. identified, was on the Symbol Search subtest of the WISC-IV, monolingual children outperformed their bilingual counterparts ($p$ = . 003). It is unclear in the literature what the impact is of being educated in a language different, from an individual's home language, in respect of test performance; for example would it be better to have an English test in a pupil's school or home language?

### 2.1.2.5.1. Child's language at school and home summary for hypotheses formulation

Based on the above literature, it was hypothesised that the mixed language children in CATS II would have a lower VCIQ compared to the monolingual children (420-422, 433, 434). It was also predicted that the bilingual children would have a higher PRIQ (420, 423-427, 433, 434) compared to the monolingual children. There would be no differences for WMIQ, PSIQ or FSIQ however, as this was sparsely reported upon in the literature. As no differences were identified for the NM subtest of the NEPSY-II (435-437), it was predicted that no further differences would be found between the participants for all of the NEPSY-II subtests.

### 2.1.2.6. Social deprivation (model 4)

There is a long established belief that those children from a higher rated social (and/or economic) backgrounds, exposed to luxuries of particular goods, services and social connections that could benefit them, are at an advantage compared to those children from a more socially deprived background (438). Social deprivation has been reported to be related to cognitive performance throughout an individual's childhood (439-441). There is a reported mean cognitive decline with decreasing social class, which was more predominant in the VC domain (442).

Differences of ability from social deprivation can be seen before the child begins schooling (440, 443). Once they get to school age, lower income families (therefore with a lower score of social deprivation) were more likely to send their child to a lower quality school (444, 445). However, if the child had a preschool enrolment before the age of 4 years, there was an association to reduced family-level influences on early reading and maths skills at age 5 years (443).

Social deprivation may affect different ages of children in different ways. This effect had been found to be reduced on children aged 5 compared to those at aged 3 years (442). White (446) agreed that the link between the background of the child and school attainment diminished with age. However, it was been found that differences in cognitive and non-cognitive skills were apparent in children of different social deprivation at ages 4 and 6 years (447) and

Smith et al. (448) identified that there were no differences for achievement between 7 and 3 year olds when controlling for social deprivation. Therefore, the impact of social deprivation may be more apparent during the first few years of schooling, longitudinal studies would be needed to investigate this further.

As well as cognition, social deprivation has also been reported to affect the child's behaviour, with those from a more socially deprived background exhibiting more behavioural difficulties (449, 450). It was noted in the literature that problems may begin to emerge in early childhood with externalising problems becoming apparent in middle childhood (451-453). In a large UK cohort sample, it was identified that those from lower socially deprived backgrounds were reported to have higher mean results on total SDQ scores for all domains, except for prosocial scale, compared to those from less socially deprived backgrounds (454). One possible reason for the correlation that has been suggested is diet (455).

### 2.1.2.6.1. Social deprivation summary for hypotheses formulation

Based on the above literature, it was hypothesised that the covariate of social deprivation (calculated by social-deprivation postcode scores from StatsWales (281) and OpenDataCommunities (282)) would have a significant effect on the child's WISC-IV and NEPSY-II scores with those from a less socially deprived background performing better than those from a more socially deprived background (439-442). In relation to the three behaviour questionnaires, it was also predicted that social deprivation would have a statistically significant effect on the outcome scores, with more frequent indications of disruptive behaviours from those of a more socially deprived background (449, 450, 454).

### 2.1.3. Hypotheses

The section above described the rationale behind each hypothesis proposed of the possible significant effects of the covariates on the three main measures. Below are tables containing the grouped hypotheses by measure, i.e. the WISC-IV, followed by the NEPSY-II and then the behavioural questionnaires (Tables 31, 32 and 33).

Table 31

*Covariate Hypotheses for the Wechsler Intelligence Scale for Children- Fourth Edition, UK (WISC-IV)*

| Covariates | WISC-IV | | | | |
|---|---|---|---|---|---|
| | VCIQ | PRIQ | WMIQ | PSIQ | FSIQ |
| Gender | Males would score higher than females | No difference | | Females would score higher than males | No difference |
| Mother age | The older the mother was at time of pregnancy, the higher the IQs would be | | | | |
| Breastfeeding | If breastfed, there would be higher IQs compared to those who were not breastfed for over 1 month | | | | |
| Where child was assessed | Those assessed at the research centre would perform better than those assessed at their homes | | | | |
| School and home language | Bilingual children would have lower a VCIQ compared to monolingual children | Bilingual children would have a higher PRIQ compared to monolingual children | No difference | | |
| Social deprivation score | Lower rating* would yield lower IQs compared to those who had a higher social deprivation score | | | | |

Note. *lower ratings=more socially deprived. VCIQ=verbal comprehension intelligent quotient, PRIQ=perceptual reasoning intelligence quotient, WMIQ=working memory intelligent quotient, PSIQ=processing speed intelligent quotient, FSIQ=full scale intelligent quotient.

Table 32

*Covariate Hypotheses for the Developmental Neuropsychological Assessment (NEPSY)-II*

| Covariates | NEPSY-II | |
| --- | --- | --- |
| | Memory | Motor-coordination |
| Gender | No difference | |
| Mother age | The older the mother was at time of pregnancy, the higher the scaled scores would be compared to those born to younger mothers | No difference |
| Breastfeeding | If breastfed, there would be higher scaled scores compared to those who were not breastfed over 1 month | |
| Where child was assessed | Those assessed at the research centre would perform better than those assessed at their homes | |
| School and home language | No difference | |
| Social deprivation score | Lower rating would yield lower scaled scores compared to those whom had a higher social deprivation score | |

Table 33

*Covariate Hypotheses for the Behavioural Questionnaires*

| Covariates | Questionnaires | | |
| --- | --- | --- | --- |
| | SDQ | Child ADHD Questionnaire | SCQ |
| Gender | No difference: except on conduct and peer problems where males would score higher | Males to have higher mean scores than females | |
| Mother age | The younger the mother was at time of pregnancy, the higher the mean scores would be | | The older the mother was, the higher the score |
| Breastfeeding | If breastfed over 1 month, there would be lower mean scores obtained compared to those who were not | | |
| Social deprivation score | A lower rating of social deprivation would yield higher scores on all questionnaires compared to a high rating | | |

Note. SDQ=strengths and difficulties questionnaire, ADHD=attention deficit hyperactivity disorder, SCQ=social communication questionnaire.

### 2.1.4. Method and Statistical Analysis

The data used in this chapter was from the CATS II study. Information about covariates was collected from the 'CATS General Questionnaire'. The dependent variable data collection has

been previously described (for the WISC-IV see chapter 1.4., Intelligence measured at age 9; CATS II data, for the NEPSY-II see chapter 1.5., Additional cognitive assessments at age 9; CATS II data, and for the behavioural questionnaires see chapter 1.6., Behavioural questionnaires at age 9; CATS II data).

The following statistical analyses were conducted in IBM SPSS statistics version 20. The indication that any of the six covariates were having a significant effect on the dependent variables was identified primarily from the MANCOVA outputs (primary analyses only). Once they had been identified, covariates were explored for their effects. If the covariates only had two groups of classification they were explored by t-test. This included gender (male or female), whether the child was breastfed for longer than one month (yes or no) and where the child was assessed (home or research centre). If the covariate had two or more groups, a MANOVA was adopted to enable investigation of either all WISC-IV, NEPSY-II or behaviour questionnaire scores in one main analysis. These included, mother age at time of consent into CATS I (data was in quartiles), the child's language spoken at school and home (English school and English home, Welsh school and English home, Welsh school and Welsh home, English school and other home language or Welsh school and other home language) and also the social deprivation score (data was quintiled). Any significant MANOVA results were explored by Bonferroni corrected post hoc tests. As discussed in chapter 1.6. (Behavioural questionnaires at age 9; CATS II data), child age was also adjusted in the primary MANCOVA analysis. This covariate did not have a significant effect on the behavioural questionnaires ($p$ = .221), and was not explored further.

An ANOVA was used for exploring the LM subtest from the NEPSY-II to keep the analysis similar to chapter 1.5. (Additional cognitive assessments at age 9; CATS II data), this was because a smaller group completed the LM subtest and by adding it to the NEPSY-II MANOVA, it would have lost overall statistical power by only including those children who completed all six NEPSY-II subtests.

### 2.1.5. Results
The following results were organised by the data collection tools; WISC-IV, NEPSY-II and behavioural questionnaires.

#### 2.1.5.1. WISC-IV
All of the covariates except for the location of where the child was assessed ($\Lambda_{ROY}$ = .022, $F$ (5, 436) = 1.891, $p$ = .095, $\eta_p^2$ = .021) had a significant effect on the IQs.

- *Gender*

As identified by the MANCOVA at the fourth level of analysis, child gender had a significant effect on IQs obtained: $\Lambda_{ROY}$ = .083, *F* (5, 436) = 7.258, *p* < .001, $\eta_p^2$ = .077. T-tests revealed that females performed significantly better than males for three IQs: WMIQ (*p* = .031*,* 95% CI [-4.612, -.216]), PSIQ (*p* = <.001*,* 95% CI [-8.554, -3.986]) and FSIQ (*p* = .029*,* 95% CI [-4.714, -.249]). VCIQ and PRIQ yielded results of *p's* >.756. See Table 34 below for further information of the gendered mean IQs.

Table 34
*Male and Female Intelligent Quotient (IQ) Means*

| IQ domain | Childs gender | N | Mean |
|---|---|---|---|
| WISC-IV VCIQ | male | 232 | 99.22 |
| | | | (12.08) |
| | female | 220 | 99.17 |
| | | | (10.49) |
| WISC-IV PRIQ | male | 232 | 104.76 |
| | | | (12.74) |
| | female | 220 | 105.13 |
| | | | (12.41) |
| WISC-IV WMIQ | male | 232 | 98.68 |
| | | | (11.79) |
| | female | 220 | 101.10 |
| | | | (11.98) |
| WISC-IV PSIQ | male | 232 | 99.95 |
| | | | (12.15) |
| | female | 220 | 106.22 |
| | | | (12.55) |
| WISC-IV FSIQ | male | 232 | 101.33 |
| | | | (12.24) |
| | female | 220 | 103.81 |
| | | | (11.89) |

Note. Standard deviations appear in parentheses below means. WISC-IV=Wechsler intelligence scale for children-fourth edition, UK, VCIQ=verbal comprehension intelligent quotient, PRIQ=perceptual reasoning intelligence quotient, WMIQ=working memory intelligent quotient, PSIQ=processing speed intelligent quotient, FSIQ=full scale intelligent quotient.

- *Mother age*

Mothers age at time of first consent into CATS I by Roy's largest root at the MANCOVA fourth level of analysis revealed a significant effect on the children's IQs, $\Lambda_{ROY}$ = .056, *F* (5, 436) = 4.912, *p* < .001, $\eta_p^2$ = .053. The exploratory MANOVA of the covariate revealed a sustained significant affect, $\Lambda_{ROY}$ = .119, *F* (5, 446) = 10.586, *p* < .001, $\eta_p^2$ = .106. At the univariate level, all IQs were significantly affected by age of mother: VCIQ- *F* (3, 448) = 12.926, *p* < .001, $\eta_p^2$ = .080, PRIQ- *F* (3, 448) = 6.573, *p* < .001, $\eta_p^2$ = .042, WMIQ- *F* (3, 448) = 5.678, *p* = .001, $\eta_p^2$ =

.057, PSIQ- $F$ (3, 448) = 11.238, $p$ < .001, $\eta_p^2$ = .070 and FSIQ- $F$ (3, 448) = 14.680, $p$ < .001, $\eta_p^2$ = .090. The overall trend was that the older the mothers were at time of consent into CATS I, i.e. during their pregnancies, the higher the mean IQs were for the children (see Figure 17 for the graph showing the means). The strongest significance was evident between quartile group one to all other ages. The CATS I sample was grouped so that group one included ages of 14-24, group two: 25-29, group three: 30-33 and group four was ages 34-49 years. From the graph it appeared that after the initial steep increase between groups one to two, the IQs did still increase, but the increase was no-where near as dramatic.
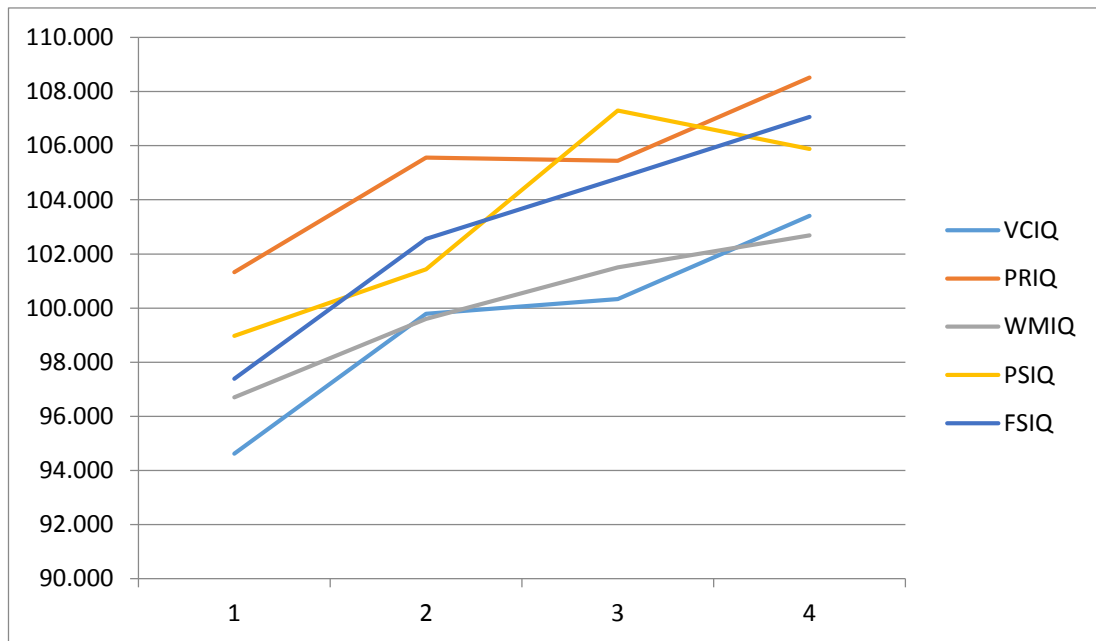


Figure 17: Graph to show trend of mean Intelligent Quotients by mother age quartiles

VCIQ=verbal comprehension intelligent quotient, PRIQ=perceptual reasoning intelligence quotient, WMIQ=working memory intelligent quotient, PSIQ=processing speed intelligent quotient, FSIQ=full scale intelligent quotient. X axis=mother age quartiles,1 lowest. Y axis=mean IQ scores.

- *Breastfeeding*

Whether the child was breastfed over one month also had a significant effect on the child's IQ scores by the MANCOVA adjusted by model four of the analysis, $\Lambda_{ROY}$ = .028, $F$ (5, 436) = 2.417, $p$ = .035, $\eta_p^2$ = .027. An exploratory t-test revealed that all IQs except for PSIQ ($p$ = .060) were significantly affected by the covariate of breastfeeding ($p$'s < .006). On average, those who were breastfed had a 5 point higher means for VCIQ, PRIQ and FSIQ compared to

those who were not, and WMIQ yielded a 3 point advantage to the breastfed over one month group. See Table 35 for further details of *p* values and confidence intervals.

Table 35
*Breastfeeding Significance and Confidence Intervals on the Dependent Variables from the Wechsler Intelligence Scale for Children-Fourth Edition, UK (WISC-IV)*

| IQ | Sig. (2-tailed) | Mean Difference | 95% Confidence Interval of the Difference | |
|---|---|---|---|---|
| | | | Lower | Upper |
| VCIQ | .000* | -4.837 | -6.958 | -2.716 |
| PRIQ | .000* | -5.042 | -7.399 | -2.684 |
| WMIQ | .006* | -3.202 | -5.470 | -.934 |
| PSIQ | .060 | -2.329 | -4.756 | .097 |
| FSIQ | .000* | -5.383 | -7.647 | -3.118 |

Note. *Significance < .05. VCIQ=verbal comprehension intelligent quotient, PRIQ=perceptual reasoning intelligence quotient, WMIQ=working memory intelligent quotient, PSIQ=processing speed intelligent quotient, FSIQ=full scale intelligent quotient.

- *Child's language at school and home*

The covariate of child's language at school and home was also significant, $\Lambda_{ROY}$ = .039, $F$ (5, 436) = 3.408, $p$ = .005, $\eta_p^2$ = .038. The exploratory MANOVA also revealed a significant effect of language of school and home on IQ, $\Lambda_{ROY}$ = .060, $F$ (5, 446) = 5.355, $p$ <.001, $\eta_p^2$ = .057. Conversely, under separate univariate analysis, none of the five IQs reached significance (all *p*'s >.084). A reason for this was that the multivariate test takes account of the correlation between the many dependent variables, therefore the data has more power to detect group differences; when compared to the univariate tests (210).

- *Social deprivation*

The final adjustment to the analysis was for social deprivation. Using Roy's largest root, there was a large significant effect on IQs for social deprivation scores, $\Lambda_{ROY}$ = .089, $F$ (5, 436) = 7.784, $p$ < .001, $\eta_p^2$ = .082. The exploratory MANOVA revealed a continued significance, $\Lambda_{ROY}$ = .146, $F$ (5, 446) = 12.999, $p$ <.001, $\eta_p^2$ = .127. At the univariate level, all IQs were still affected by the covariates (all *p*'s <.001). The overall trend was that the more socially deprived the participants were (i.e. those with a lower quintiled score, see Tables 9 and 10 for further information), the worse the child's performance on the IQ domains was (see Figure 18). There did appear to be a spike of VCIQ for the social deprivation score classification two, but this may have been a type 1 error as it was only apparent on this domain.
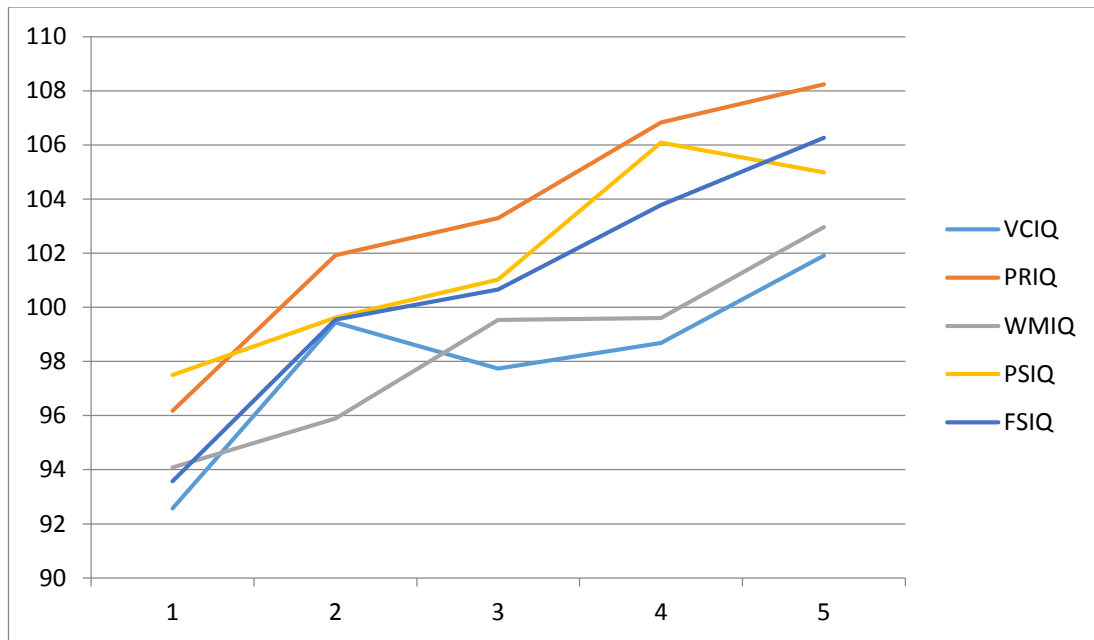
Figure 18: Graph to show trend of mean Intelligent Quotients by social deprivation quintiles

VCIQ=verbal comprehension intelligent quotient, PRIQ=perceptual reasoning intelligence quotient, WMIQ=working memory intelligent quotient, PSIQ=processing speed intelligent quotient, FSIQ=full scale intelligent quotient. X axis= social deprivation score quintiles, 1 represents the most socially deprived, Y axis=mean IQs.

### 2.1.5.2. NEPSY-II

As can be seen in Table 36 for the LM ANCOVA (as adjusted for by model four of the analysis), only child gender and social deprivation had a significant effect on the combined working and LTM subtest from the NEPSY-II. Gender was investigated by an independent samples t-test. Females performed significantly better than males, receiving a mean of 11.30 for LM compared to the males achieving 10.28 ($p$ = .001, 95% CI [-1.643, -.403]. The effects of the social deprivation were investigated by ANOVA (as five groups of quintiled ratings), $F$ (4, 312) = 3.802, $p$ = .005, $\eta_p^2$ = .046. The social deprivation scores revealed that the more deprived the participants were, the worse they performed (see Table 37 below).

Table 36
*Covariate Effects on List Memory and List Memory Delayed Subtest*

| Source | df (error=307) | F | Sig. | Partial Eta Squared |
|---|---|---|---|---|
| Child Gender | 1 | 10.982 | .001* | .035 |
| Mother Age Quartiles | 1 | .398 | .528 | .001 |
| Social Deprivation Quintiles | 1 | 11.264 | .001* | .035 |
| Where Assessed | 1 | .009 | .923 | .000 |
| Language at school & home | 1 | .004 | .950 | .000 |
| Breast fed ≥ 1month | 1 | 1.020 | .313 | .003 |

Note. *Significance < .05. NEPSY-II= developmental neuropsychological assessment, df=degrees of freedom.

Table 37
*Social Deprivation Quintiles and Means for the Subtest List Memory and List Memory Delayed, from the Developmental Neuropsychological Assessment (NEPSY-II)*

| Social Deprivation Quintiles of English and Welsh. Score 1 low-5 | N | Mean |
|---|---|---|
| 1 | 44 | 9.59 (3.00) |
| 2 | 40 | 10.30 (2.80) |
| 3 | 48 | 10.52 (3.22) |
| 4 | 63 | 11.08 (2.65) |
| 5 | 122 | 11.34 (2.62) |
| Total | 317 | 10.79 (2.85) |

Note. Standard deviations appear in parentheses below means.

For the MANCOVA analysis involving the other NEPSY-II tests, only gender ($\Lambda_{ROY}$ = .041, *F* (5, 397) = 3.236, *p* = .007, $\eta_p^2$ = .039) and language spoken at school and home ($\Lambda_{ROY}$ = .046, *F* (5, 397) = 3.675, *p* = .003, $\eta_p^2$ = .044) had a significant effect on the participants. All other covariates did not have a significant effect on the NEPSY-II scores (*p*'s > .111); where the child was assessed, mother age at consent in CATS I, whether the child was breastfed for longer than one month, and the social deprivation score.

- *Gender*

Gender was assessed for the five NEPSY-II subtests by an independent samples t-test. Similar to the ANCOVA, it was females who outperformed males on FTDH ($p$ =.043, 95% CI [-.599, -.009]) with females achieving a mean of 12.26 compared to males achieving on average 11.95. Additionally, the subtest NM was significant, $p$ =.001, 95% CI [-1.447, -.395], with females generating a mean score of 11.79 to the males' score of 10.87. The t-test revealed non-significant differences for gender ($p's$ >.417) on the remaining NEPSY-II tests (MD, MDD and FTNDH).

- *Child's language at school and home*

For language at school, a MANOVA was adopted to investigate the covariate for the five tests. Using Roy's largest root significance was still seen across the CATS II group: $\Lambda_{ROY}$ = .051, $F$ (5, 406) = 4.134, $p$ = .001, $\eta_p^2$ = .048. At the univariate level, it was found that it was the FTDH subtest that was significantly affected by the child's language at school: $F$ (4, 407) = 3.490, $p$ = .008, $\eta_p^2$ = .033 (other subtests generated $p$'s >.057). As can be seen in Table 38 below, the highest mean attained was from those that attended a Welsh school, and spoke Welsh at home. Separate (Bonferroni corrected) t-tests could not have been computed on this covariate as only one participant attended a Welsh school with other language (not English or Welsh) spoken at home.

Table 38
*Mean Score of Fingertip Tapping Dominant Hand by Language Spoken at School and Home*

| Language at school | N | Mean |
|---|---|---|
| English school and English at home | 330 | 11.97 (1.55) |
| Welsh school and English at home | 72 | 12.51 (1.41) |
| Welsh school and Welsh at home | 10 | 13.20 (1.32) |
| English school and other language (not Welsh or English) spoken at home | 3 | 12.67 (1.15) |
| Welsh school and other language (not English or Welsh) spoken at home | 1 | 11.00 (*) |

Note. Standard deviations appear in parentheses below means. *no standard deviation available as only one participant with this language classification. NEPSY-II= developmental neuropsychological assessment.

### 2.1.5.3. Behavioural questionnaires

The MANCOVA revealed that child gender ($\Lambda_{ROY}$ = .158, $F$ (9, 450) = 7.916, $p$ < .001, $\eta_p^2$ = .137) and social deprivation ($\Lambda_{ROY}$ = .065, $F$ (9, 450) = 3.229, $p$ = .001, $\eta_p^2$ = .061) had a significant effect on the child behavioural questionnaires; the other covariates of mother age

at time of consent into CATS I and whether the mother breastfed over one month, were non-significant (*p*'s > .052).

- *Gender*

The significant effect of gender on the child behaviour questionnaires was explored further by a t-test. There were significant interactions for all questionnaire outcomes except SDQ emotion and SDQ peer problems (*p*'s > .110).  Table 39 displays the means, significance and confidence intervals for the remaining questionnaires. As is seen in the table, males performed significantly worse than females across the board (a higher score indicated a greater behavioural difficulties).

Table 39

*Significant Differences Between Genders for the Behavioural Questionnaire*

| Questionnaire | Gender | N | Mean | Sig. | 95% Confidence Interval | |
|---|---|---|---|---|---|---|
| | | | | | Lower | Upper |
| SDQ Conduct | Male | 236 | 1.59 (1.75) | .002 | .166 | .740 |
| | Female | 233 | 1.13 (1.39) | | | |
| SDQ Hyperactivity | Male | 236 | 3.82 (2.90) | .001 | .378 | 1.351 |
| | Female | 233 | 2.96 (2.44) | | | |
| SDQ Total Difficulties | Male | 236 | 9.10 (6.55) | .023 | .177 | 2.350 |
| | Female | 233 | 7.83 (5.37) | | | |
| SDQ Prosocial* | Male | 236 | 8.40 (2.00) | < .001 | -1.092 | -.472 |
| | Female | 233 | 9.18 (1.36) | | | |
| ADHD Inattention | Male | 237 | 7.29 (6.14) | < .001 | 1.105 | 3.137 |
| | Female | 232 | 5.17 (5.01) | | | |
| ADHD Overactivity | Male | 237 | 3.09 (2.82) | < .001 | .932 | 1.855 |
| | Female | 232 | 1.70 (2.23) | | | |
| ADHD Impulsivity | Male | 237 | 3.85 (3.20) | .001 | .400 | 1.480 |
| | Female | 232 | 2.91 (2.73) | | | |
| ADHD Total | Male | 237 | 14.24 (11.06) | < .001 | 2.655 | 6.254 |
| | Female | 232 | 9.78 (8.65) | | | |
| SCQ | Male | 234 | 5.33 (4.14) | < .001 | .960 | 2.336 |
| | Female | 232 | 3.68 (3.38) | | | |

Note. Standard deviations appear in parentheses below means. *females had a higher mean, prosocial scale was the only one with 'reverse' marking, i.e. a higher score indicated a more desirable outcome. SDQ=strengths and difficulties questionnaire, ADHD=attention deficit hyperactivity disorder, SCQ=social communication questionnaire.

- *Social deprivation*

The significant effect of social deprivation was explored by a MANOVA; $\Lambda_{ROY}$ = .106, $F$ (9, 456) = 5.371, $p$ = > .001, $\eta_p^2$ = .096. It was identified that all, except three mean scores, were

significantly affected by social deprivation; only SDQ prosocial, ADHD inattention and ADHD total scores were not significantly affected ($p's > .075$). See the graph below (Figure 19) for the general trend of the significantly affected questionnaire domain scores.
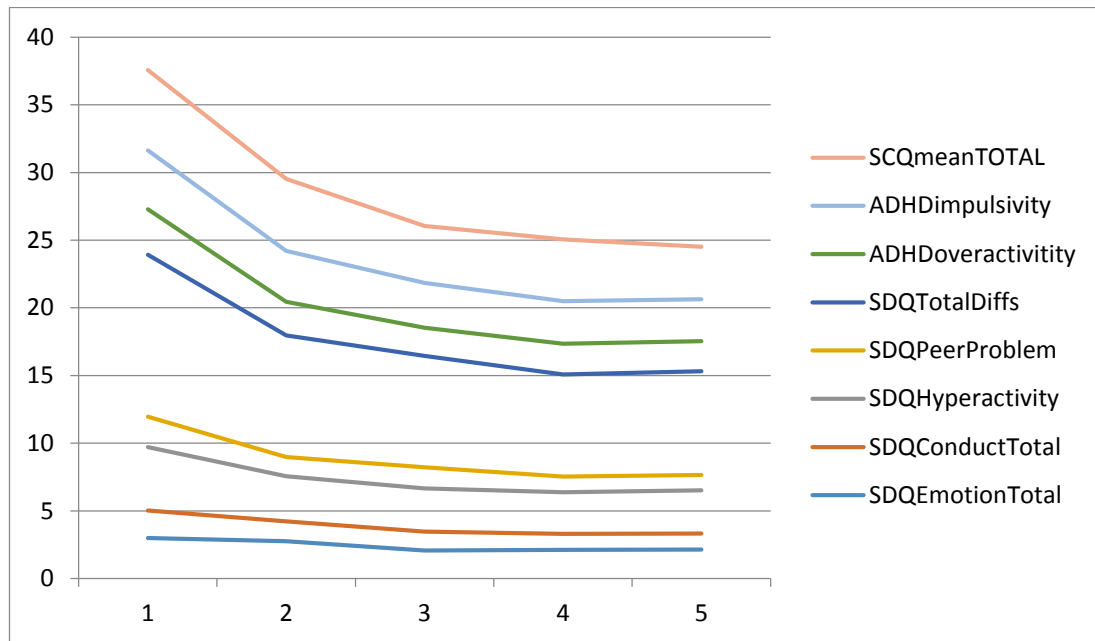


Figure 19: Graph to show mean scores by social deprivation score

NOTE: Only those questionnaire domains that were significantly affected by social deprivation score are presented. Scale on the Y axis is arbitrary as it includes the total scores of the three questionnaires. Reader is advised to look at one particular scale at a time for a true reflection of the relationship between social deprivation ratings to questionnaire outcome. SDQ=strengths and difficulties questionnaire, ADHD=attention deficit hyperactivity disorder, SCQ=social communication questionnaire. X axis= social deprivation score quintiles, a score of 1 represented a more socially deprived rating, Y axis= means of questionnaires.

### 2.1.6. Discussion

The current chapter investigated whether the covariates (child gender, whether the mother breastfed over one month, age of mother at time of consent into CATS I, where the child was assessed, language of the child's school and home, and social deprivation) had any effect on the dependent variables from the WISC-IV, NEPSY-II and behavioural questionnaires. It was hypothesised that all of the covariates would have had some effect to the dependent variables, this was confirmed apart from the covariate of where the child was assessed. The biggest effects to the WISC-IV, NEPSY-II and behaviour questionnaires, were child gender and social deprivation. Discussions below are organised by the dependent variable measures.

It was hypothesised that the WISC-IV scores would be significantly affected by child gender. The hypotheses generated in Table 31 based on the literature, were all unfounded except for PSIQ, where females performed significantly higher compared to males (328, 332). This result was also identified for WMIQ and FSIQ. The hypotheses of breastfeeding over one month yielding higher IQs in offspring compared to those who were not breastfed (380-384, 386-390, 402, 403), the older the mother at time of pregnancy the higher the IQs (360, 361, 363, 364) and the less socially deprived the participants were, the higher the IQs (439-442), were all confirmed by the above analysis. It was predicted that those who were assessed in the research centre would have higher IQ scores than those assessed at their homes, as the home environment would include more distraction, and therefore not be an optimal testing space (216, 230, 250). The testing environment appeared to have no effect at the multivariate level to IQs measured on the WISC-IV ($p$ = .095). The final hypothesis of the effects of bilingualism on IQ was not confirmed at the multivariate level. It was predicted that bilingual children would have lower VCIQs than monolingual children (420-422, 433, 434), but perhaps have a higher PRIQ compared to monolingual children (420, 423-427, 433, 434), with no difference between the other IQ scores. There was a difference at the multivariate level but not at the univariate level, a possible reason for this was that the data had more power to detect group differences at the multivariate level (210).

Table 32 displayed the anticipated outcome of no gender differences for the NEPSY-II, as it was evidenced in the literature that no previous differences had been identified for memory or motor-coordination (332). Females outperformed the males on LM, FTDH and NM subtests. This added validity to the WISC-IV gender findings, as significant differences for WMIQ were identified with females outperforming males. As mentioned in chapter 1.5. (Additional cognitive assessments at age 9; CATS II data), the LM subtest was analysed separately, consequently the effects of the covariates on the NEPSY-II were also analysed separately in the current chapter. Social deprivation rating was predicted to have a significant effect on the NEPSY-II scores (439-442), this was confirmed in the LM analysis, but confirmed other literature of a null hypothesis in the main MANCOVA of the NEPSY-II subtests (442, 446-448). As the literature had only identified language differences for the NM subtest of the NEPSY-II (435-437), it was predicted that there would be no differences on the NEPSY-II; as the covariate was analysed at the multivariate level, a subtle difference would be undetectable unless pronounced. Upon analysis, I identified that those who attended Welsh schools and who spoke Welsh at home, had higher FTDH results than those from other language combinations. As mentioned in the results section, no post hoc analysis

could be completed as one of the groups only contained one participant, it could be inferred then that this result was a false positive as the analysis was not conducted on groups with close/equal numbers in. No significant effects of whether the mother breastfed over one month or where the child was assessed were identified here, which contradicted the literature ((380-384, 386-390, 402, 403) and (216, 230, 250), respectively). The covariate of mother age at time of consent into CATS I, was predicted to have no effect on motor coordination (363), this was confirmed in the current analysis. However for the memory tests, it was predicted that the older the mother was, the higher the scores would be for the offspring alike to the WISC-IV predictions; this was unfounded in the current analysis.

Hypotheses for the child behavioural questionnaires can be found in Table 33. Gender had a significant effect on all of the questionnaires, and on every domain measured. The males received significantly higher scores, indicating 'worse' behaviour (except for SDQ prosocial where scoring was inverted), compared to females. This was confirmed in the literature for ADHD and ASCs (128, 341-345, 347-350), and subsequently was adopted as the hypothesis. For the SDQ, no differences for gender were anticipated, except for peer problems and conduct where males would score higher (338, 339). The current analysis confirmed the previous literature, but also found males presenting more difficulties across the domains of the SDQ. The only other covariate that had a significant effect was social deprivation. It was predicted that the lower the rating, the 'worse' the behaviours (i.e. higher scores) would be for all of the questionnaires (449, 450, 454). This was confirmed in the analysis, and Figure 19 displays a graph to illustrate the trend of scores by social deprivation. The rating appeared to have no effect on SDQ prosocial, ADHD inattention or ADHD total scores. Mother age at time of consent into CATS I and whether the mother breastfed over one month did not have a significant effect on the behavioural questionnaires. This contradicted the literature, it had been found that if the mother was young at time of pregnancy, there would be more behavioural difficulties reported (361, 365-367, 370), but for ASC traits it was an inverse relationship, i.e. the older the mother, the more likely it was that the offspring would have ASC traits (371-376). Similarly the breastfeeding data contradicted the literature, studies showing that better behaviour would be seen in those who breastfed for a period (404, 410, 412-414). However, it is worth noting that the measures here were questionnaires and perhaps further investigations of assessments for specific behaviour difficulties or SENs may yield a link to breastfeeding.

The WISC-IV was the only measure in CATS II significantly affected by the covariates of age of mother at time of consent into CATS I and whether the mother breastfed over one month.

Given the literature discussed in the introduction evidencing effects from mother age and breastfeeding, this was surprising. One of the reasons no differences were identified could have been because the NEPSY-II comparison included subtests, rather than the WISC-IV which was analysed at the domain level (e.g. VCIQ takes account of three subtests). Conversely, the behavioural questionnaires were domain level as each domain took account of at least five items for a total score.

As mentioned, gender appeared to be the covariate with the largest effect. For cognitive differences, it was always females scoring higher than males. Chen et al. (248) found no differences between males and females on the WISC-V whereas Camarata and Woodcock (332) did identify differences. These two studies were based on American data, thus we can compare findings, but with caution. The behavioural questionnaires completed by the mothers, had males rated as presenting worse behaviour than females. As ADHD and ASCs were measured as more prevalent in males than females (128, 341-345, 347-350), this was not surprising, but with the extension of the SDQ finding a similar discrepancy between the genders, male specific traits may extend into more generalised behaviour.

### 2.1.6.1.  Limitations

One of the limitations of this chapter was that there was a lot of multiple testing. The WISC-IV, NEPSY-II and behavioural questionnaire measures were analysed separately so that the work here would be more comparable to that found in the main CATS II analysis. However, multiple testing in this way could have falsely identified a difference as significant when it should not have been.

If these analyses were conducted again, adjustments could have been made when exploring the six covariates. For example, social deprivation has been previously discussed as an important factor to control for when exploring the effects of maternal age (364), as previously significant findings may not be once adjusted. Furthermore, the covariates and the dependent variables have all been analysed as one complete cohort; analyses were not repeated for the three separate groups of participants. Whilst this has added power to the calculations from a larger cohort, any subtle differences between the groups will have been lost. The results cannot be generalised as the SGTF groups make up around half of the participants.

A limitation of the breastfeeding findings was that maternal education was not controlled for in the analysis. This could have greatly affected the cognition findings (399-401); therefore, they can only be discussed with caution. I did note the importance of maternal

education when I was collecting the CATS II data; unfortunately, this was identified too late in the project and my modified version of the CATS General Questionnaire to include this data was not submitted as an ethical amendment (see appendix 8).

### 2.1.6.2. Conclusions

It was hypothesised that all of the covariates (child gender, mother age at time of consent into CATS I, whether the mother breastfed over one month, where the child was assessed, language at school and home, and social deprivation) would have some effect on the dependent measures of the WISC-IV, NEPSY-II and the behavioural questionnaires. Child gender was reported to have the biggest effect, with females outperforming males on the majority of the cognitive measures, and males scoring higher (indicating a 'worse' behaviour) on the questionnaires. Social deprivation was the second covariate that appeared to have the strongest effect (as displayed in the partial eta squared scores), with scores of being less socially deprived being associated to more desirable outcomes. Where the child was assessed was the only covariate that did not affect the dependent variables; this confounder was not included in the questionnaire analysis.

### 2.1.7. Chapter Summary

This chapter has discussed the findings from the exploratory analysis that investigated the extent to which the covariates had a significant impact on the dependent variables. The literature debating for effects or no effects for each covariate was discussed firstly, then based on the evidence, a series of hypotheses were devised. The analysis was based on the final model MANCOVAs from the WISC-IV (see chapter 1.4., Intelligence measured at age 9; CATS II data), the NEPSY-II (see chapter 1.5., Additional cognitive assessments at age 9; CATS II data), and the behavioural questionnaires (see chapter 1.6., Behavioural questionnaires at age 9; CATS II data); any significance noted from these were explored further by exploratory t-tests and MANOVAs. The results were discussed in respect of the data collection measure and closely tied to the literature. The following chapter explores the associations between the age 3 and age 9 IQs of the CATS treated and untreated SGTF offspring.

## 2.2. IQ comparison between ages 3 and 9; children from the CATS sample

### 2.2.1. Chapter Overview

This chapter contains the second exploratory analysis using data from the CATS cohort; the treated and untreated SGTF groups and the offspring's IQ scores from age 3 (WPPSI-III) and age 9 (WISC-IV). The CATS study, as well as being the first to explore the effect of treatment on SGTF, was also the first to follow up the children assessing them at two time points which presents a unique dataset. The chapter contains a literature review of the development of intelligence over an individual's lifespan, the stability of IQ in young childhood and also a study this chapter's analysis can be compared to.

### 2.2.2. Introduction

#### 2.2.2.1. Intelligence development over the life span

When considering intelligence over an individual's life span, as well as having an understanding of what intelligence is (see pages 5-6), it is important to review how different aspects of intelligence change with age (456). Furthermore, it is important to consider the estimated stability between the multiple IQ measures as this could affect the continuity IQ scores between two time points (457).

##### 2.2.2.1.1. Changes in intelligence

The overall IQ measure from intelligence tests, FSIQ, is often referred to as a representation of 'g'. This general intelligence factor can be subdivided into two categories, fluid intelligence (gf) where we make understandings amongst stimuli, consider the implications and draw our own conclusions/inferences, and crystallised intelligence (gc) which is the measure of the breadth and depth of knowledge about our culture, i.e. general knowledge (458). It has been proposed that the stability of gc maintains and (can) improve through adulthood (459). Conversely, gf peaks in late adolescence/early adulthood then declines, and this may be a result of maturation of the aging brain (460). Schroeders et al. (456) found evidence for this theory, with gc displaying stronger age related gains compared to gf (ages 11-19, n > 10,000). A steady decline in average scores on IQ tests after young adulthood has been identified (461).

The age differentiation hypothesis postulates that general ability gradually breaks down into a group of fairly distinct aptitudes as age increases, i.e. verbal and performance IQ would, with age, become less correlated (462). Studies have identified that as ages increase (> 3 years), correlations decline, suggestive that there was a continuous reduction of the 'g' factor

across ages (463, 464). However, others have struggled to identify this 'loss' in association with increasing age (465, 466).

### 2.2.2.1.2. The Flynn Effect

As mentioned firstly in chapter 1.2. (Re-analysis of intelligence at age 3; UK CATS I cohort), the 'Flynn Effect' is a phenomenon of increasing IQ scores over the years; i.e. as time goes on, people appear to perform better on IQ tests (217-223); the average increase appears to be 3 IQ points per decade, although a recent meta-analysis of 285 studies found an average mean increase of 2.31 points per decade (467). Beaujean and Sheng (468) investigated the Flynn Effect in all versions of the WPPSI and WISC from the norming samples and found larger effects (0.61 and 0.73 increase of points per year, respectively). Caution has been advised when comparing decade increases of IQ as some studies may include children of slightly below or above average intelligence; as Lynn (301) found greater IQ increases in the lower range of ability (fifth percentile) when compared to those in the 95$^{th}$ percentile. There is no one cause for the Flynn Effect and many papers have tried to explain this in relation to: the general environment (e.g., (469)), specific environmental aspects such as education (e.g., (470)) or nutrition (e.g., (306)) or to biological factors (e.g., (471)) and the idea of 'heterosis' whereby genetically unrelated parents have children with IQs that are slightly higher than those born to the general population (472). This has been critiqued recently as not the leading cause as developmental status has also been discussed as important, such as demographic mobility, family structure, education and health (473).

A reversal of the Flynn Effect is now being measured. Teasdale and Owen (474) found a reversal of the Flynn Effect in their study based on primarily Danish men. It was found that between time points 1988-1998 IQ increased by roughly 1.5 IQ points, but from 1998-2003/4 IQ points dropped by 1.5 points. Similarly, a decline was also reported in Finland; 25,000 males were assessed and from 1988-97 an average increase of 4 IQ points per decade was found, but from 1997-2009 a decline of 2 IQ points per decade was recorded (475). In France, 79 participants were measured using adult Wechsler scales and a drop of 3.8 points was identified between 1999 and 2008-9 (476). Shayer (477) found in the UK a decline in IQ amongst 11-12 year olds of 12 IQ points from 1975-2003, which represents an average of 4.3 IQ points per decade. These four studies could indicate that across Western Europe, the Flynn Effect is being called into question.

Testing the stability of childhood IQ measurements is important for the validity for the test scores (478), and also the consistency of the scores over time (479). Long-term stability of IQ scores are critical as 'high-stake' decisions are based upon the results, i.e. long term SEN placements (480), therefore it is essential that the initial placement decisions are based upon stable IQ results, as an incorrect SEN placement could be harmful to some (481). Intelligence is critiqued as a stable construct (482, 483), but of these studies few have measured IQ in infancy.

However, some argue that the 'g' factor is not identifiable in young children/infants (465, 484), due to an increase in stability of intelligence as individuals make the transition from childhood to adolescence (485). In one study children aged 2-5, 5-8 and 9-12 years were compared with an increasing correlation strength identified ($r$ = .32, .70, .85 respectively) (486). Other early studies also identified little to no associations between IQ measured in infancy to those measured in early-middle childhood (487). Yang et al. (488) assessed 313 pre-school children twice with delayed cognitive profiles (FSIQ < 85) with a follow-up mean of 38.6 months. A mean difference of 7.4 points from time point one to two was found which was significant ($r$ = .43-.5). However, this was only a moderate correlation and participants weren't all administered the same intelligence measure (a total of three were used to collect the data).

Other studies have found a good stability in childhood to adulthood IQ measurements. IQs measured at age 11 have been associated to those at age 77 in a sample of 101 ($r$ = .63) (489), and up to age 90 ($r$ = .54) (490). Mortensen et al. (491) assessed slightly younger children at age 9.5 (n = 26) with the WISC-III and again at age 23.5 with the WAIS and identified stability estimates of $r$ = .86, .86 and .89 (for verbal, performance and FSIQs respectively). Testing at age 6 compared to age 11 in a larger sample (n = 717) and also with a Wechsler scale, found very strong correlations ($r$ = .85) (492). Testing at similar ages with the WISC-II (test interval 2.83 years, n = 642) also yielded good stability (493). Finally, Watkins and Smith (480) analysed within this age group with the WISC-IV identifying coefficients ranging from $r$ = .65-.82 (FSIQ was the highest). However, 25% has FSIQs varying by ≥ 10 points and the VCIQ, PRIQ, WMIQ and PSIQ also varied (29%, 39%, 37% and 44% respectively, mixture of ± mean scores). It could be concluded that WISC-IV scores are not consistent with long test-retest intervals. This latter study highlighted how investigating the means may not provide the 'full picture' of IQ stability.

The causes of IQ change in childhood (beyond unreliability) remain unclear (492). Cherney et al. (494) have proposed age appropriate explanations as to why IQ during young childhood might not be stable. Firstly, that the transition from infancy to early childhood is where cognitive ability rapidly changes as the individual's language develops. Secondly, the period covering early to middle childhood (around age 7) is where an individual develops concrete operational thoughts coupled with an increase in formal academic tasks undertaken.

### 2.2.2.3. Comparative Study

As part of the validation of WISC-IV, it was compared against scores on the WPPSI-III ((78), pp. 64). Both assessment measures were administered to 182 children aged 6-7 years with a testing interval of 9-62 days (mean of 22 days). It was found that the mean FSIQ had a small difference of 0.2 points: the WISC-IV scores were higher. The verbal IQ from the WPPSI-III and the VCIQ from the WISC-IV had a significant positive strong association ($r$ = .76). The performance IQ from the WPPSI-III and the PRIQ from the WISC-IV had a similar association, ($r$ = .74) and the FSIQs were very strongly associated ($r$ = .85). The difference between the tests was non-significant: ($F$[1, 163] = 1.0, $p$ = .32) based on a 2 (test: WISC-IV, WPPSI-III) X 3 (scale: VCIQ-VCIQ, PRIQ-performance IQ, FSIQ-FSIQ) X 2 (test order) MANOVA; and ($F$[1, 169] = .25, $p$ = .62) based on a 2 (test: WISC-IV, WPPSI-III) X 3 (scale: VCIQ-VCIQ, PRIQ-performance IQ, PSIQ-PSIQ) X 2 (test order) MANOVA (78). It was also identified that the two norms from the batteries were highly consistent with one another.

### 2.2.3. Aims/hypothesis

The aim of the current analysis was to explore how similar and associated the scores from the WPPSI-III and the WISC-IV were from the CATS cohort. Based on the literature that cognitive measures administered to younger children may not be good predictors of later outcomes (465, 484-488), the chapter aim and hypotheses were;

i.    There would be a weak-moderate association between the IQs (VCIQ, PRIQ and FSIQ): based on (480, 486, 488, 490)
ii.   Age 3 IQ would be significantly different to age 9 IQ: based on (465, 484-488)

### 2.2.4. Method and statistical analysis

The data used in the current chapter was taken from the CATS I and CATS II IQ tests. The collection method of the data can be found on pages 30-32 for the age 3 IQs (WPPSI-III), and pages 53-56 and 62-64 for the IQ data I collected at around age 9 of offspring (WISC-IV). All of the data was collected using Wechsler scales, and the WPPSI-III and the WISC-IV had similar structures. The FSIQs and verbal IQ (WPPSI-III) and VCIQ (WISC-IV) were similar, the WPPSI-III also generated a performance IQ, and for the analysis here this was compared to

the WISC-IV's PRIQ. For the rest of this chapter, the verbal domain IQs from the WISC-IV and WPPSI-III will be referred to as VCIQ, and similarly, the performance domain IQs will be referred to as PRIQ.

The analysis for the current chapter was completed after the data was collected for the CATS II project; this was to avoid any potential bias by unblinding the treated and untreated SGTF groups. The WPPSI-III and WISC-IV data had been cleaned previously (pages 32-35 and 64-71, respectively) and checks for normality conducted alongside. Whilst no differences were identified between the cognitive outcomes for the treated and untreated SGTF groups, the groupings were included as the between-subjects factor.

The data was analysed in IBM SPSS Statistics Version 20. In total, there were 196 participants (114 treated SGTF, 82 untreated SGTF). The analysis was two-fold to explore the chapter's hypotheses.

The first analysis was to explore the first hypothesis (i) by Pearson correlations, and adjusted partial correlations were executed to explore the associations whilst taking account of the following covariates: corrected for child gender, mother age at time of consent into CATS I, whether the mother breastfed over one month and social deprivation. These four covariates were selected for adjustments out of the possible six CATS II confounders as 'place of assessment' would not have been relative to the WPPSI-III assessments as all were conducted at home. Similarly, 'language at school and home' at age 3 was not collected (see page 59 for further information on the covariates used in this thesis which were the same as used in CATS II (280)). The correlations were carried out separately for both of the SGTF groups.

The second analysis was to explore the significant results from the correlations, and also the second hypotheses (ii). The unadjusted model was analysed firstly by a repeated measures MANOVA; the within-subjects factor was the IQ test (i.e. WPPSI-III and WISC-IV), with measures being VCIQ, PRIQ and FSIQ. The main analysis reported in this chapter was the findings from the repeated measures MANCOVA; adjusted by the above four covariates used in the partial correlations.

### 2.2.5. Results

Descriptive statistics for the groups can be found in Table 40; output from the repeated measures MANCOVA. Note that total number has dropped (from 196 to 194) due to adjustments and missing data.

Table 40

*Means and Standard Deviations of Intelligent Quotient (IQ) Scores for the Treated and Untreated Suboptimal Gestational Thyroid Function (SGTF) Groups from Wave One and Wave Two*

|  | Participant Group | Mean | N |
|---|---|---|---|
| WPPSI-III VCIQ | Treated SGTF | 109.91 (11.37) | 113 |
|  | Untreated SGTF | 110.10 (11.47) | 81 |
|  | Total | 109.99 (11.38) | 194 |
| WISC-IV VCIQ | Treated SGTF | 97.74 (10.02) | 113 |
|  | Untreated SGTF | 100.46 (13.24) | 81 |
|  | Total | 98.88 (11.52) | 194 |
| WPPSI-III PRIQ | Treated SGTF | 108.96 (13.64) | 113 |
|  | Untreated SGTF | 109.68 (13.16) | 81 |
|  | Total | 109.26 (13.41) | 194 |
| WISC-IV PRIQ | Treated SGTF | 104.56 (12.28) | 113 |
|  | Untreated SGTF | 106.40 (13.29) | 81 |
|  | Total | 105.32 (12.71) | 194 |
| WPPSI-III FSIQ | Treated SGTF | 110.84 (12.09) | 113 |
|  | Untreated SGTF | 111.38 (11.47) | 81 |
|  | Total | 111.07 (11.80) | 194 |
| WISC-IV FSIQ | Treated SGTF | 101.90 (12.02) | 113 |
|  | Untreated SGTF | 103.74 (13.07) | 81 |
|  | Total | 102.67 (12.47) | 194 |

Note. Standard deviations appear in parentheses below means. WPPSI-III=Wechsler preschool and primary scale of intelligence-third edition, UK, WISC-IV=Wechsler intelligence scale for children-fourth edition, UK, VCIQ=verbal comprehension intelligent quotient, PRIQ=perceptual reasoning intelligence quotient, FSIQ=full scale intelligent quotient, SGTF=suboptimal gestational thyroid function.

The initial analysis explored the unadjusted (n = 114) and adjusted (n = 113, with 107 degrees of freedom) associations between the treated SGTF group for the VC, PR and FS IQs (see Table 41). All associations were positive and significant, $r$'s were > .293, with the majority displaying moderate relationships; the FSIQs had the strongest relationships ($r$'s > .635).

Table 41
*Correlations for the Treated Suboptimal Gestational Thyroid Function Group of IQs from wave one (WPPSI) and wave two (WISC)*

|  |  | WISC VCIQ | | WISC PRIQ | | WISC FSIQ | |
|---|---|---|---|---|---|---|---|
|  |  | Unad.* | Ad.** | Unad. | Ad. | Unad. | Ad. |
| WPPSI VCIQ | Correlation | .575** | .546 | .424** | .381 | .585** | .543 |
|  | Sig. | .000 | .000 | .000 | .000 | .000 | .000 |
| WPPSI PR IQ | Correlation | .326** | .293 | .561** | .550 | .552** | .518 |
|  | Sig. | .000 | .002 | .000 | .000 | .000 | .000 |
| WPPSI FSIQ | Correlation | .530** | .497 | .588** | .562 | .676** | .635 |
|  | Sig. | .000 | .000 | .000 | .000 | .000 | .000 |

Note. *Unad. = unadjusted model, Pearson correlation. **Ad. = adjusted model, partial correlation. WPPSI=Wechsler preschool and primary scale of intelligence-third edition, UK, WISC=Wechsler intelligence scale for children-fourth edition, UK, VCIQ=verbal comprehension intelligent quotient, PRIQ=perceptual reasoning intelligence quotient, FSIQ=full scale intelligent quotient, SGTF=suboptimal gestational thyroid function.

The untreated SGTF group's associations were calculated for the IQs measured by the WPPSI-III and the WISC-IV with an unadjusted, Pearson correlation (n = 82) and adjusted partial correlation (n = 81, with 75 degrees of freedom) to take account of the covariates. Table 42 displays the associations, again the strongest correlated was the FSIQ ($r$'s > .595) with all associations being significant and positive.

Table 42

*Correlations for the Untreated Suboptimal Gestational Thyroid Function Group of*
*Intelligent Quotients (IQs) from Wave One (WPPSI) and Wave Two (WISC)*

| | | WISC VCIQ | | WISC PRIQ | | WISC FSIQ | |
|---|---|---|---|---|---|---|---|
| | | Unad.* | Ad.** | Unad. | Ad. | Unad. | Ad. |
| WPPSI | Correlation | .536** | .479 | .396** | .300 | .497** | .392 |
| VCIQ | Sig. | .000 | .000 | .000 | .008 | .000 | .000 |
| WPPSI | Correlation | .445** | .365 | .684** | .633 | .624** | .548 |
| PR IQ | Sig. | .000 | .001 | .000 | .000 | .000 | .000 |
| WPPSI | Correlation | .585** | .522 | .661** | .603 | .678** | .595 |
| FSIQ | Sig. | .000 | .000 | .000 | .000 | .000 | .000 |

Note. *Unad. = unadjusted model, Pearson correlation. **Ad. = adjusted model, partial correlation. WPPSI=Wechsler preschool and primary scale of intelligence-third edition, UK, WISC=Wechsler intelligence scale for children-fourth edition, UK, VCIQ=verbal comprehension intelligent quotient, PRIQ=perceptual reasoning intelligence quotient, FSIQ=full scale intelligent quotient, SGTF=suboptimal gestational thyroid function.

### 2.2.5.2. Repeated measures analysis

For the unadjusted, repeated measures MANOVA, there were no significant differences between the treated and untreated SGTF groups ($p = .616$). IQ tests taken at age 3 and age 9 were significantly different; $\Lambda_{ROY} = 1.024$, $F (3, 192) = 65.510$, $p < .001$, $\eta_p^2 = .506$. Univariate analysis revealed the VCIQ had significantly changed; $F (1, 194) = 188.948$, $p < .001$, $\eta_p^2 = .493$, the PRIQ was significantly different; $F (1, 194) = 21.069$, $p < .001$, $\eta_p^2 = .098$, and also the FSIQ had significantly changed; $F (1, 194) = 132.961$, $p < .001$, $\eta_p^2 = .407$. As can be seen in Table 40, the IQs had all dropped over time. The repeated measures MANOVA also took account of the change between each IQ test for the groups and calculated to see if there was any difference, i.e. a group and IQ test interaction; no significance was identified ($p = .300$).

The adjusted analysis took account of the four covariates. Similar to the unadjusted model, there were no significant differences between the groups ($p = .378$). The IQ tests were again significantly different; $\Lambda_{ROY} = .111$, $F (3, 186) = 6.870$, $p < .001$, $\eta_p^2 = .100$. Univariate analysis revealed the VCIQ had significantly changed; $F (1, 188) = 4.619$, $p = .033$, $\eta_p^2 = .024$, PRIQ was not significantly different between age 3 and age 9 ($p = .288$), but FSIQ was; $F (1, 188) = 13.720$, $p < .001$, $\eta_p^2 = .068$. Figures 20 and 21 below show the drops in mean IQ over the two time points for first the treated and also the untreated SGTF groups. The analysis also explored any adjusted differences between the groups with an interaction for IQ test, results remained non-significant ($p = .224$).
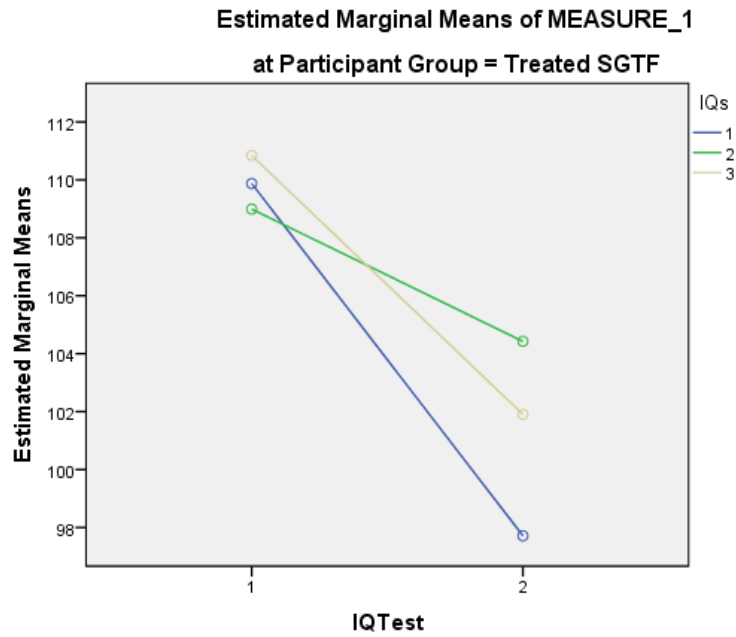
Estimated Marginal Means of MEASURE_1

at Participant Group = Treated SGTF

Covariates appearing in the model are evaluated at the following values: Gender = 1.48, Social Deprivation Quintile (1 low) = 3.71, Was the child breastfed for over 1 month = .60, Mother age at consent into CATS (pregnancy). Quarters, 1 younger = 2.31

Figure 20: Graph of mean drop in intelligent quotient (IQ) over time for the treated Suboptimal Gestational Thyroid Function (SGTF) group

IQs 1 = Verbal Comprehension IQ, 2 = Perceptual Reasoning IQ, 3 = Full Scale IQ. IQ test 1 = Wechsler Preschool and Primary Scale of Intelligence-third edition, UK, 2 = Wechsler Intelligence Scale for Children-fourth edition, UK. X axis=IQ test, Y axis=estimated marginal means of the IQ domains per IQ test.

**Estimated Marginal Means of MEASURE_1**

**at Participant Group = Untreated SGTF**

Covariates appearing in the model are evaluated at the following values: Gender = 1.48, Social Deprivation Quintile (1 low) = 3.71, Was the child breastfed for over 1 month = .60, Mother age at consent into CATS (pregnancy). Quarters, 1 younger = 2.31
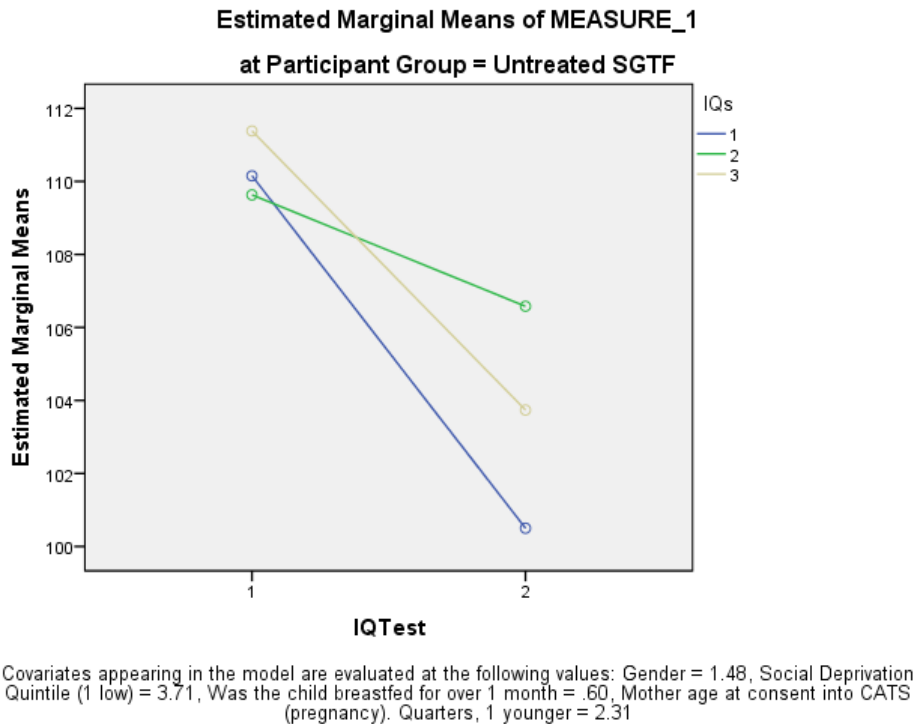
Figure 21: Graph of mean drop in intelligent quotient (IQ) over time for the untreated Suboptimal Gestational Thyroid Function (SGTF) group

IQs 1 = Verbal Comprehension IQ, 2 = Perceptual Reasoning IQ, 3 = Full Scale IQ. IQ test 1 = Wechsler Preschool and Primary Scale of Intelligence-third edition, UK, 2 = Wechsler Intelligence Scale for Children-fourth edition, UK. X axis=IQ test, Y axis=estimated marginal means of the IQ domains per IQ test.

### 2.2.6.    Discussion

The aim of the current chapter was to explore the associations between ages 3 (WPPSI-III) and 9 (WISC-IV) IQ scores from those in the treated and untreated SGTF groups from the CATS sample. The first hypothesis was to explore whether there would be a weak-moderate association between the IQs (based on previous literature (480, 486, 488, 490)). This was confirmed by the partial correlations, between the groups they were all positive and significant ($p$'s < .008) and were weak to strong associations ($r$'s = .293-.684). The second hypothesis was that the age 3 IQ would be significantly different to the age 9 IQ (based on the literature of younger age IQs not being stable (465, 484-488)). The main analysis confirmed this and revealed, from an adjusted repeated measures MANCOVA, that the VCIQ ($p$ = .033) and FSIQs ($p$ < .001) were significantly different between the two time points.

The correlations identified in the results section were all positive relationships. The analysis in this chapter needed to extend beyond correlation calculations as they only informed the

direction of the IQs, i.e. if a child had a lower age 3 IQ they would also have a lower age 9 IQ. This trend did not indicate how the statistics had changed overtime. It was clear from the means in Table 40 that the WPPSI-III scores were much higher than the WISC-IV, which was also why a repeated measured MANCOVA was implemented. The weakest associations for the treated SGTF group were between the WPPSI-III PRIQ and WISC-IV VCIQ, for the untreated SGTF group it was between the WPPSI-III VCIQ and WISC-IV PRIQ. A reason for this may have been because these were separate domains and thus measuring two types of cognitive functioning. The strongest relationship for the treated SGTF group was between the FSIQs, as a representation of g; this being the most stable over the 6 year period it contradicted other findings that found no association over childhood (487). As can be seen by the treated (Table 41) and untreated (Table 42) SGTF group correlations, the covariates had an effect on all of the interactions and reduced the strength of all the associations.

The comparative study in the WISC-IV manual ((78), pp. 64) found positive strong and very strong significant associations, these were more highly correlated than the current results. A reason for this could be due to the fact the WISC-IV was administered prior to its publication, and in CATS II, the WISC-IV was administered around ten years post publication (i.e. a decade post standardisation). However, the means for the WISC-IV were around the proposed population mean of 100, and subsequently were not subject to the Flynn Effect, which proposed a gain of around three IQ points (217-223). It appeared to be the WPPSI-III means that were much higher than the population. The WPPSI-III was published in 2003 and was used in CATS I during 2006-2010, so the Flynn Effect could explain the high means of 106-109 (page 41). Further investigation of calculated means in the sample being higher than those measured earlier in the study would be an interesting reflection. The reversal of the Flynn Effect (474-477) may be having an effect on the IQs, as it was proposed that more recently, Western countries previously measured increases in IQ (due to the Flynn effect), were now slowly decreasing. This phenomenon has not been widely reported yet, so it is difficult to generalise the hypothesis here, especially when considering the sample in the current analysis were not from the normal GTF group.

As well as the correlations from the study conducted during the validation testing of the WISC-IV providing a baseline comparison ((78), pp. 64), the significance of the difference between the scores on the tests were also calculated, revealing a non-significant difference which contradicted the main findings here, apart from the PRIQ domain. The WISC-IV findings (chapter 1.4., Intelligence measured at age 9; CATS II data) previously reported this higher PRIQ compared to the other domains of VC, WM and PS, which have been confirmed

by other literature (222, 297-300). The graphs for the treated (Figure 20) and untreated (Figure 21) display how the PRIQs dropped over time, as with the VC and FSIQs, but this decrease was nowhere near as substantial. Comparative study findings found that the WPPSI-III and WISC-IV were extremely similar in results (FSIQ results were 0.2 points different with the WISC-IV scoring higher), the current analysis revealed a mean difference of 8.4 points, but as mentioned, for the treated SGTF group the FSIQs held the strongest associations.

As previously reported (200), the CATS I study was the first in the world to explore the effect of treatment on SGTF in an RCT and now also, being the first to explore the effects of SGTF whilst following up the offspring. It is difficult to compare the findings as no such similar studies have been planned in the UK context at this time. The findings from the repeated measures MANCOVA were not significant when calculated for any differences between the treated and untreated SGTF groups ($p$ = .378), and also when an interaction was included for the two IQ tests ($p$ = .224). These results were anticipated as means between the groups were similar at age 3 (see chapter 1.2., Re-analysis of intelligence at age 3; UK CATS I cohort) and age 9 (see chapter 1.4., Intelligence measured at age 9; CATS II data). The latter result suggests that both groups' IQs changed equally over time.

As part of data cleaning of the WISC-IV data in chapter 1.4. (Intelligence measured at age 9; CATS II data), an outlier was identified and subsequently removed (see pages 64-71), as the individual scored < -3 for z-scores on 4 out of 5 IQs. Upon unblinding, it came to light, that this participant was from one of the SGTF groups and had taken part in IQ testing at age 3. The participant was excluded from the analysis in the current chapter, due to the WISC-IV IQ results being removed, but her data was included in chapter 1.2. (Re-analysis of intelligence at age 3; UK CATS I cohort) as her scores obtained were not flagged as being outliers. As reported, the WPPSI-III mean scores were on the whole, significantly higher than the WISC-IV scores. The removed CATS II participant at age 3 scored 116 for verbal IQ, 102 for performance IQ and 110 for her FSIQ. The participant's scores for the WISC-IV were 87 for VCIQ (dropped 29 points), 59 for PRIQ (dropped 43 points), (59 for WMIQ, 50 for PSIQ) and a FSIQ of 58 (dropped 52 points). Points to consider alongside these reductions are, when the WISC-IV assessment had taken place the child may not have wanted to comply with the testing which may have resulted in poorer results (284), as well as motivational factors (227-229), and the child's anxiety levels (287) (more information can be found on pages 79-81). It would be difficult to summarise that this child may have 'missed' a diagnosis of having a learning difficulty at a young age, if the WPPSI-III results found during CATS I were presented

to the school, this stability of the earlier IQ results may be weak and the importance of this in relation to SEN placement has been previously discussed (480). However, without a third time point it would be challenging to confirm whether the age 3 IQ was not stable compared to the age 9 data.

As the CATS IQs dramatically reduced over the follow-up period, another possible explanation for this could be the phenomenon of "regression to the mean"; was the regression a true reflection of IQ drops over-time, or was it a product of the reliability of the measuring instrument? An early discussion for this phenomenon was that this result could be a product of measurement error (495). There was also evidence to suggest that individuals with IQ scores in the 'outer ranges' of the bell curve (for example, IQs > 130) were more 'susceptible' to this phenomena (496). Veiel and Koopman (497) identified, in a premorbid IQ context, that there was a tendency for an estimate of IQ to be low (as measured on intelligence tests) if an individual's true IQ was above the population mean, and the effect would be inversed for true IQs below the population mean. This would suggest, that the age 3 to age 9 IQs in the CATS cohort may be a biased estimate of change of the children's IQs.

### 2.2.6.1.    Limitations

The current analysis included a number of limitations. Firstly, performance IQ from the WPPSI-III was compared to the PRIQ on the WISC-IV. The WPPSI-III score may have been a broader representation of this domain to encompass as much data as possible in a short testing time. On reflection, these two domains may not have been good comparisons of one another, although were identified as the best match for this analysis.

As only the SGTF groups were included in CATS I, only these participants could be included in the current analysis. It was unfortunate that the normal GTF were not included in CATS I, as this group of participants would have been a better representation of the wider population. Overall a third IQ measurement may have generated a better interpretation of IQ stability during childhood.

The comparative study published in the WISC-IV (78) comparing it to the WPPSI-III was useful as the identical versions were used in the current analysis, although CATS I and II adopted the UK editions. Sample size was similar (the current analysis contained 14 more participants), but it could be argued we did not have a 'typically developing' sample. Furthermore, at time of the WISC-IV data collection, the tool was around a decade out of date; also the WPPSI-III will have been out of date for assessments during the latter parts of the CATS I data collection. When the comparison study was conducted, the WPPSI-III had just

been published (2003) and the WISC-IV was also new, which could account for the similar reported IQs.

### 2.2.6.2. Conclusions

This chapter aimed to explore how similar and associated the scores from the WPPSI-III and the WISC-IV would be for the treated and untreated SGTF groups. It was found that all scores were positivity correlated, and the difference between the PRIQs was not significant (when adjusted) between age 3 and age 9. However, the VC and FSIQs were significantly different between the two time points. As all of the IQs measured at age 3 were high, it could be concluded that IQ measurement around this time (i.e. late infancy to very early childhood), does not yield stable results. It was possible that the WPPSI-III may have been subject to the Flynn Effect, whereas the WISC-IV may have succumbed to the new phenomena of the Reversed Flynn Effect. Also, the phenomenon of "regression to the mean" may have contributed to the drop in IQs. If future studies within the CATS cohort were to be conducted, a third IQ test may be beneficial to explore whether the age 3 or age 9 results persist as significantly different.

### 2.2.7. Chapter Summary

This chapter has discussed the findings from the exploratory analysis into the comparison of the age 3 and age 9 IQs of those children from the treated and untreated SGTF groups. All IQs were significantly positively correlated with associations ranging from weak to strong. Following adjustments for the repeated measures MANCOVA, the VC and FSIQs were significantly different between age 3 and 9; PRIQ was not. No differences were identified between the two groups, inclusive of group with an interaction of IQ test. The chapter covered a literature detailing how intelligence has been evidenced to change over time, how stability of IQ tests in very young children may not be best practice, and also a brief review of a comparative study. The research aim followed along with two hypotheses which were explored in the results section. A discussion with links to the literature closed the chapter. The following section is the main discussion for this thesis

# 3. General Discussion

### 3.1.     Overview

This final section of the thesis summarises all the results from the previous eight chapters. Overviews of the findings in the first two sections are considered in the context of the literature. Implications of the findings are discussed as well as the main limitations of CATS I and II. Two possible future studies are discussed with this section closing on final conclusions.

### 3.2.     Main findings

This section covers the main findings from all the analyses conducted in the thesis. The aims and hypothesis are covered for each chapter, as well as the primary and secondary analyses (where relevant). Links are drawn across the findings where applicable, as well as to the CATS II analysis.

#### 3.2.1.   Section 1 Overview

##### 3.2.1.1.     Summary of, 'Re-analysis of intelligence at age 3; UK CATS I cohort'

This first results chapter presented the analysis I conducted of the UK data from CATS I, excluding the Italian sample, an analysis that had not previously been done or presented (200). This also acted as the pilot to the analysis I planned to conduct to explore the data I had collected for CATS II. Based on the previous published results, it was hypothesised that there would be no significant differences between the IQs at age 3 for those from the treated and untreated SGTF groups (n = 607).

As predicted, no difference was identified for mean IQs in the adjusted model ($p$ = .688). The exploration of the continuous outcome was different to the CATS I publication by sample, and it also controlled for variables which may have had an impact on the scores; child gender, mother's age at time of consent into CATS I, and a measure of the participant's social deprivation. These were not adjusted in CATS I as the study was an RCT and therefore should have been free from bias. The data I had collected for CATS II was adjusted for covariates, therefore making my own CATS I analysis similar. The results that treatment for SGTF made no difference to childhood IQ may have been a first hint that being born to a mother with SGTF had no effect on offspring intelligence, if being born to a mother with euthyroid function was the same for offspring intelligence as being born to a mother with SGTF, why would treatment for SGTF have an impact? This theory of SGTF not having an effect to offspring intelligence has been found in Henrichs et al.'s (53) epidemiology study, as well as a number of other recent studies (91-93, 95, 96). Also, the finding of treatment having no effect on offspring IQ was previously identified in Haddow et al.'s work (54); a study which included a small number of women who received treatment (n = 14).

The secondary analysis of a binary outcome was conducted so that the work in this thesis would be comparable to the published CATS I study (200); odds of IQ ≤ 85 (1 SD below the mean of 100). This also included an unadjusted (chi-square) and adjusted model (multinomial logistic regression); which again developed the analysis forwards from the original study by controlling for variables. It was found that those from the untreated SGTF group were 3.335 times more likely to score ≤ 85 for FSIQ compared to those from the treated SGTF group. This finding was suggestive that treatment, did in-fact, have some benefit to offspring and appeared to improve IQ. Caution was advised however, as analysing for binary outcomes has been described as not being as valid as continuous analysis (i.e. the primary analysis from this chapter) (211-213).

The CATS I paper (200) also described how the analysis was conducted on an 'intention to treat' analysis; i.e. those who were originally randomised to the treatment branch of the study were kept in the treated SGTF group, even if they were suspected of not complying with the levothyroxine medication. For the CATS II study the analysis was executed with an 'on-treatment' plan whereby those that violated the treatment protocol were subsequently allocated for analysis purposes to the untreated SGTF group (n = 2); this was based on CATS I treated SGTF pregnancy data when the mothers were followed up twice after initial consent.

### 3.2.1.2.    Summary of, 'Intelligence measured at age 9; CATS II data'

This chapter, 1.4., was the first analysis of the cognitive data I had collected for CATS II. This chapter would be comparable to CATS II, as described in Hales et al. (280), IQ was the primary outcome. The aims were to explore whether there would be any difference between the untreated SGTF group and the normal GTF group, based on previous studies it was hypothesised that the normal GTF group would perform better (34, 49-57, 90, 91). The second aim was to explore the IQ of the treated and untreated SGTF groups at age 9, hypothesising that there would be no difference based on chapter 1.2. (Re-analysis of intelligence at age 3; UK CATS I cohort) findings and CATS I results (200). Four hundred and fifty two participants' IQ data were used in the analysis.

The primary analysis for this chapter was comparing the three groups of participants (normal GTF, treated and untreated SGTF) in a MANCOVA. The finding at the multivariate level was that there were no significant differences of VCIQ, PRIQ, WMIQ, PSIQ or FSIQ between the groups ($p = .353$). Within the sample of CATS II, it can thus be inferred that SGTF did not have an impact on childhood intelligence at age 9; which was supported in some studies (53, 56,

91-93, 95, 96). As there was no difference between the untreated SGTF and the normal GTF group, it could be inferred that treatment of SGTF would not yield any difference, as SGTF was reported as not being detrimental for the offspring IQ at age 9. This chapter was also the first to introduce the four models of analysis (for continuous outcomes); so that the findings from this thesis could be comparable to the cognitive CATS II analysis (see study protocol where the models were first proposed (280)).

The secondary analysis, similar to chapter 1.2. (Re-analysis of intelligence at age 3; UK CATS I cohort), was to explore possible differences between the groups based on binary outcomes. This was deemed important to execute in the thesis, as the primary outcome for CATS II was odds of IQ ≤ 85; therefore binary calculations made this work comparable to CATS II findings. I chose continuous variables as my primary outcome, since the validity of binary outcomes have been questioned in the literature. Binary outcomes could underestimate the variability in a sample and indicate that characteristics are different when in fact they could be similar (211-213). The exploratory chi-squares and regressions revealed no significant differences between the groups. As a binary difference was identified at age 3 for FSIQ between the treated and untreated SGTF groups, and if we accept the regression result as not being a type one error (see Austin and Brunner (213) for further details), it could be inferred that any differences from treatment may only be apparent in very early childhood.

### 3.2.1.3.    Summary of, 'Additional cognitive assessments at age 9; CATS II data'

As well as IQ data, I also collected additional cognitive data from the CATS II participants. As with CATS II, the NEPSY-II assessments were treated as secondary to the main outcome of childhood IQ (280). The aims were to explore whether the non-significant differences for IQs would extend to the additional cognitive tests; based on previous research it was hypothesised that I would fail to reject this null hypothesis (34, 51, 53, 54, 56, 91-96, 111, 113, 114). The second aim was to explore whether there would be any differences between the SGTF groups to the normal GTF group, hypothesising that specifically the untreated SGTF would perform less well than the normal GTF group (supported by (10, 32, 49-51, 53-58, 103-106, 110, 112)). Four hundred and sixteen participants were assessed using the NEPSY-II following their WISC-IV assessment.

The primary analysis from this chapter was to explore the differences between the groups by an ANCOVA and MANCOVA. As previously discussed (page 91), LM was analysed separately as a reduced number of participants completed the subtest. The LM analysis revealed a non-significant difference between the groups ($p$ = .613) as was the case with the

other subtests that were analysed together in a MANCOVA ($p$ = .212). As this persistent lack of difference was identified between the SGTF and normal GTF groups, it could be inferred that an underactive thyroid during pregnancy does not affect the child's cognition at age 9; which Haddow et al. (54), Willougby et al. (103-105) and Saurez-Rogrigues et al. (51) found for similar aged children. With the small CATS II groups, it would have been difficult to have the power to identify a mean difference between the groups. As can be seen in Table 21, all of the NEPSY-II subtest means were higher in the normal GTF group (similar to the WISC-IV findings), which warrants the continued suggestion that with larger groups this difference may have been more apparent, or even achieved significance.

The secondary analysis was assessing odds for binary cut offs. The NEPSY-II data generated scaled scores, but similarly the cut-off of ≤ 1 SD was adopted, therefore it was the odds of scoring ≤ 7. It was identified that the treated SGTF group were 2.257 times more likely to score ≤ 7 compared to the untreated SGTF. Similarly, the normal GTF were 1.899 times less likely to score below the threshold compared to the treated SGTF group; we can infer from these findings that the treated SGTF performed significantly worse than both groups. Conversely, this finding was not evident on any other of the WM tests from the NEPSY-II, nor was it identified on the WM domain of the WISC-IV. The exploratory chapter 2.4. (Memory score comparisons; data from the WISC-IV and NEPSY-II assessments) investigated associations between the WISC-IV and NEPSY-II, and found that for the MD subtest, it was only associated to letter-number sequencing. This evidenced how MD was not entirely similar to the WISC-IV WM subtests, which may suggest a type one error. This second analysis highlighted the challenges from binary outcomes (211-213), and was why the primary analyses focused on continuous outcomes in this thesis.

### 3.2.1.4. Summary of, 'Behavioural questionnaires at age 9; CATS II data'

The final data collection I completed for CATS II was the child behavioural questionnaires completed by the mothers. These included the SDQ (assessed general behaviour across five domains), the Child ADHD Questionnaire (assessed three different aspects of ADHD-typical behaviours) and the SCQ (assessed for ASC traits). The aim of the research was to quantify differences between the three groups of participants that might extend beyond cognition. The hypothesis, that there would be no differences between the treated and untreated SGTF groups, were based upon the chapter 1.4. (Intelligence measured at age 9; CATS II data) and chapter 1.5. (Additional cognitive assessments at age 9; CATS II data) findings. It was also predicted that the untreated SGTF group would have more behavioural problems as rated by the questionnaires compared to those from the normal GTF group; as identified in other

studies (19, 44, 45, 143, 148-151, 170, 175-180). Four hundred and seventy one mothers in CATS II completed the questionnaires and were included in the analysis.

The primary analysis was investigating for differences at the continuous level; MANCOVA. Similar to the previous models, a three-staged analysis was executed so that findings would be comparable to CATS II results. At the multivariate level a difference was detected ($p$ = .006), and from post hoc analysis that was Bonferroni corrected, it was found that the treated SGTF group had significantly higher scores on the SCQ (indicating more behaviours indicative of ASCs) compared to the untreated SGTF group ($p$ = .047). It was inferred from this finding, that within the CATS cohort, treatment of SGTF may increase ASC behaviours as it could be seen in the graph of the SCQ means (Figure 16) that the treated SGTF group had, undoubtedly, much higher scores compared to the other two groups. As was previously discussed (page 120), it may not have been the levothyroxine treatment that incurred the higher rating on the SCQ, but some form of psychological effect from taking medication during their pregnancies. However, there is new evidence that over-treatment and consequently higher levels of T4 during pregnancy could have detrimental effects for the offspring (37).

The secondary analysis was centred on binary outcomes. The SCQ finding was replicated during this analysis with the normal GTF being 4.132 times less likely to score above the threshold ($\geq$ 15 [SCQ published cut-off to warrant further investigating for a diagnosis of ASCs (233)]) compared to the treated SGTF group. It was interesting that the comparison was not between the treated and untreated SGTF groups similar to the continuous outcomes analysis, but this added validity to the idea that binary outcomes may increase the risk of a false positive result (213). Following adjustments for the multinomial logistic regressions, two more significant differences were identified; the normal GTF group were less likely to score $\geq$ 1 SD (indicating more behavioural difficulties) compared to the treated SGTF group for Overactivity and Impulsivity (2.027 OR and 2.060 OR, respectively). These findings were not replicated from the primary analysis, but suggest that treatment for SGTF incurred higher scoring for offspring (indicating less desirable behaviour) on the questionnaires. Offspring cognition at age 9 and the behaviour questionnaires could be a future analysis to explore within this cohort.

Additional analyses (appendix 10) identified that there was an over-treatment of levothyroxine in the treated SGTF group; this was evident from the mean T4 scores compared to current 'normal reference ranges'. Mothers with high T4 values (identified as T4 > 97.5[th]

174

percentile of the complete UK CATS I cohort = 17.7 pmol/L) were found to have offspring with higher mean scores and scores > 2 SDs compared to the rest of the CATS II cohort and the normal GTF sub-group (respectively). This finding identified a need for clinicians to closely monitor levothyroxine dosage throughout pregnancy.

### 3.2.2. Section 2 Overview

#### 3.2.2.1. Summary of, 'Significant effects from the covariates; CATS II data'

This chapter was the first exploratory analysis conducted on the CATS II data. The aims were to explore whether any of the six covariates (child gender, whether the mother breastfed over one month, the age of the mother at time of consent into CATS I, where the child was assessed, language at the child's school and home and social deprivation) had significant effects on the dependent variables of CATS II; WISC-IV, NEPSY-II and child behaviour questionnaires. The hypotheses were based on the literature discussed in the chapter and the following paragraphs will discuss the findings in light of the data collection tools. The number of 'groups' for each covariate (for example, gender had two- male or female) dictated whether the exploratory analysis was to be a t-test or MANOVA. The initial significant effects of covariates were identified from the final model of analyses for the WISC-IV, NEPSY-II and questionnaires. The participant sizes were identical to those found in the respective results chapters for the tools; for the WISC-IV see chapter 1.4. (Intelligence measured at age 9; CATS II data), NEPSY-II see chapter 1.5. (Additional cognitive assessments at age 9; CATS II data) and for the behavioural questionnaires see chapter 1.6. (Behavioural questionnaires at age 9; CATS II data). Participants were kept in one group for the analyses in this chapter.

Further details of the hypotheses for the WISC-IV in respect of the covariates can be found in Table 31. The alternate hypothesis proposed by Goldbeck et al. (333) that males would outperform females for VCIQ, was rejected following analysis. For PRIQ, WMIQ and FSIQ it was predicted that there would be no differences between males and females (332, 335), whereas for PSIQ females were anticipated to perform significantly better (328, 332). The analysis identified that females performed significantly better than males for WMIQ, FSIQ and PSIQ; yielding the biggest difference. For the effects of breastfeeding, based on the literature it was hypothesised that those who were breastfed over one month would achieve higher IQs compared to those who were not (as widely investigated by: (380-384, 386-390, 402, 403)). The older the mother was at time of pregnancy, it was predicted the offspring IQs would be higher at age 9 (360, 361, 363, 364), and also the higher the social deprivation score

(indicating less social deprivation), the higher the IQs would be (439-442). These three alternate hypotheses were confirmed in the analysis from this chapter. This added validity to the study demonstrating that the sample for CATS II was somewhat representative of the general population. The place of the WISC-IV assessment was anticipated to have a significant effect on the results (216, 230, 250), this hypothesis was rejected based upon the findings ($p$ = .095). The final covariate was for the child's language of school and home, it was hypothesised that VCIQ of bilingual children would be lower compared to those who were monolingual (420-422, 433, 434), this was not identified by the results. Furthermore, the bilingual children would perform better on PRIQ compared to the monolingual children (420, 423-427, 433, 434), which was also not apparent in the results. A possible reason for this finding was that the groups were underpowered to see these effects. As can be seen in table 13 in chapter 1.4. (Intelligence measured at age 9; CATS II data), 77.7-87.1% of children in each of the participant groups attended English speaking schools and also spoke English at home. Therefore the number of monolingual children vastly outweighed the number of bilingual children for this exploratory analysis.

Specific to the NEPSY-II, child gender, child's language at school and home, and social deprivation rating had a significant effect on subtest scores. The gender finding contradicted the proposed null hypothesis (332), again it was identified that females out performed males. The social deprivation rating alternate hypothesis was confirmed; those from more socially deprived backgrounds performed worse compared to those from a less socially deprived background (442, 446-448). The null hypothesis of language having an effect was adopted as no effects were found for NM subtest in the literature (435-437); this was unfounded in the current data. The hypotheses that mother age would have a significant effect on the scores (360, 361, 363, 364), as well as breastfeeding (380-384, 386-390, 402, 403) and where the NESPY-II assessment took place (216, 230, 250), were not confirmed in this chapter's analysis. Further details of the hypotheses that were adopted for the analysis can be found in Table 32.

Finally, the behaviour questionnaires were analysed with respect to the covariates; excluding where the child was assessed and language at school and home, as these were considered not to affect ongoing childhood behaviour. The only null hypothesis devised was for SDQ, except for males scoring higher on peer problems based on the literature (338, 339). For the Child ADHD Questionnaire and SCQ, males were predicted to score higher compared to females (128, 341-345, 347-350). The analysis revealed that males performed worse (with higher scores) on *all* of the questionnaire domains ($p$'s < .023). Social deprivation score was

the only other covariate to have a significant effect on the questionnaires. The findings confirmed the hypothesis that the lower the rating, the higher the scores would be (indicating 'worse' behaviour) (449, 450, 454). It was anticipated that breastfeeding over one month would have an effect on the questionnaires as it was widely reported in the literature (404, 410, 412-414); however this was not confirmed in the CATS II sample. The final covariate of maternal age yielded two hypotheses; for SDQ and Child ADHD Questionnaires, the younger the mother was the higher the mean scores would be (361, 365-367, 370), with the effect reversed for ASC traits (371-376). Both alternate hypotheses were rejected based on the findings. Further details of the hypotheses that were adopted for the analysis can be found in Table 33.

The covariate of gender had the biggest effect on the offspring scores for WISC-IV, NEPSY-II and the child behaviour questionnaires. Upon reflection, the widest literature available was on gender differences, and is something, I predict, that will continue to be reported upon for many decades to come. A number of the hypotheses were confirmed which supported past studies and aided the generalisability of the current research. Conversely, some of the hypotheses were not confirmed. One of the reasons for this may be that the exploratory analyses were on the grouped data set, i.e. the normal GTFs were mixed with the SGTF groups. With around 50% of the sample being in the SGTF groups, it could be argued that these individual's may not necessarily be 'typically developing' and therefore were not comparable to those individuals included in the studies discussed in the Introduction section of the chapter 2.1. (Significant effects from the covariates; CATS II data); although we have no evidence that the SGTF group were not 'typically developing'. However, with the initial identification of the covariates having an effect on the dependent variables being derived from the final model MANCOVAs, splitting the participants into their respective groups may have incurred difficulties associated with multiple testing.

### 3.2.2.2. Summary of, 'IQ comparison between ages 3 and 9; children from the CATS sample'

The aim of this chapter was to explore the associations between the WPPSI-III and WISC-IV IQ assessments that were administered to children in the treated and untreated SGTF groups. Age 3 IQs were again calculated in their 'uncorrected' form, and not by the corrected IQs around 100 that were presented in the CATS I paper (200). Previous literature discussing cognitive measures conducted on very young children concluded that they do not make good predictions for later life outcomes (465, 484, 485). Key studies supporting this literature have found significant drops in IQ overtime (Yang et al. (488)), infancy to middle childhood IQs not

being associated (Anderson (487)) and increasing correlations of IQ scores with age (Honzik et al. (486)). Based on this, it was hypothesised that there would be a weak-moderate positive significant association between the IQs (VCIQ, PRIQ and FSIQ), but that age 3 IQs would be significantly different to those measured at age 9. One hundred and six participants' data was used for this chapter's analysis.

To explore the first hypothesis, Pearson and partial correlations were calculated for both of the groups. The adjusted correlations took account of four of the six covariates, excluding place of assessment and child's language at school and home. The predicted weak-moderate significant associations were based on a number of studies (480, 486, 488, 490). This was confirmed between the groups with all associations being positive and significant ($p$'s < .008), and weak to strong associations reported ($r$'s = .293-.684). It was inferred from these findings that the two Wechsler scales included similar constructs. Furthermore, the associations were apparent despite only those from the SGTF groups being included in the analysis, i.e. it could be argued that these may be 'atypically developing' children.

The second analysis was quantifying the differences between the two IQ scores by repeated measures MANOVA and an adjusted MANCOVA. As mentioned, an alternate hypothesis was anticipated based on previous literature confirming IQ from young children being unstable (465, 484-488). Both models confirmed these findings, reporting that the age 9 IQs had significantly decreased from age 3; PRIQ lost significance following adjustments. VCIQ had decreased by 11.11 points ($p$ = .033) and FSIQ by 8.4 points ($p$ < .001) (see Table 40 for further information on the means). This reported decline in IQs overtime was attributed as possibly being reflective of either the Flynn Effect, or the phenomena of "regression to the mean". The WPPSI-III may have been subject to the Flynn Effect of increasing IQs over time (217-223), whilst the WISC-IV results could be attributed to the new found phenomena of the Reversal Flynn Effect recently found to occur in European countries (474-477). The regression to the mean in the age 9 IQ results may have been a product of bias in the estimate of change, or a measurement error. This latter phenomenon highlights how there may have been a bias with the CATS II sample.

One of the main draw backs of this analysis was that the data available was only for those from the SGTF groups. None of the normal GTF offspring were assessed at age 3, and these participants would have been most relevant when generalising the results. However, no WISC-IV or NEPSY-II differences were found between the three groups at age 9 so it would be unclear how different the results of those from the normal GTF group might have been.

The repeated measures MANCOVA found no differences between the IQ tests of the SGTF groups, which confirmed chapter 1.2. (Re-analysis of intelligence at age 3; UK CATS I cohort) and chapter 1.4. (Intelligence measured at age 9; CATS II data) findings of there being no differences between these two groups.

### 3.2.3. Conclusions

The outcome from CATS II was that there were no cognitive differences between the normal GTF and both SGTF groups. For behaviour however, the MANCOVA identified that the treated SGTF group had significantly higher SCQ means compared to the untreated SGTF group; inferring that the treated SGTF were more likely to present behaviours related to ASCs. This finding could be critiqued as the evidence was based upon a mother-completed questionnaire, but the sustained effect of treatment on behaviour from maternal gestational thyroid function was measured around 9 years post-partum; therefore a long time since the exposure to the treatment for SGTF.

## 3.3. Limitations

### 3.3.1. Challenges from CATS I

In critiquing CATS II and the research in this thesis, challenges presented by wave one of the study also need to be discussed. A recurring theme throughout this thesis was the somewhat broad definition of SGTF; including those with either TSH above the 97.5[th] percentile of the cohort (GSH), or T4 below the 2.5[th] percentile (gestational hypothyroxinaemia), or both (overt hypothyroidism). Nearly all the studies in the 'General introduction' (chapter 1.1.), used different reference ranges to define an underactive thyroid during pregnancy. One such example that was different to CATS, was Klein et al. (90) that identified GSH as having a TSH in the 99.85[th] percentile of the cohort. This more stringent GSH definition could have been why Klein and colleagues identified a difference compared to those with normal thyroid function. For intelligence, the literature has tended to explore either a high TSH or low T4; those that have found an effect (49, 54, 56, 90, 91) and no effect (53, 56) of GSH, and maternal hypothyroxinaemia (evidence of there being an effect to offspring intelligence (34, 49-53) and no effect (91, 92, 94-96)).

The percentile cut-offs were adjusted throughout the recruitment stage of CATS I as more participants joined the study; this was a pragmatic approach based on the population. This led to the untreated SGTF classification being 'more hypothyroid' than the treated SGTF group. This was because those who initiated treatment and were recruited at the beginning of the study will have started their levothyroxine therapy immediately based upon percentile

ranges of a smaller cohort. Within CATS II, this subsequently lead to participants from the normal GTF group being moved into the untreated SGTF group, to help make the percentile ranges more comparable between the groups; 30 participants were moved.

As discussed in the CATS I paper (200), participants were recruited during their first visit to the antenatal clinic in hospital; which generally fell towards the end of the first trimester (median of 13 weeks and three days). Soon after this, the foetus' thyroid has been reported to work independently (5). It could be inferred, that a more 'critical time' of the foetus being dependent on the mothers circulating thyroid hormones, would be during the first trimester; i.e. before treatment for SGTF was initiated in CATS I. Therefore, CATS may have been challenged to measure any benefits of treatment from the outset because treatment was initiated too late.

The treated SGTF group were followed up during pregnancy at six weeks post initial consent and at 30 weeks gestation. The starting dose was 150 µg per day of levothyroxine, and the follow up appointments were to track any fluctuations in the mothers TSH and T4; an indication that dose may have needed to been adjusted. It would have been difficult to categorically conclude that mothers took their medication every day. Within the supporting documents in the CATS I paper, graphs of the mean changes in T4 and TSH were presented to show that change towards more euthyroid function did occur (200). Compliance was explored, as a few participants (two of which continued with CATS into wave two) were suspected of not initiating therapy. CATS I analysed data on an 'intention to treat' model, and an exploratory analysis did remove those who were suspected of non-compliance but no difference was identified (200).

Further to treatment, the levothyroxine drug itself was not given to participants from the study, but a prescription for it. The make of the levothyroxine drug was not investigated in CATS I and this may have been a potential confounder as absorption rates of T4 have been demonstrated to vary between brands (498). Caffeine has also been recently evidenced to affect absorption rates of levothyroxine (499, 500), which was unknown when CATS I began recruitment and data collection.

As previously discussed, an opportunity may have been overlooked by not assessing the normal GTF offspring in wave one. This would have made CATS I more comparable to other studies that could not identify a link between underactive thyroid function during pregnancy and euthyroid function (i.e. the comparison would be between the untreated SGTF and the normal GTF groups) (53, 56, 91-93, 95, 96). The CATS I findings identified no difference for

treatment of SGTF, but if the normal GTF were found to have similar IQs in wave one as in CATS II, it may have been concluded treatment was no benefit to the offspring, but SGTF made no difference to cognition either.

### 3.3.2. Challenges from CATS II

Recruitment for CATS II was a challenge from the beginning of the data collection period. From August 2011 to the end of 2012, recruitment was low at an average of around four participants per month. In 2013 (when the cognitive assessments began) this rate doubled, 2014 saw an average of 12 participants per month, and in the final year of recruitment the rate was around 18 per month. This increase of recruitment could have been due to a number of factors; there were changes within the CATS II recruitment team from August 2014, a Facebook page was developed which was another mode of communication for the participants, the option of home visits were introduced in September 2013, and also the Welsh Demographics Service was latterly introduced which enabled the team to 'prompt' responses from participants after initial invitation to the research by telephone calls. Importantly, the Patient Data Register was used which ensured up-to-date addresses for the participants; this was only adopted for the SGTF groups, utilisation of these services required amendments to ethical approval. If all of these alterations were in place from the start of CATS II, the recruitment target of 480 for cognitive assessments may have been surpassed.

Once participants were booked in for their appointment at the research centre or at their homes, there was the further issue of appointments not being kept. This posed a time issue in respect of the home visits, particularly if the participants lived far from the research centre. The team attempted to overcome this issue by the text reminders, which also offered a simple method of cancelling the appointment. As can be seen in the demographic tables on pages 70, 90 and 108, the untreated SGTF group were the most challenging to revisit; specifically their attendance to the research centre was much lower than the treated SGTF and normal GTF groups. In retrospect, it would have been interesting to have kept track of participants cancelling appointments and explored if there was a group bias involved.

To gain a better perspective on what motivated participants to take part, a follow-up questionnaire could have been given to participants at the end of their appointments/following the return of their postal pack. This could have also generated feedback on how the structure of the study could have been improved, which may have helped recruitment. Also exploring what motivated individuals to participate may have

identified a possible confounder of participants taking part because they may have believed something to be 'wrong' developmentally/behaviourally with their child.

Specific to the cognitive assessments, the environments were not considered ideal, although these were the best options for the study to keep things feasible for participants. For research centre visits, the assessments occurred at the end of the visit. A higher performance may have been scored if the assessments were conducted earlier in the visits as the child would have not have fatigued from proceeding demands i.e. giving a blood or saliva sample, waiting for his/her mother to complete the informed consent and medical history with a clinician or from lying still for the bone density scans. However, as discussed in Table 8, there was a time constraint and the only feasible time to have the assessments was at the end of the visit. This was similar to the difficulties of the home assessments, as achieving the ideal 'distraction free environment' (216) was challenging in someone's home.

As research centre visits occurred during one visit, it was natural that the cognitive assessments were also completed during one visit. In retrospect, perhaps the cognitive assessments could have been completed on separate days to the initial research centre visits (i.e. collecting consent, blood/saliva samples, and bone scans etc.), or the WISC-IV and the NEPSY-II assessments could have been divided by a defined break of, for example, collecting a saliva sample. Fatigue of the child was something I was constantly aware of as the examiner, and was most frequently listed as a reason for not commencing with the additional NEPSY-II tests (in total 36/452 [8%] did not complete any of the NEPSY-II, see page 88 for further details). Placing the NEPSY-II straight after the WISC-IV was the correct decision, as the mean scores for the NEPSY-II were not below the standardised means of ten (230); it could be argued that no 'underperformance' was apparent. A sensitivity analysis of those who did not complete the NEPSY-II compared to those who did, of IQs may have been interesting. Commonly, the child was fatigued due to a long testing time of the WISC-IV, which could have indicated a superior performance on the IQ test (78), it could therefore be hypothesised that the more intelligent children in the CATS II cohort did not complete the NEPSY-II.

As previously discussed on page 80, administering different intelligence tests could produce a range of results (289, 290), thus perhaps by only adopting one it may not be a true representation of that individual's functioning. For the measure of intelligence of the offspring in CATS II, the WISC-IV could have been perhaps used in conjunction with other data. For example, data from the National Pupil Database could have been explored as this

was routine data collected by schools which would have captured the child's *ongoing* performance, behaviour and sociability; this data would have avoided any potential bias of maternal reporting. With the increasing recruitment of CATS II, and time cost of home visits and scoring data, this data would have been difficult for me to gather during the data collection phase of the CATS II study. However, having this educational attainment information may have been useful to explore the correlation to the child's IQ. In retrospect, attainment information could have been collected post hoc by the proposed 'follow-up' questionnaires mentioned above.

Specific to the analysis of the cognitive data, one MANCOVA could have been executed for the WISC-IV and NEPSY-II data instead of two separate analyses. From a statistical standpoint, this may have been preferable as it would have reduced multiple testing. If this thesis was not going to be comparable to the CATS II results, this analysis may have been different; as the planned analysis was for separate comparisons between the WISC-IV and NEPSY-II (280). However, one MANCOVA would have meant smaller participant groups as the LM NEPSY-II subtest was not completed by many.

Throughout all analyses (except for significant effects of the covariates' investigations), adjustments were made to try and control for any extraneous effects. Randomisation from the CATS I study was inevitably compromised, for example some of the normal GTF group were recruited based upon their proximity to the research centre, similarly those from the SGTF groups who had moved further than three hours away were difficult to follow-up in CATS II. The age of the children in the normal GTF group were significantly older than those from the SGTF groups, which was indicative of this broken randomisation; on average they were around 0.235 years older, which is equivalent to just under three months. The social deprivation score as a covariate could have been collated better by the 'General CATS Questionnaire', for example collecting information on parental education, whether their home was bought, income levels etc. (see Appendix 8: Revised CATS general questionnaire, this could not be used unfortunately as it was devised too close to the end of the study).

### 3.4.    Future studies
- CATS 0

To overcome many of the limitations discussed above, a restart on the entire project could be considered. One of the main benefits of this would be that serum samples to test thyroid function would be collected earlier in pregnancy, and also treatment could be initiated earlier. This could occur by recruiting local surgeries into the project so that when women

visit their general practitioner to confirm the pregnancy, samples could be collected here. The SGTF definition could be split into exploring GSH and maternal hypothyroxinaemia, with clear defined reference ranges based on the most up-to-date UK/European guidelines available.

Tracking the mothers' diet and when levothyroxine was taken during the day could be conducted; including caffeine consumption. Also, following the offspring's development and cognition throughout childhood would be interesting, as well as following the development of those born to the normal GTF group. Data from the National Pupil Database could be utilised as well as perhaps using the teacher version of the SDQ to minimise any bias from maternal reporting. CATS 0 could administer questionnaires and cognitive assessments electronically to aid speed of scoring, electronic questionnaire could be designed to 'flag' unanswered questions to the mother to ensure high completion rates.

Clear targets of how many participants need to be assessed per month might aid numbers, also allowing for participants not attending visits would be beneficial. More contact with participants could be established between waves of the study which would help with attrition rates; e.g. news leaflets, birthday cards, information on relevant study publications. Finally, use of social media could also be adopted, as well as a Facebook page for the mothers; one could also be developed for the children in addition to using other social network avenues.

More information about the participants 'general home life' would be beneficial when considering adjustments for analyses. For example, information on diet (specific for the physical aspects of CATS II), information on parental education, as well as the child's educational attainments would be useful. As mentioned, a follow-up questionnaire could be used to help develop and improve the project.

- CATS III; brain imaging

Thyroid hormone action in the developing brain is dependent on an important gene called deiodinase 2 (D2), as well as the mother's thyroid hormone levels (25). D2 is involved in the deiodination of T4 to T3, and is expressed in the brain (and also the pituitary, thyroid gland and skeletal muscle) (6). Rodent studies have shown how exposure to low thyroid hormone in utero may lead to hypo-myelination of white matter tracts (501, 502); but exploration in humans is limited. In the exploratory analysis for CATS II, it was found that children carrying a common variant of the D2 gene (homozygous for Thr92Ala) and who were also born to

mothers with low T4, were at an increased risk of FSIQ ≤ 85 and treatment appeared to alleviate some of these lower IQs.

At time of submission of this thesis, a small grant had been awarded to conduct 12 magnetic resonance imaging scans on the CATS children. These scans will be conducted on six children from the untreated SGTF (i.e. low maternal T4) homozygous for Thr92Ala, compared to six children from the normal GTF without the specific genotype. This small study only planned to scan males to reduce any gender effects, and would act as a pilot for a much larger study. An application for this larger study was also submitted at time of completion of this thesis. This study proposed 80 scans; 40 children born to mothers with low T4, 20 with the homozygous Thr92Ala genotype and 20 without, would be compared to 40 children born to mothers with normal T4 levels, again 20 with the Thr92Ala genotypes and 20 without. Repeat cognitive testing was also planned to ensure contemporaneous capture with the neuroimaging readouts. It was planned that the children would undergo characterisation of white matter microstructure in vivo; which would generate a quantification of their myelination. This potential third wave of the study was planned to occur around age 11 of the children to avoid any potential confounders from puberty.

The research aims of this planned future study are to explore:

i. Whether differences would exist between children exposed to gestational low thyroid hormone or not; if so, whether these would be restricted to the adverse D2 genotype?
ii. Whether any differences in cognition were attributable to differences in white matter microstructure; if so, were they specific to myelin metrics and in which anatomical pathways?
iii. Whether any differences identified would be alleviated by maternal thyroxine supplementation.

### 3.5.    Final conclusions and summary

The work described in this thesis failed to generate evidence for SGTF having a negative impact on offspring's cognition at age 9 years. Treatment of SGTF also had no effect to cognition, but was shown to 'worsen' behaviour as measured by questionnaires completed by the mothers. Whilst the exploratory analyses generated interesting findings of the significant effects of the covariates and how the IQs significantly changed between wave one and two, no further evidence was identified between the three main groups in the study.

CATS was the first RCT in the world to explore the effect of treatment in a large cohort, further large studies exploring underactive thyroid function during pregnancy are needed to determine whether universal screening and treatment should be sought during pregnancy

or not. With the finding from the behavioural questionnaires, additional exploration is needed to better quantify whether treatment for SGTF could be detrimental for the child's behaviour.

## References

1.      Trumpff C, De Schepper J, Tafforeau J, Van Oyen H, Vanderfaeillie J, Vandevijvere S. Mild iodine deficiency in pregnancy in Europe and its consequences for cognitive and psychomotor development of children: a review. J Trace Elem Med Biol. 2013;27(3):174-83.

2.      Oppenheimer JH, Schwartz HL, Mariash CN, Kinlaw WB, Wong NC, Freake HC. Advances in our understanding of thyroid hormone action at the cellular level. Endocrine Reviews. 1987;8(3):288-308.

3.      Gibson RS. Principles of Nutritional Assessment 2nd Ed. . Oxford, UK: Oxford University Press; 1990.

4.      Manzano J, Bernal J, Morte B. Influence of thyroid hormones on maturation of rat cerebellar astrocytes. International Journal of Developmental Neuroscience. 2007;25(3):171-9.

5.      Daniels GH, Dayan CM. Thyroid Disorders. Oxford, UK: Health Press; 2006.

6.      Friesema EC, Jansen J, Visser TJ. Thyroid hormone transporters. Biochemical Society transactions. 2005;33(Pt 1):228-32.

7.      Szkudlinski MW, Fremont V, Ronin C, Weintraub BD. Thyroid-stimulating hormone and thyroid-stimulating hormone receptor structure-function relationships. Physiological Reviews. 2002;82(2):473-502.

8.      Truter I. Clinical review: Hyper- and hypothyroidism. SA Pharmaceutical Journal. 2011;78(6):10-4.

9.      Stricker RT, Echenard M, Eberhart R, Chevailler MC, Perez V, Quinn FA, et al. Evaluation of maternal thyroid function during pregnancy: the importance of using gestational age-specific reference intervals. Eur J Endocrinol. 2007;157(4):509-14.

10.     Bougma K, Aboud FE, Harding KB, Marquis GS. Iodine and mental development of children 5 years old and under: a systematic review and meta-analysis. Nutrients. 2013;5(4):1384-416.

11.     Vanderpump MPJ, Lazarus JH, Smyth PP, Laurberg P, Holder RL, Boelaert K, et al. Iodine status of UK schoolgirls: A cross-sectional survey. The Lancet. 2011;377(9782):2007-12.

12.     Glinoer D, De Nayer P, Bourdoux P, Lemone M, Robyn C, Van Steirteghem A, et al. Regulation of maternal thyroid during pregnancy. Journal of Clinical Endocrinology and Metabolism. 1990;71(2):276-87.

13.     Glinoer D. The regulation of thyroid function in pregnancy: Pathways of endocrine adaptation from physiology to pathology. Endocrine Reviews. 1997;18(3):404-33.

14.     Lazarus JH, Premawardhana LDKE. Screening for thyroid disease in pregnancy. Journal of Clinical Pathology. 2005;58(5):449-52.

15.     Bernal J, Nunez J. Thyroid hormones and brain development. European Journal of Endocrinology. 1995;133(4):390-8.

16.     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(2):147-209.

17.     Zimmermann MB. Iodine deficiency. Endocrine Reviews. 2009;30(4):376-408.

18.     Parkes IL, Schenker JG, Shufaro Y. Thyroid disorders during pregnancy. Gynecol Endocrinol. 2012;28(12):993-8.

19.     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(11):1365-74.

20.     Ritchie JWA, Taylor PM. Role of the System L permease LAT1 in amino acid and iodothyronine transport in placenta. Biochemical Journal. 2001;356(3):719-25.

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

22.     Rovet JF. The role of thyroid hormones for brain development and cognitive function. Endocrine Development. 2014;26:26-43.

23.     De Escobar GM. The role of thyroid hormone in fetal neurodevelopment. Journal of Pediatric Endocrinology and Metabolism. 2001;14(SUPPL. 6):1453-62.

24.     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(6):1719-22.

25.     Zoeller RT, Rovet J. Timing of thyroid hormone action in the developing brain: clinical observations and experimental findings. J Neuroendocrinol. 2004;16(10):809-18.

26.     Bunevičius R, Prange Jr AJ. Thyroid disease and mental disorders: Cause and effect or only comorbidity? Current Opinion in Psychiatry. 2010;23(4):363-8.

27.     Zoeller RT, Rovet J. Timing of thyroid hormone action in the developing brain: clinical observations and experimental findings. Journal of Neuroendocrinology. 2004;16(10):809-18.

28.     Shapira SK, Hinton CF, Held PK, Jones E, Harry Hannon W, Ojodu J. Single newborn screen or routine second screening for primary congenital hypothyroidism. Molecular Genetics and Metabolism. 2015;116(3):125-32.

29.     Delange F. Iodine deficiency as a cause of brain damage. Postgraduate medical journal. 2001;77(906):217-20.

30.     Vermiglio F, Lo Presti VP, Castagna MG, Violi MA, Moleti M, Finocchiaro MD, et al. Increased risk of maternal thyroid failure with pregnancy progression in an iodine deficient area with major iodine deficiency disorders. Thyroid : official journal of the American Thyroid Association. 1999;9(1):19-24.

31.     Wasserman EE, Pillion JP, Duggan A, Nelson K, Rohde C, Seaberg EC, et al. Childhood IQ, hearing loss, and maternal thyroid autoimmunity in the Baltimore Collaborative Perinatal Project. Pediatric Research. 2012;72(5):525-30.

32.     Hynes KL, Otahal P, Hay I, Burgess JR. Mild iodine deficiency during pregnancy is associated with reduced educational outcomes in the offspring: 9-year follow-up of the gestational iodine cohort. The Journal of clinical endocrinology and metabolism. 2013;98(5):1954-62.

33.     Bath SC, Steer CD, Golding J, Emmett P, Rayman MP. Effect of inadequate iodine status in UK pregnant women on cognitive outcomes in their children: Results from the Avon Longitudinal Study of Parents and Children (ALSPAC). The Lancet. 2013;382(9889):331-7.

34.     Berbel P, Mestre JL, Santamaría A, Palazón I, Franco A, Graells M, et al. 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(5):511-9.

35.     Velasco I, Carreira M, Santiago P, Muela JA, Garcia-Fuentes E, Sanchez-Munoz B, et al. Effect of iodine prophylaxis during pregnancy on neurocognitive development of children during the first two years of life. J Clin Endocr Metab. 2009;94(9):3234-41.

36.     Murcia M, Rebagliato M, Iñiguez C, Lopez-Espinosa MJ, Estarlich M, Plaza B, et al. Effect of iodine supplementation during pregnancy on infant neurodevelopment at 1 year of age. American Journal of Epidemiology. 2011;173(7):804-12.

37.     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.

38.     Cooper DS. Subclinical hypothyroidism. Thyroid Research and Practice. 2013;10(4):9.

39.     Wilson S, Parle JV, Roberts LM, Roalfe AK, Hobbs FD, Clark P, et al. Prevalence of subclinical thyroid dysfunction and its relation to socioeconomic deprivation in the elderly: a community-based cross-sectional survey. The Journal of clinical endocrinology and metabolism. 2006;91(12):4809-16.

40.     Owen PJ, Lazarus JH. The treatment of post-partum thyroid disease. Journal of endocrinological investigation. 2003;26(4):290-1.

41.     Stagnaro-Green A, Abalovich M, Alexander E, Azizi F, Mestman J, Negro R, et al. Guidelines of the American Thyroid Association for the diagnosis and management of thyroid disease during pregnancy and postpartum. Thyroid : official journal of the American Thyroid Association. 2011;21(10):1081-125.

42.     Krassas GE, Poppe K, Glinoer D. Thyroid function and human reproductive health. Endocrine Reviews. 2010;31(5):702-55.

43.     Canaris GJ, Manowitz NR, Mayor GH, Ridgway EC. The Colorado thyroid disease prevalence study. Archives of internal medicine. 2000;160(4):526-34.

44.     Ghassabian A, Henrichs J, Tiemeier H. Impact of mild thyroid hormone deficiency in pregnancy on cognitive function in children: Lessons from the Generation R Study. Best Practice and Research: Clinical Endocrinology and Metabolism. 2014;28(2):221-32.

45.     Modesto T, Tiemeier H, Peeters RP, Jaddoe VV, Hofman A, Verhulst FC, et al. Maternal mild thyroid hormone insufficiency in early pregnancy and attention-deficit/hyperactivity disorder symptoms in children. JAMA Pediatrics. 2015;169(9):838-45.

46.     Marchioni E, Fumarola A, Calvanese A, Piccirilli F, Tommasi V, Cugini P, et al. Iodine deficiency in pregnant women residing in an area with adequate iodine intake. Nutrition. 2008;24(5):458-61.

47.     Hiéronimus S, Bec-Roche M, Ferrari P, Chevalier N, Fénichel P, Brucker-Davis F. Iodine status and thyroid function of 330 pregnant women from Nice area assessed during the second part of pregnancy. Annales d'Endocrinologie. 2009;70(4):218-24.

48.     Ruiz AM, Martínez EG, Rodríguez MAR, de Llano JMA, Hernández CF, Rodríguez MM, et al. Prevalence of iodine deficiency in

pregnant women in the health area of palencia (spain). Endocrinologia y Nutricion. 2009;56(10):452-7.

49.    Li Y, Shan Z, Teng W, Yu X, Li Y, Fan C, et al. Abnormalities of maternal thyroid function during pregnancy affect neuropsychological development of their children at 25-30 months. Clinical endocrinology. 2010;72(6):825-9.

50.    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(3):282-8.

51.    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(6):435-8.

52.    Ghassabian A, El Marroun H, Peeters RP, Jaddoe VW, Hofman A, Verhulst FC, et al. 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(7):2383-90.

53.    Henrichs J, Bongers-Schokking JJ, Schenk JJ, Ghassabian A, Schmidt HG, Visser TJ, et al. Maternal thyroid function during early pregnancy and cognitive functioning in early childhood: The generation R study. Journal of Clinical Endocrinology and Metabolism. 2010;95(9):4227-34.

54.    Haddow JE, Palomaki GE, Allan WC, Williams JR, Knight GJ, Gagnon J, et al. Maternal thyroid deficiency during pregnancy and subsequent neuropsychological development of the child. The New England journal of medicine. 1999;341(8):549-55.

55.    Williams F, Watson J, Ogston S, Hume R, Willatts P, Visser T, et al. Mild maternal thyroid dysfunction at delivery of infants born </=34 weeks and neurodevelopmental outcome at 5.5 years. The Journal of clinical endocrinology and metabolism. 2012;97(6):1977-85.

56.    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(3):291-5.

57.    Pop VJ, Kuijpens JL, van Baar AL, Verkerk G, van Son MM, de Vijlder JJ, et al. Low maternal free thyroxine concentrations during early pregnancy are associated with impaired psychomotor development in infancy. Clin Endocrinol. 1999;50(2):149-55.

58.     Bath SC, Steer CD, Golding J, Emmett P, Rayman MP. Effect of inadequate iodine status in UK pregnant women on cognitive outcomes in their children: results from the Avon Longitudinal Study of Parents and Children (ALSPAC). Lancet. 2013;382(9889):331-7.
59.     De Escobar GM, Obregon MJ, del Rey FE. Maternal thyroid hormones early in pregnancy and fetal brain development. Best Pract Res Clin Endocrinol Metab. 2004;18(2):225-48.
60.     Boeck P. Intelligence, Where to Look, Where to Go? Journal of Intelligence. 2013;1(1):5-24.
61.     Sternberg RJ. The concept of intelligence and its role in lifelong learning and success. American Psychologist. 1997;52(10):1030-7.
62.     Sternberg RJ, Grigorenko EL. The general factor of intelligence: How general is it? Mahwah, New Jersey: Psychology Press; 2002.
63.     Carroll JB. Human cognitive abilities: A survey of factor-analytic studies. Cambridge: Cambridge University Press; 1993.
64.     Johnson W, Bouchard Jr TJ. The structure of human intelligence: It is verbal, perceptual, and image rotation (VPR), not fluid and crystallized. Intelligence. 2005;33(4):393-416.
65.     Hampshire A, Highfield RR, Parkin BL, Owen AM. Fractionating Human Intelligence. Neuron. 2012;76(6):1225-37.
66.     Schmidt FLaH, J. E. Select on intelligence. In: Locke EA, editor. The blackwell handbook of organizational principles. Oxford: Blackwell; 2000. p. 3-14.
67.     Ang S, Van Dyne L. Handbook of cultural intelligence. New York: Routledge; 2015.
68.     Jensen AR. Genetics and education. London,UK: Methuen; 1972.
69.     Gottfredson LS. Intelligence: is it the epidemiologists' elusive" fundamental cause" of social class inequalities in health? Journal of personality and social psychology. 2004;86(1):174.
70.     Hogan R. In defense of personality measurement: New wine for old whiners. Human Performance. 2005;18(4):331-41.
71.     Jensen AR. The g factor: The science of mental ability: Praeger Westport, CT; 1998.
72.     Judge TA, Colbert AE, Ilies R. Intelligence and leadership: a quantitative review and test of theoretical propositions. Journal of Applied Psychology. 2004;89(3):542.
73.     Neisser U, Boodoo G, Bouchard Jr TJ, Boykin AW, Brody N, Ceci SJ, et al. Intelligence: knowns and unknowns. American psychologist. 1996;51(2):77.
74.     Roberts BW, Kuncel NR, Shiner R, Caspi A, Goldberg LR. The power of personality: The comparative validity of personality traits, socioeconomic status, and cognitive ability for predicting important

life outcomes. Perspectives on Psychological Science. 2007;2(4):313-45.

75.    Sternberg RJ, Grigorenko E, Bundy DA. The predictive value of IQ. Merrill-Palmer Quarterly. 2001;47(1):1-41.

76.    Snowling MJ, Bishop DVM, Stothard SE, Chipchase B, Kaplan C. Psychosocial outcomes at 15 years of children with a preschool history of speech-language impairment. Journal of Child Psychology and Psychiatry and Allied Disciplines. 2006;47(8):759-65.

77.    Koenen KC, Moffitt TE, Roberts AL, Martin LT, Kubzansky L, Harrington H, et al. Childhood IQ and adult mental disorders: A test of the cognitive reserve hypothesis. American Journal of Psychiatry. 2009;166(1):50-7.

78.    Wechsler D. Wechsler intelligence scale for children–Fourth Edition (WISC-IV). San Antonio, TX: The Psychological Corporation; 2003.

79.    Bayley N. Bayley Scale of Infant and Toddler Devlopment- third edition: Technical manual. . San Antonio, TX: Harcourt Asssessment; 2006.

80.    Albers CA, Grieve AJ. Review of Bayley Scales of Infant and Toddler Development. Journal of Psychoeducational Assessment. 2007;25(2):180-90.

81.    Plomin R, Spinath FM. Intelligence: Genetics, Genes, and Genomics. Journal of Personality and Social Psychology. 2004;86(1):112-29.

82.    McGue M, Bouchard T.J, Jr., Iacono WG, Lykken DT. Behavioral genetics of cognitive ability: A life-span perspective. In: Plomin RaM, G. E. , editor. Nature, Nurture and Psychology. Washington, DC, US: American Psychological Association; 1993. p. 59-76.

83.    Brans RGH, Kahn R, Schnack HG, Van Baal GCM, Posthuma D, Van Haren NEM, et al. Brain plasticity and intellectual ability are influenced by shared genes. Journal of Neuroscience. 2010;30(16):5519-24.

84.    Petanjek Z, Judaš M, Šimić G, Rašin MR, Uylings HBM, Rakic P, et al. Extraordinary neoteny of synaptic spines in the human prefrontal cortex. Proceedings of the National Academy of Sciences of the United States of America. 2011;108(32):13281-6.

85.    Shaw P, Greenstein D, Lerch J, Clasen L, Lenroot R, Gogtay N, et al. Intellectual ability and cortical development in children and adolescents. Nature. 2006;440(7084):676-9.

86.    Bartels M, Rietveld MJH, Van Baal GCM, Boomsma DI. Genetic and environmental influences on the development of intelligence. Behavior Genetics. 2002;32(4):237-49.

87.     Brant AM, Haberstick BC, Corley RP, Wadsworth SJ, Defries JC, Hewitt JK. The developmental etiology of high IQ. Behavior Genetics. 2009;39(4):393-405.

88.     Haworth CMA, Wright MJ, Luciano M, Martin NG, De Geus EJC, Van Beijsterveldt CEM, et al. The heritability of general cognitive ability increases linearly from childhood to young adulthood. Molecular Psychiatry. 2010;15(11):1112-20.

89.     Lenroot RK, Schmitt JE, Ordaz SJ, Wallace GL, Neale MC, Lerch JP, et al. Differences in genetic and environmental influences on the human cerebral cortex associated with development during childhood and adolescence. Human Brain Mapping. 2009;30(1):163-74.

90.     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(1):18-20.

91.     Su PY, Huang K, Hao JH, Xu YQ, Yan SQ, Li T, et al. 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(10):3234-41.

92.     Craig WY, Allan WC, Kloza EM, Pulkkinen AJ, Waisbren S, Spratt DI, et al. Mid-gestational maternal free thyroxine concentration and offspring neurocognitive development at age two years. J Clin Endocr Metab. 2012;97(1):E22-8.

93.     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.

94.     Wechsler D. Preschool and Primary Scale of Intelligence (; WPPSI-III). San Antonio, TX: The Psychological Corporation; 2002.

95.     Grau G, Aguayo A, Vela A, Aniel-Quiroga A, Espada M, Miranda G, et al. 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.

96.     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.

97.     Oerbeck B, Sundet K, Kase BF, Heyerdahl S. Congenital hypothyroidism: no adverse effects of high dose thyroxine treatment on adult memory, attention, and behaviour. Archives of disease in childhood. 2005;90(2):132-7.

98. Wheeler SM, Willoughby KA, McAndrews MP, Rovet JF. Hippocampal size and memory functioning in children and adolescents with congenital hypothyroidism. Journal of Clinical Endocrinology and Metabolism. 2011;96(9):E1427-E34.

99. Gilbert ME, Paczkowski C. Propylthiouracil (PTU)-induced hypothyroidism in the developing rat impairs synaptic transmission and plasticity in the dentate gyrus of the adult hippocampus. Developmental Brain Research. 2003;145(1):19-29.

100. Lavado-Autric R, Ausó E, García-Velasco JV, Del Carmen Arufe M, Escobar del Rey F, Berbel P, et al. Early maternal hypothyroxinemia alters histogenesis and cerebral cortex cytoarchitecture of the progeny. Journal of Clinical Investigation. 2003;111(7):1073-82.

101. Svoboda E, McKinnon MC, Levine B. The functional neuroanatomy of autobiographical memory: A meta-analysis. Neuropsychologia. 2006;44(12):2189-208.

102. Cabeza R, Prince SE, Daselaar SM, Greenberg DL, Budde M, Dolcos F, et al. Brain activity during episodic retrieval of autobiographical and laboratory events: An fMRI study using a novel photo paradigm. Journal of Cognitive Neuroscience. 2004;16(9):1583-94.

103. 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.

104. 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(4):419-29.

105. Willoughby KA, McAndrews MP, Rovet JF. Effects of maternal hypothyroidism on offspring hippocampus and memory. Thyroid : official journal of the American Thyroid Association. 2014;24(3):576-84.

106. Pharoah POD, Connolly KJ. Relationship between maternal thyroxine levels during pregnancy and memory function in childhood. Early Human Development. 1991;25(1):43-51.

107. Williams FLR, Watson J, Ogston SA, Visser TJ, Hume R, Willatts P. Maternal and umbilical cord levels of T4, FT4, TSH, TPOAb, and TgAb in term infants and neurodevelopmental outcome at 5.5 years. Journal of Clinical Endocrinology and Metabolism. 2013;98(2):829-38.

108.   Man EB. Thyroid function in pregnancy and infancy. Maternal hypothyroxinemia and retardation of progeny. CRC critical reviews in clinical laboratory sciences. 1972;3(2):203-25.

109.   Klein RZ, Mitchell ML. Maternal hypothyroidism and child development. A review. Hormone Research. 1999;52(2):55-9.

110.   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(3):172-9.

111.   Radetti G, Gentili L, Paganini C, Oberhofer R, Deluggi I, Delucca A. Psychomotor and audiological assessment of infants born to mothers with subclinical thyroid dysfunction in early pregnancy. Minerva Pediatrica. 2000;52(12):691-8.

112.   Henrichs J, Ghassabian A, Peeters RP, Tiemeier H. Maternal hypothyroxinemia and effects on cognitive functioning in childhood: How and why? Clin Endocrinol. 2013;79(2):152-62.

113.   Pop VJ, Kuijpens JL, van Baar AL, Verkerk G, van Son MM, de Vijlder JJ, et al. Low maternal free thyroxine concentrations during early pregnancy are associated with impaired psychomotor development in infancy. Clin Endocrinol. 1999;50(2):149-55.

114.   Noten AME, Loomans EM, Vrijkotte TGM, van de Ven PM, van Trotsenburg ASP, Rotteveel J, et al. Maternal hypothyroxinaemia in early pregnancy and school performance in 5-year-old offspring. Eur J Endocrinol. 2015;173(5):563-71.

115.   Glinoer D, Delange F. The potential repercussions of maternal, fetal, and neonatal hypothyroxinemia on the progeny. Thyroid. 2000;10(10):871-87.

116.   Forrest D. The developing brain and maternal thyroid hormone: Finding the links. Endocrinology. 2004;145(9):4034-6.

117.   Goldey ES, Kehn LS, Rehnberg GL, Crofton KM. Effects of developmental hypothyroidism on auditory and motor function in the rat. Toxicology and Applied Pharmacology. 1995;135(1):67-76.

118.   Xue-Yi C, Xin-Min J, Zhi-Hong D, Rakeman MA, Ming-Li Z, O'Donnell K, et al. Timing of vulnerability of the brain to iodine deficiency in endemic cretinism. New Engl J Med. 1994;331(26):1739-44.

119.   Delange F, de Benoist B, Pretell E, Dunn JT. Iodine deficiency in the world: Where do we stand at the turn of the century? Thyroid : official journal of the American Thyroid Association. 2001;11(5):437-47.

120. Debruyne F, Vanderschueren-Lodeweyckx M, Bastijns P. Hearing in congenital hypothyroidism. International Journal of Audiology. 1983;22(4):404-9.

121. Rovet J, Walker W, Bliss B, Buchanan L, Ehrlich R. Long-term sequelae of hearing impairment in congenital hypothyroidism. Journal of Pediatrics. 1996;128(6):776-83.

122. Vanderschueren-Lodeweyckx M, Debruyne F, Dooms L, Eggermont E, Eeckels R. Sensorineural hearing loss in sporadic congenital hypothyroidism. Archives of Disease in Childhood. 1983;58(6):419-22.

123. Gottschalk B, Richman RA, Lewandowski L. Subtle speech and motor deficits of children with congenital hypothyroid treated early. Developmental Medicine and Child Neurology. 1994;36(3):216-20.

124. Wasserman EE, Nelson K, Rose NR, Eaton W, Pillion JP, Seaberg E, et al. Maternal thyroid autoantibodies during the third trimester and hearing deficits in children: an epidemiologic assessment. Am J Epidemiol. 2008;167(6):701-10.

125. Association AP. Diagnostic and statistical manual of mental disorders American Psychiatric Association. Washington, DC: American Psychiatric Pub; 1994. 471-5 p.

126. Rappley MD. Attention deficit-hyperactivity disorder. New Engl J Med. 2005;352(2):165-73.

127. Polanczyk G, de Lima MS, Horta BL, Biederman J, Rohde LA. The worldwide prevalence of ADHD: a systematic review and metaregression analysis. The American journal of psychiatry. 2007;164(6):942-8.

128. Thapar A, Cooper M, Jefferies R, Stergiakouli E. What causes attention deficit hyperactivity disorder? Archives of Disease in Childhood. 2012;97(3):260-5.

129. Getahun D, Jacobsen SJ, Fassett MJ, Chen W, Demissie K, Rhoads GG. Recent trends in childhood attention-deficit/hyperactivity disorder. JAMA Pediatrics. 2013;167(3):282-8.

130. Thapar A, Cooper M, Eyre O, Langley K. Practitioner review: what have we learnt about the causes of ADHD? Journal of Child Psychology and Psychiatry. 2013;54(1):3-16.

131. Thapar A, Holmes J, Poulton K, Harrington R. Genetic basis of attention deficit and hyperactivity. The british journal of psychiatry. 1999;174(2):105-11.

132. Faraone SV, Perlis RH, Doyle AE, Smoller JW, Goralnick JJ, Holmgren MA, et al. Molecular genetics of attention-deficit/hyperactivity disorder. Biological psychiatry. 2005;57(11):1313-23.

133.   Nikolas MA, Burt SA. Genetic and environmental influences on ADHD symptom dimensions of inattention and hyperactivity: a meta-analysis. Journal of Abnormal Psychology. 2010;119(1):1.

134.   Morrison JR, Stewart MA. The psychiatric status of the legal families of adopted hyperactive children. Archives of general psychiatry. 1973;28(6):888-91.

135.   Cantwell DP. Genetics of hyperactivity. Journal of child psychology and psychiatry. 1975;16:261-4.

136.   Cunningham L, Cadoret RJ, Loftus R, Edwards JE. Studies of adoptees from psychiatrically disturbed biological parents: Psychiatric conditions in childhood and adolescence. The British Journal of Psychiatry. 1975;126(6):534-49.

137.   Alberts-Corush J, Firestone P, Goodman JT. Attention and impulsivity characteristics of the biological and adoptive parents of hyperactive and normal control children. American Journal of Orthopsychiatry. 1986;56(3):413.

138.   Gizer IR, Ficks C, Waldman ID. Candidate gene studies of ADHD: a meta-analytic review. Human genetics. 2009;126(1):51-90.

139.   Asherson P, Brookes K, Franke B, Chen W, Gill M, Ebstein RP, et al. Confirmation that a specific haplotype of the dopamine transporter gene is associated with combined-type ADHD. American Journal of Psychiatry. 2007;164(4):674-7.

140.   Grady DL, Harxhi A, Smith M, Flodman P, Spence MA, Swanson JM, et al. Sequence variants of the DRD4 gene in autism: further evidence that rare DRD4 7R haplotypes are ADHD specific. American Journal of Medical Genetics Part B: Neuropsychiatric Genetics. 2005;136(1):33-5.

141.   Neale BM, Medland SE, Ripke S, Asherson P, Franke B, Lesch K, et al. Meta-analysis of genome-wide association studies of attention-deficit/hyperactivity disorder. Journal of the American Academy of Child & Adolescent Psychiatry. 2010;49(9):884-97.

142.   Rutter M. Genes and behavior: Nature-nurture interplay explained. London: Blackwell Publishing; 2006.

143.   Hauser P, Zametkin AJ, Martinez P, Vitiello B, Matochik JA, Mixson JA, et al. Attention deficit-hyperactivity disorder in people with generalized resistance to thyroid hormone. New Engl J Med. 1993;328(14):997-1001.

144.   Weiss RE, Stein MA, Duck SC, Chyna B, Phillips W, O'Brien T, et al. Low intelligence but not attention deficit hyperactivity disorder is associated with resistance to thyroid hormone caused by mutation R316H in the thyroid hormone receptor beta gene. The Journal of Clinical Endocrinology & Metabolism. 1994;78(6):1525-8.

145.  Valentine J, Rossi E, O'LEARY P, Parry TS, Kurinczuk JJ, Sly P. Thyroid function in a population of children with attention deficit hyperactivity disorder. Journal of paediatrics and child health. 1997;33(2):117-20.

146.  Spencer T, Biederman J, Wilens T, Guite J, Harding M. ADHD and thyroid abnormalities: a research note. Journal of Child Psychology and Psychiatry. 1995;36(5):879-85.

147.  Stein MA, Weiss RE. Thyroid function tests and neurocognitive functioning in children referred for attention deficit/hyperactivity disorder. Psychoneuroendocrinology. 2003;28(3):304-16.

148.  Päkkilä F, Männistö T, Pouta A, Hartikainen AL, Ruokonen A, Surcel HM, et al. The impact of gestational thyroid hormone concentrations on ADHD symptoms of the child. Journal of Clinical Endocrinology and Metabolism. 2014;99(1):E1-E8.

149.  Ghassabian A, Bongers-Schokking JJ, De Rijke YB, Van Mil N, Jaddoe VWV, De Muinck Keizer-Schrama SMPF, et al. Maternal thyroid autoimmunity during pregnancy and the risk of attention deficit/hyperactivity problems in children: The generation r study. Thyroid : official journal of the American Thyroid Association. 2012;22(2):178-86.

150.  Mathew JL, Mukherjee SB, Sukumar SP, Bhadada S. Maternal mild thyroid insufficiency and risk of attention deficit hyperactivity disorder. Indian Pediatrics. 2015;52(9):797-801.

151.  Vermiglio F, Lo Presti VP, Moleti M, Sidoti M, Tortorella G, Scaffidi G, et al. 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(12):6054-60.

152.  Beversdorf DQ, Manning SE, Hillier A, Anderson SL, Nordgren RE, Walters SE, et al. Timing of prenatal stressors and autism. Journal of autism and developmental disorders. 2005;35(4):471-8.

153.  Rodier PM. Environmental causes of central nervous system maldevelopment. Pediatrics. 2004;113(Supplement 3):1076-83.

154.  Lauritsen MB. Autism spectrum disorders. European Child and Adolescent Psychiatry. 2013;22(SUPPL.1):S37-S42.

155.  Association AP. Diagnostic and statistical manual of mental disorders, (DSM-5®). Washington, DC: American Psychiatric Pub; 2013.

156.  Baron-Cohen S, Scott FJ, Allison C, Williams J, Bolton P, Matthews FE, et al. Prevalence of autism-spectrum conditions: UK

school-based population study. The British Journal of Psychiatry. 2009;194(6):500-9.

157. Society TNA. Statistics: how many people have autistic spectrum disorders? 2014 [Available from: http://www.autism.org.uk/about/what-is/myths-facts-stats.aspx.

158. Fischbach GD, Lord C. The Simons Simplex Collection: a resource for identification of autism genetic risk factors. Neuron. 2010;68(2):192-5.

159. Tansey KE, Brookes KJ, Hill MJ, Cochrane LE, Gill M, Skuse D, et al. Oxytocin receptor (OXTR) does not play a major role in the aetiology of autism: genetic and molecular studies. Neuroscience letters. 2010;474(3):163-7.

160. Constantino JN, Lajonchere C, Lutz M, Gray T, Abbacchi A, McKenna K, et al. Autistic social impairment in the siblings of children with pervasive developmental disorders. American Journal of Psychiatry. 2014;163(2).

161. Christian SL, Brune CW, Sudi J, Kumar RA, Liu S, Karamohamed S, et al. Novel submicroscopic chromosomal abnormalities detected in autism spectrum disorder. Biological psychiatry. 2008;63(12):1111-7.

162. Matuszek G, Talebizadeh Z. Autism Genetic Database (AGD): a comprehensive database including autism susceptibility gene-CNVs integrated with known noncoding RNAs and fragile sites. BMC medical genetics. 2009;10(1):102.

163. Szatmari P, Paterson AD, Zwaigenbaum L, Roberts W, Brian J, Liu X, et al. Mapping autism risk loci using genetic linkage and chromosomal rearrangements. Nature genetics. 2007;39(3):319-28.

164. Volk HE, Lurmann F, Penfold B, Hertz-Picciotto I, McConnell R. Traffic-related air pollution, particulate matter, and autism. JAMA psychiatry. 2013;70(1):71-7.

165. Croen LA, Grether JK, Yoshida CK, Odouli R, Hendrick V. Antidepressant use during pregnancy and childhood autism spectrum disorders. Archives of general psychiatry. 2011;68(11):1104-12.

166. Zerbo O, Qian Y, Yoshida CK, Grether JK, Van de Water J, Croen LA. Maternal infection during pregnancy and autism spectrum disorders. Journal of autism and developmental disorders. 2013;45(12):4015-25.

167. Surén P, Roth C, Bresnahan M, Haugen M, Hornig M, Hirtz D, et al. Association between maternal use of folic acid supplements and risk of autism spectrum disorders in children. Jama. 2013;309(6):570-7.

168.  De Escobar GM, Obregón MJ, Del Rey FE. Role of thyroid hormone during early brain development. Eur J Endocrinol. 2004;151(Suppl 3):U25-U37.

169.  Bernal J, Guadaño-Ferraz A, Morte B. Perspectives in the study of thyroid hormone action on brain development and function. Thyroid : official journal of the American Thyroid Association. 2003;13(11):1005-12.

170.  Hoshiko S, Grether JK, Windham GC, Smith D, Fessel K. Are thyroid hormone concentrations at birth associated with subsequent autism diagnosis? Autism Research. 2011;4(6):456-63.

171.  Soldin OP, Lai S, Lamm SH, Mosee S. Lack of a relation between human neonatal thyroxine and pediatric neurobehavioral disorders. Thyroid : official journal of the American Thyroid Association. 2003;13(2):193-8.

172.  Cohen DJ, Young JG, Lowe TL, Harcherik D. Thyroid hormone in autistic children. Journal of autism and developmental disorders. 1980;10(4):445-50.

173.  Chen SW, Zhong XS, Jiang LN, Zheng XY, Xiong YQ, Ma SJ, et al. Maternal autoimmune diseases and the risk of autism spectrum disorders in offspring: A systematic review and meta-analysis. Behavioural Brain Research. 2016;296:61-9.

174.  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(2):151-7.

175.  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.

176.  Román GC, Ghassabian A, Bongers-Schokking JJ, Jaddoe VWV, Hofman A, Rijke YB, et al. Association of gestational maternal hypothyroxinemia and increased autism risk. Annals of neurology. 2013;74(5):733-42.

177.  Reuss ML, Paneth N, Pinto-Martin JA, Lorenz JM, Susser M. The relation of transient hypothyroxinemia in preterm infants to neurologic development at two years of age. New Engl J Med. 1996;334(13):821-7.

178.  Korzeniewski SJ, Pinto-Martin JA, Whitaker AH, Feldman JF, Lorenz JM, Levy SE, et al. Association Between Transient Hypothyroxinaemia of Prematurity and Adult Autism Spectrum Disorder in a Low-Birthweight Cohort: An Exploratory Study. Paediatric and perinatal epidemiology. 2013;27(2):182-7.

179. Román GC. Autism: Transient in utero hypothyroxinemia related to maternal flavonoid ingestion during pregnancy and to other environmental antithyroid agents. Journal of the Neurological Sciences. 2007;262(1-2):15-26.

180. Yau VM, Lutsky M, Yoshida CK, Lasley B, Kharrazi M, Windham G, et al. Prenatal and Neonatal Thyroid Stimulating Hormone Levels and Autism Spectrum Disorders. Journal of autism and developmental disorders. 2015;45(3):719-30.

181. Clairman H, Skocic J, Lischinsky JE, Rovet J. Do children with congenital hypothyroidism exhibit abnormal cortical morphology? Pediatric Research. 2015;78(3):286-97.

182. Wheeler SM, McAndrews MP, Sheard ED, Rovet J. Visuospatial associative memory and hippocampal functioning in congenital hypothyroidism. Journal of the International Neuropsychological Society. 2012;18(1):49-56.

183. Kier EL, Truwit CL. The normal and abnormal genu of the corpus callosum: An evolutionary, embryologic, anatomic, and MR analysis. American Journal of Neuroradiology. 1996;17(9):1631-45.

184. Raybaud C. The corpus callosum, the other great forebrain commissures, and the septum pellucidum: Anatomy, development, and malformation. Neuroradiology. 2010;52(6):447-77.

185. Paul LK. Developmental malformation of the corpus callosum: A review of typical callosal development and examples of developmental disorders with callosal involvement. Journal of Neurodevelopmental Disorders. 2011;3(1):3-27.

186. Gravel C, Hawkes R. Maturation of the corpus callosum of the rat: I. Influence of thyroid hormones on the topography of callosal projections. Journal of Comparative Neurology. 1990;291(1):128-46.

187. Samadi A, Skocic J, Rovet JF. Children born to women treated for hypothyroidism during pregnancy show abnormal corpus callosum development. Thyroid : official journal of the American Thyroid Association. 2015;25(5):494-502.

188. Si SS, Qian M, Chen ZP, Ding WJ, Yang HC, Yan YQ, et al. Relationship between maternal thyroid function during the 1st and 2nd gestational trimester and child brain and neural development. Chinese Journal of Endemiology. 2012;31(3):259-62.

189. Thung SF, Funai EF, Grobman WA. The cost-effectiveness of universal screening in pregnancy for subclinical hypothyroidism. American journal of obstetrics and gynecology. 2009;200(3):267 e1-7.

190. Jouyandeh Z, Hasani-Ranjbar S, Qorbani M, Larijani B. Universal screening versus selective case-based screening for thyroid disorders in pregnancy. Endocrine. 2014;48(1):116-23.

191. Morreale de Escobar G, Jesús Obregón M, Escobar del Rey F. Is Neuropsychological Development Related to Maternal Hypothyroidism or to Maternal Hypothyroxinemia? The Journal of Clinical Endocrinology & Metabolism. 2000;85(11):3975-87.

192. Gharib H, Tuttle RM, Baskin HJ, Fish LH, Singer PA, McDermott MT. Consensus statement: Subclinical thyroid dysfunction: A joint statement on management from the American Association of Clinical Endocrinologists, the American Thyroid Association, and the Endocrine Society. Journal of Clinical Endocrinology and Metabolism. 2005;90(1):581-5.

193. Abalovich M, Amino N, Barbour LA, Cobin RH, De Groot LJ, Glinoer D, et al. Clinical practice guideline: Management of thyroid dysfunction during pregnancy and postpartum: An endocrine society clinical practice guideline. Journal of Clinical Endocrinology and Metabolism. 2007;92(8):S1-S47.

194. Vaidya B, Anthony S, Bilous M, Shields B, Drury J, Hutchison S, et al. Brief report: Detection of thyroid dysfunction in early pregnancy: Universal screening or targeted high-risk case finding? Journal of Clinical Endocrinology and Metabolism. 2007;92(1):203-7.

195. Vaidya B. Management of hypothyroidism in pregnancy: we must do better. Clin Endocrinol (Oxf). 2013;78(3):342-3.

196. Van den Boogaard E, Vissenberg R, Land JA, van Wely M, van der Post JAM, Goddijn M, et al. Significance of (sub) clinical thyroid dysfunction and thyroid autoimmunity before conception and in early pregnancy: a systematic review. Human reproduction update. 2011;17(5):605-19.

197. Negro R, Schwartz A, Gismondi R, Tinelli A, Mangieri T, Stagnaro-Green A. Universal screening versus case finding for detection and treatment of thyroid hormonal dysfunction during pregnancy. The Journal of Clinical Endocrinology & Metabolism. 2010;95(4):1699-707.

198. 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 [Internet]. 2013; (5). Available from: http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD007752.pub3/abstract.

199.   Gynecology ACoOa. ACOG practice bulletin: Thyroid disease in pregnancy - Number 37, August 2002. International Journal of Gynecology and Obstetrics. 2002;79(2):171-80.

200.   Lazarus JH, Bestwick JP, Channon S, Paradice R, Maina A, Rees R, et al. Antenatal thyroid screening and childhood cognitive function. The New England journal of medicine. 2012;366(6):493-501.

201. Moleti M, Vermiglio F, Trimarchi F. Maternal isolated hypothyroxinemia: To treat or not to treat? Journal of endocrinological investigation. 2009;32(9):780-2.

202.   De Escobar GM, Obregón MJ, Del Rey FE. Iodine deficiency and brain development in the first half of pregnancy. Public Health Nutrition. 2007;10(12 A):1554-70.

203.   Ghassabian A, Tiemeier H. Is measurement of maternal serum TSH sufficient screening in early pregnancy? A case for more randomized trials. Clin Endocrinol (Oxf). 2012;77(6):802-5.

204.   De Groot L, Abalovich M, Alexander EK, Amino N, Barbour L, Cobin RH, et al. Management of thyroid dysfunction during pregnancy and postpartum: an Endocrine Society clinical practice guideline. J Clin Endocrinol Metab. 2012;97(8):2543-65.

205.   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(2):76-94.

206. Wechsler D. Wechsler Preschool and Primary Scale of Intelligence - Third Edition (WPPSI-IIIUK). Oxford: The Psychological Corporation; 2003.

207.   Wechsler D. Preschool and Primary Scale of Intelligence. New York: The Psychological Corporation; 1967.

208. Wechsler D. Wechsler Preschool and Primary Scale of Intelligence - Fourth Edition (WPPSI-IV UK). London: Pearson Assessment; 2013.

209.   Lichtenberger EO. General measures of cognition for the preschool child. Mental Retardation and Developmental Disabilities Research Reviews. 2005;11(3):197-208.

210.   Field A. Discovering statistics using IBM SPSS statistics; fourth edition. London: Sage; 2013.

211.   Altman DG, Royston P. The cost of dichotomising continuous variables. British Medical Journal. 2006;332(7549):1080.

212.   McClelland G, Lynch JG, Irwin JR, Spiller SA, Fitzsimons GJ. Median Splits, Type II Errors, and False Positive Consumer Psychology: Don't Fight the Power. Journal of Comsumer Psychology. 2015;53:1-51.

213. Austin PC, Brunner LJ. Inflation of the type I error rate when a continuous confounding variable is categorized in logistic regression analyses. Statistics in medicine. 2004;23(7):1159-78.

214. Julvez J, Alvarez-Pedrerol M, Rebagliato M, Murcia M, Forns J, Garcia-Esteban R, et al. Thyroxine levels during pregnancy in healthy women and early child neurodevelopment. Epidemiology. 2013;24(1):150-7.

215. 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 & Metabolism. 2009;94(2):497-503.

216. Kamphaus RW. Clinical Assessment of Child and Adolescent Intelligence (2nd edition). New York: Springer Science and Business Media; 2005.

217. Howard RW. Objective evidence of rising population ability: a detailed examination of longitudinal chess data. Personality and Individual Differences. 2005;38(2):347-63.

218. Meisenberg G, Lawless E, Lambert E, Newton A. The Flynn effect in the Caribbean: Generational change of cognitive test performance in Dominica. Mankind Quarterly. 2005;46(1):29.

219. Lynn R. What has caused the Flynn effect? Secular increases in the Development Quotients of infants. Intelligence. 2009;37(1):16-24.

220. Hiscock M. The Flynn effect and its relevance to neuropsychology. Journal of clinical and experimental neuropsychology. 2007;29(5):514-29.

221. te Nijenhuis J, van der Flier H. Is the Flynn effect on g?: A meta-analysis. Intelligence. 2013;41(6):802-7.

222. Flynn JR. Massive IQ gains in 14 nations: What IQ tests really measure. Psychological bulletin. 1987;101(2):171.

223. Flynn JR. Are we getting smarter; Rising IQ in the Twenty-First Century. Cambridge: Cambridge University Press; 2012.

224. Bishop EG, Cherny SS, Corley R, Plomin R, DeFries JC, Hewitt JK. Development genetic analysis of general cognitive ability from 1 to 12 years in a sample of adoptees, biological siblings, and twins. Intelligence. 2003;31(1):31-49.

225. Petrill SA, Hewitt JK, Cherny SS, Lipton PA, Plomin R, Corley R, et al. Genetic and environmental contributions to general cognitive ability through the first 16 years of life. Developmental Psychology. 2004;40(5):805-12.

226. Hoekstra RA, Bartels M, Boomsma DI. Longitudinal genetic study of verbal and nonverbal IQ from early childhood to young adulthood. Learning and Individual Differences. 2007;17(2):97-114.

227. Duckworth AL, Quinn PD, Lynam DR, Loeber R, Stouthamer-Loeber M. Role of test motivation in intelligence testing. Proceedings of the National Academy of Sciences. 2011;108(19):7716-20.

228. Steinmayr R, Spinath B. Sex differences in school achievement: What are the roles of personality and achievement motivation? European journal of personality. 2008;22(3):185-209.

229. Freund PA, Holling H. Who wants to take an intelligence test? Personality and achievement motivation in the context of ability testing. Personality and Individual Differences. 2011;50(5):723-8.

230. Korkman M, Kirk, U. and Kemp, S. . NEPSY-II: A developmental neuropsychological assessment. San Antonio, TX: The Psychological Corporation; 2007.

231. Goodman R. The Strengths and Difficulties Questionnaire: a research note. Journal of child psychology and psychiatry. 1997;38(5):581-6.

232. 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(12):1528-36.

233. Rutter M, Bailey A, Lord C. The social communication questionnaire: Manual. Los Angeles, USA: Western Psychological Services; 2003.

234. Kaufman AS, Flanagan DP, Alfonso VC, Mascolo JT. Review of Wechsler Intelligence Scale for Children, (WISC-IV). Journal of Psychoeducational Assessment. 2006;24(3):278-95.

235. Archer RP, Buffington-Vollum JK, Stredny RV, Handel RW. A survey of psychological test use patterns among forensic psychologists. Journal of personality assessment. 2006;87(1):84-94.

236. Bowden SC. Theoretical convergence in assessment of cognition. Journal of Psychoeducational Assessment. 2013:0734282913478035.

237. Rabin LA, Barr WB, Burton LA. Assessment practices of clinical neuropsychologists in the United States and Canada: A survey of INS, NAN, and APA Division 40 members. Archives of Clinical Neuropsychology. 2005;20(1):33-65.

238. Sternberg RJ. Intelligence. Wiley Interdisciplinary Reviews: Cognitive Science. 2012;3(5):501-11.

239. Flanagan DP, McGrew KS, Ortiz SO. The Wechsler Intelligence Scales and Gf-Gc theory: A contemporary approach to interpretation: Allyn & Bacon; 2000.

240. Keith TZ, Fine JG, Taub GE, Reynolds MR, Kranzler JH. Higher order, multisample, confirmatory factor analysis of the Wechsler Intelligence Scale for Children—Fourth Edition: What does it measure. School Psychology Review. 2006;35(1):108-27.

241. Watkins MW. Orthogonal higher order structure of the Wechsler Intelligence Scale for Children. Psychological Assessment. 2006;18(1):123.

242. Wechsler D. Wechsler-Bellevue Intelligence Scale. New York: The Psychological Corporation; 1939.

243. Wechsler D. Wechsler-Bellevue intelligence Scale, Form II. New York, US: The Psychological Corporation; 1946.

244. Wechsler D. Wechsler Intelligence Scale for Children. New York: The Psychological Corporation; 1949.

245. Wechsler D. The Wechsler Intelligence Test for Children–Revised. San Antonio, TX: The Psychological Corporation; 1974.

246. Wechsler D. Manual for the Wechsler intelligence scale for children-III. New York: The Psychological Corporation; 1991.

247. Wechsler D. Wechsler Intelligence Scale for Children-Fifth Edition. Technical and interpretive manual. Bloomington MN: Pearson; 2014.

248. Chen H, Zhang O, Raiford SE, Zhu J, Weiss LG. Factor invariance between genders on the Wechsler Intelligence Scale for Children-Fifth Edition. Personality and Individual Differences. 2015;86:1-5.

249. Na SD, Burns TG. Wechsler Intelligence Scale for Children-V: Test Review. Applied Neuropsychology: Child. 2015(ahead-of-print):1-5.

250. Wechsler D. Wechsler Intelligence Scale for Children- Fourth UK Edition: Administration and scoring manual. London: Pearson Assessment; 2004.

251. Anastasi A, and Urbina, S. Psychological Testing (7th ed.). Upper Saddle River, NJ: Prentice Hall; 1997.

252. Korkman M, Kirk U, Kemp SJ. NEPSY: A developmental neuropsychological assessment. San Antonio, TX: Psychological Corporation; 1998.

253. Ahmad SA, Warriner EM. Review of the NEPSY: A developmental neuropsychological assessment. The Clinical Neuropsychologist. 2001;15(2):240-9.

254. Brooks BL, Sherman EMS, Strauss E. NEPSY-II: A developmental neuropsychological assessment. Child Neuropsychology. 2009;16(1):80-101.

255. Davis JL, Matthews RN. NEPSY-II Review: Korkman, M., Kirk, U., & Kemp, S. (2007). NEPSY—Second Edition (NEPSY-II). San Antonio,

TX: Harcourt Assessment. Journal of Psychoeducational Assessment. 2010;28(2):175-82.

256. Elliott CD. Differential ability scales-Second edition. San Antonio, TX: Harcourt Assessment; 2007.

257. Wechsler DaN, J. A. The Wechsler nonverbal scale of ability. San Antonio, TX: Harcourt Assessment; 2006.

258. 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(2):S95-S103.

259. Rutter M. A children's behaviour questionnaire for completion by teachers: preliminary findings. Child Psychology & Psychiatry & Allied Disciplines. 1967;8(1):1-11.

260. Goodman R. A modified version of the Rutter parent questionnaire including extra items on children's strengths: a research note. Journal of Child Psychology and Psychiatry. 1994;35(8):1483-94.

261. Caci H, Morin AJS, Tran A. Investigation of a bifactor model of the Strengths and Difficulties Questionnaire. European child & adolescent psychiatry. 2015;24(10):1291-301.

262. Russell G, Rodgers LR, Ford T. The strengths and difficulties questionnaire as a predictor of parent-reported diagnosis of autism spectrum disorder and attention deficit hyperactivity disorder. PloS one. 2013;8(12).

263. Goodman R. Psychometric properties of the strengths and difficulties questionnaire. Journal of the American Academy of Child & Adolescent Psychiatry. 2001;40(11):1337-45.

264. Glazebrook C, Hollis C, Heussler H, Goodman R, Coates L. Detecting emotional and behavioural problems in paediatric clinics. Child: care, health and development. 2003;29(2):141-9.

265. Goodman R, Meltzer HM, Bailey V. The Strengths and Difficulties Questionnaire: A pilot study on the validity of the self-report version. European child & adolescent psychiatry. 1998;7(3):125-30.

266. Curvis W, McNulty S, Qualter P. The validation of the self-report Strengths and Difficulties Questionnaire for use by 6-to 10-year-old children in the UK. British Journal of Clinical Psychology. 2014;53(1):131-7.

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

268.  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(2):189.

269.  Woerner W, Fleitlich-Bilyk B, Martinussen R, Fletcher J, Cucchiaro G, Dalgalarrondo P, et al. The Strengths and Difficulties Questionnaire overseas: evaluations and applications of the SDQ beyond Europe. European Child & Adolescent Psychiatry. 2004;13(2):ii47-ii54.

270.  DuPaul GJ. Parent and teacher ratings of ADHD symptoms: psychometric properties in a community-based sample. Journal of Clinical Child and Adolescent Psychology. 1991;20(3):245-53.

271.  Conners CK. Psychological Assessment Of Children With Minimal Brain Dysfunction. Annals of the New York Academy of Sciences. 1973;205(1):283-302.

272.  Cooper AM, Michels R. Diagnostic and statistical manual of mental disorders, revised (DSM-III-R). American Journal of Psychiatry. 1988;145(10):1300-1.

273.  Collett BR, Ohan JL, Myers KM. Ten-year review of rating scales. V: scales assessing attention-deficit/hyperactivity disorder. Journal of the American Academy of Child & Adolescent Psychiatry. 2003;42(9):1015-37.

274.  Le Couteur A, Rutter M, Lord C, Rios P, Robertson S, Holdgrafer M, et al. Autism diagnostic interview: A standardized investigator-based instrument. Journal of autism and developmental disorders. 1989;19(3):363-87.

275.  Lord C, Rutter M, Le Couteur A. Autism Diagnostic Interview-Revised: a revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. Journal of autism and developmental disorders. 1994;24(5):659-85.

276.  Berument SK, Rutter M, Lord C, Pickles A, Bailey A. Autism screening questionnaire: diagnostic validity. The British Journal of Psychiatry. 1999;175(5):444-51.

277.  Wilkinson L. A best practice guide to assessment and intervention for autism and Asperger syndrome in schools. London: Jessica Kingsley Publishers; 2010.

278.  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(3):469-80.

279.  Eaves LC, Wingert H, Ho HH. Screening for autism Agreement with diagnosis. Autism. 2006;10(3):229-42.

280.   Hales C, Channon S, Taylor PN, Draman MS, Muller I, Lazarus J, et al. The second wave of the Controlled Antenatal Thyroid Screening (CATS II) study: the cognitive assessment protocol. Bmc Endocr Disord. 2014;14(1):95.

281.   Govenment W. WIMD 2014 revised - 12 August 2015 2014 [Available                                                               from: https://statswales.wales.gov.uk/Catalogue/Community-Safety-and-Social-Inclusion/Welsh-Index-of-Multiple-Deprivation/WIMD-2014.

282.   OpenDataCommunities. 2010 Deprivation mapper 2010 [Available                                                                from: http://apps.opendatacommunities.org/showcase/deprivation.

283.   Pierce CA, Block RA, Aguinis H. Cautionary note on reporting eta-squared values from multifactor ANOVA designs. Educational and psychological measurement. 2004;64(6):916-24.

284.   W. R. Individual differences in personality and motivation: "non-cognitive" determinants of cognitive performance. . In: Baddeley A. D WAD, editor. Attention: Selection, Awareness, and Control A tribute to Donald Broadbent. Oxford: Oxford University Press; 1993. p. 346-73.

285.   Young JL. Test or toy? Materiality and the measurement of infant intelligence. History of psychology. 2015;18(2):103.

286.   Meijer J, Oostdam R. Effects of instruction and stage-fright on intelligence testing. European journal of psychology of education. 2011;26(1):143-61.

287.   Moutafi J, Furnham A, Tsaousis I. Is the relationship between intelligence and trait Neuroticism mediated by test anxiety? Personality and Individual Differences. 2006;40(3):587-97.

288.   Boake C. From the Binet–Simon to the Wechsler–Bellevue: Tracing the history of intelligence testing. Journal of clinical and experimental neuropsychology. 2002;24(3):383-405.

289.   Kaufman AS. IQ testing 101. New York, US: Springer Publishing Co; 2009.

290.   Kranzler JH, Floyd RG. Assessing intelligence in children and adolescents: A practical guide. London: Guilford Press; 2013.

291.   Bernstein D. Essentials of psychology. Wadsworth: Cengage Learning; 2013.

292.   Friedman NP, Miyake A, Corley RP, Young SE, DeFries JC, Hewitt JK. Not all executive functions are related to intelligence. Psychological science. 2006;17(2):172-9.

293.   Heaven PCL, Ciarrochi J. When IQ is not everything: Intelligence, personality and academic performance at school. Personality and Individual Differences. 2012;53(4):518-22.

294. Poropat AE. A meta-analysis of the five-factor model of personality and academic performance. Psychological bulletin. 2009;135(2):322.

295. Whitaker S, Wood C. The Distribution of Scaled Scores and Possible Floor Effects on the WISC-III and WAIS-III. Journal of Applied Research in Intellectual Disabilities. 2008;21(2):136-41.

296. Orsini A, Pezzuti L, Hulbert S. Beyond the floor effect on the Wechsler Intelligence Scale for Children–4th Ed.(WISC-IV): calculating IQ and Indexes of subjects presenting a floored pattern of results. Journal of Intellectual Disability Research. 2015;59(5):468-73.

297. Wheeler LR. A comparative study of the intelligence of East Tennessee mountain children. Journal of Educational Psychology. 1942;33:321-34.

298. Flynn JR. What is intelligence?: Beyond the Flynn effect. Cambridge: Cambridge University Press; 2007.

299. Lynn R. Differential rates of secular increase of five major primary abilities. Social Biology. 1990;37(1-2):137-41.

300. Lynn R, Hampson S. The rise of national intelligence: Evidence from Britain, Japan and the USA. Personality and individual differences. 1986;7(1):23-32.

301. Lynn R. Fluid intelligence but not vocabulary has increased in Britain, 1979–2008. Intelligence. 2009;37(3):249-55.

302. Lynn R. The role of nutrition in secular increases in intelligence. Personality and individual Differences. 1990;11(3):273-85.

303. Lynn R. Nutrition and intelligence. In: Vernon PA, editor. Biological approaches to the study of intelligence. Norwood, NJ: Ablex; 1993.

304. Lynn R. In support of nutrition theory. In: Neisser U, editor. The rising curve. Washington, DC: American Psychological Association; 1998.

305. Arija V, Esparó G, Fernández-Ballart J, Murphy MM, Biarnés E, Canals J. Nutritional status and performance in test of verbal and non-verbal intelligence in 6 year old children. Intelligence. 2006;34(2):141-9.

306. Colom R, Lluis-Font JM, Andrés-Pueyo A. The generational intelligence gains are caused by decreasing variance in the lower half of the distribution: Supporting evidence for the nutrition hypothesis. Intelligence. 2005;33(1):83-91.

307. Pérez-Lobato R, Ramos R, Calvente I, Arrebola JP, Ocón-Hernández O, Dávila-Arias C, et al. Thyroid status and its association

with cognitive functioning in healthy boys at 10 years of age. Eur J Endocrinol. 2015;172(2):129-39.

308. Wilcoxon JS, Nadolski GJ, Samarut J, Chassande O, Redei EE. Behavioral inhibition and impaired spatial learning and memory in hypothyroid mice lacking thyroid hormone receptor α. Behavioural brain research. 2007;177(1):109-16.

309. van Boxtel MPJ, Menheere PPCA, Bekers O, Hogervorst E, Jolles J. Thyroid function, depressed mood, and cognitive performance in older individuals: the Maastricht Aging Study. Psychoneuroendocrinology. 2004;29(7):891-8.

310. Wheeler SM, Willoughby KA, McAndrews MP, Rovet JF. Hippocampal size and memory functioning in children and adolescents with congenital hypothyroidism. The Journal of Clinical Endocrinology & Metabolism. 2011;96(9):E1427-E34.

311. Gilbert ME, Rovet J, Chen Z, Koibuchi N. Developmental thyroid hormone disruption: Prevalence, environmental contaminants and neurodevelopmental consequences. NeuroToxicology. 2012;33(4):842-52.

312. Rivas M, Naranjo JR. Thyroid hormones, learning and memory. Genes, Brain, & Behavior. 2007;6 Suppl 1:40-4.

313. Kline P. Handbook of psychological testing (second edition). London: Routledge; 2013.

314. 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(2):287-300.

315. Bongers-Schokking JJ, De Muinck Keizer-Schrama SMPF. Influence of timing and dose of thyroid hormone replacement on mental, psychomotor, and behavioral development in children with congenital hypothyroidism. Journal of Pediatrics. 2005;147(6):768-74.

316. Ghassabian A, Bongers-Schokking JJ, Henrichs J, Jaddoe VWV, Visser TJ, Visser W, et al. Maternal thyroid function during pregnancy and behavioral problems in the offspring: The generation R study. Pediatric Research. 2011;69(5 PART 1):454-9.

317. Talge NM, Neal C, Glover V. Antenatal maternal stress and long-term effects on child neurodevelopment: how and why? Journal of Child Psychology and Psychiatry. 2007;48(3-4):245-61.

318. O'Connor TG, Heron J, Golding J, Beveridge M, Glover V. Maternal antenatal anxiety and children's behavioural/emotional

problems at 4 years Report from the Avon Longitudinal Study of Parents and Children. The British Journal of Psychiatry. 2002;180(6):502-8.

319. Linnet KM, Dalsgaard S, Obel C, Wisborg K, Henriksen TB, Rodriguez A, et al. Maternal lifestyle factors in pregnancy risk of attention deficit hyperactivity disorder and associated behaviors: review of the current evidence. American Journal of Psychiatry. 2003;160(6):1028-40.

320. 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(1):45-54.

321. Chan S, Kachilele S, Hobbs E, Bulmer JN, Boelaert K, McCabe CJ, et al. Placental iodothyronine deiodinase expression in normal and growth-restricted human pregnancies. Journal of Clinical Endocrinology and Metabolism. 2003;88(9):4488-95.

322. Koopdonk-Kool JM, De Vijlder J, Veenboer GJ, Ris-Stalpers C, Kok JH, Vulsma T, et al. Type II and type III deiodinase activity in human placenta as a function of gestational age. The Journal of Clinical Endocrinology & Metabolism. 1996;81(6):2154-8.

323. Mortimer R, Galligan J, Cannell G, Addison R, Roberts M. Maternal to fetal thyroxine transmission in the human term placenta is limited by inner ring deiodination. The Journal of Clinical Endocrinology & Metabolism. 1996;81(6):2247-9.

324. Emerson C, Bambini G, Alex S, Castro M, Roti E, Braverman L. The Effect of Thyroid Dysfunction and Fasting on Placenta Inner Ring Deiodinase Activity in the Rat*. Endocrinology. 1988;122(3):809-16.

325. Liu J, Lynn R. Factor structure and sex differences on the Wechsler Preschool and Primary Scale of Intelligence in China, Japan and United States. Personality and Individual Differences. 2011;50(8):1222-6.

326. Halpern DF, LaMay ML. The Smarter Sex: A Critical Review of Sex Differences in Intelligence. Educational Psychology Review. 2000;12(2):229-46.

327. Fairweather H, Butterworth G. The WPPSI at four years: a sex difference in verbal-performance discrepancies. The British journal of educational psychology. 1977;47(1):85-90.

328. Kaiser SM, Reynolds CR. Sex differences on the Wechsler preschool and primary scale of intelligence. Personality and Individual Differences. 1985;6(3):405-7.

329. Keith TZ, Reynolds MR, Patel PG, Ridley KP. Sex differences in latent cognitive abilities ages 6 to 59: Evidence from the Woodcock–

Johnson III tests of cognitive abilities. Intelligence. 2008;36(6):502-25.

330. Nyborg H. Sex-related differences in general intelligence g, brain size, and social status. Personality and Individual Differences. 2005;39(3):497-509.

331. Reynolds MR, Keith TZ, Ridley KP, Patel PG. Sex differences in latent general and broad cognitive abilities for children and youth: Evidence from higher-order MG-MACS and MIMIC models. Intelligence. 2008;36(3):236-60.

332. Camarata S, Woodcock R. Sex differences in processing speed: Developmental effects in males and females. Intelligence. 2006;34(3):231-52.

333. Goldbeck L, Daseking M, Hellwig-Brida S, Waldmann HC, Petermann F. Sex differences on the German Wechsler intelligence test for children (WISC-IV). Journal of Individual Differences. 2015;31:22-8.

334. Weiss EM, Deisenhammer EA, Hinterhuber H, Marksteiner J. Gender differences in cognitive functions. Fortschritte der Neurologie-Psychiatrie. 2005;73(10):587-95.

335. Chen H, Zhu J. Factor invariance between genders of the Wechsler Intelligence Scale for Children–Fourth edition. Personality and Individual Differences. 2008;45(3):260-6.

336. Majeres RL. Sex differences in phonological coding: Alphabet transformation speed. Intelligence. 2007;35(4):335-46.

337. Mulenga K, Ahonen T, Aro M. Performance of Zambian children on the NEPSY: A pilot study. Developmental neuropsychology. 2001;20(1):375-83.

338. Van Roy B, Grøholt B, Heyerdahl S, Clench-Aas J. Self-reported strengths and difficulties in a large Norwegian population 10-19 years: Age and gender specific results of the extended SDQ-questionnaire. European Child and Adolescent Psychiatry. 2006;15(4):189-98.

339. 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(3):466-81.

340. Becker A, Rothenberger A, Sohn A, group Bs. Six years ahead: a longitudinal analysis regarding course and predictive value of the Strengths and Difficulties Questionnaire (SDQ) in children and adolescents. European child & adolescent psychiatry. 2014;24:715-25.

341. Makransky G, Bilenberg N. Psychometric Properties of the Parent and Teacher ADHD Rating Scale (ADHD-RS): Measurement Invariance Across Gender, Age, and Informant. Assessment. 2014;21(6):694-705.

342. Keshavarzi Z, Bajoghli H, Mohamadi MR, Holsboer-Trachsler E, Brand S. Attention deficit hyperactivity disorder in children is found to be related to the occurrence of ADHD in siblings and the male gender, but not to birth order, when compared to healthy controls. International Journal of Psychiatry in Clinical Practice. 2014;18(4):272-9.

343. Ghanizadeh A, Abotorabi-Zarchi M, Mohammadi MR, Firoozabadi A. Birth order and sibling gender ratio of a clinical sample of children and adolescents diagnosed with attention deficit hyperactivity disorder. Iranian Journal of Psychiatry. 2012;7(3):109-13.

344. Biederman J, Faraone SV. Attention-deficit hyperactivity disorder. Lancet. 2005;366(9481):237-48.

345. 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.

346. Biederman J, Mick E, Faraone SV, Braaten E, Doyle A, Spencer T, et al. Influence of gender on attention deficit hyperactivity disorder in children referred to a psychiatric clinic. American Journal of Psychiatry. 2002;159(1):36-42.

347. Broderick N, Wade JL, Meyer JP, Hull M, Reeve RE. Model Invariance Across Genders of the Broad Autism Phenotype Questionnaire. Journal of Autism and Developmental Disorders. 2015;45(10):3133-47.

348. Rosenberg RE, Law JK, Yenokyan G, McGready J, Kaufmann WE, Law PA. Characteristics and concordance of autism spectrum disorders among 277 twin pairs. Archives of Pediatrics and Adolescent Medicine. 2009;163(10):907-14.

349. Postorino V, Fatta LM, De Peppo L, Giovagnoli G, Armando M, Vicari S, et al. Longitudinal comparison between male and female preschool children with autism spectrum disorder. Journal of Autism and Developmental Disorders. 2015;45(7):2046-55.

350. Halladay AK, Bishop S, Constantino JN, Daniels AM, Koenig K, Palmer K, et al. Sex and gender differences in autism spectrum disorder: Summarizing evidence gaps and identifying emerging areas of priority. Molecular Autism. 2015;6(1).

351. Baio J. Prevalence of Autism Spectrum Disorder Among Children Aged 8 Years — Autism and Developmental Disabilities

Monitoring Network, 11 Sites, United States, 2010. Surveillance Summaries. 2014;63:1-21.

352. Singer L. Thoughts about sex and gender differences from the next generation of autism scientists Understanding the links between sex/gender and autism Dr Meng-Chuan Lai. Molecular Autism. 2015;6(1).

353. Bray I, Gunnell D, Smith GD. Advanced paternal age: How old is too old? Journal of Epidemiology and Community Health. 2006;60(10):851-3.

354. Cannon M. Contrasting effects of maternal and paternal age on offspring intelligence the clock ticks for men too. PLoS Medicine. 2009;6(3):0237-8.

355. Jacobsson B, Ladfors L, Milsom I. Advanced maternal age and adverse perinatal outcome. Obstetrics & Gynecology. 2004;104(4):727-33.

356. Cleary-Goldman J, Malone FD, Vidaver J, Ball RH, Nyberg DA, Comstock CH, et al. Impact of maternal age on obstetric outcome. Obstetrics & Gynecology. 2005;105(5, Part 1):983-90.

357. Delbaere I, Verstraelen H, Goetgeluk S, Martens G, De Backer G, Temmerman M. Pregnancy outcome in primiparae of advanced maternal age. European Journal of Obstetrics & Gynecology and Reproductive Biology. 2007;135(1):41-6.

358. Radin N. Primary-caregiving fathers in intact families. In: Gottfried AEaGAW, editor. Redefining families. New York: Springer; 1994. p. 11-54.

359. Fuligni ASaB-G, J. Measuring mother and father shared caregiving: an analysis using the Panel Study of Income Dynamics-Child Development Supplement. In: Day RaL, M., editor. Conceptualizing and measuring father involvement. Mahwah, NJ: Erlbaum; 2004. p. 341-57.

360. Saha S, Barnett AG, Foldi C, Burne TH, Eyles DW, Buka SL, et al. Advanced paternal age is associated with impaired neurocognitive outcomes during infancy and childhood. PLoS Medicine. 2009;6(3):0303-11.

361. Tearne JE. Older maternal age and child behavioral and cognitive outcomes: A review of the literature. Fertil Steril. 2015;103(6):1381-91.

362. Stein Z, Susser M. The risks of having children in later life. British Medical Journal. 2000;320(7251):1681-2.

363. Edwards RD, Roff J. Negative effects of paternal age on children's neurocognitive outcomes can be explained by maternal education and number of siblings. PloS one. 2010;5(9).

364.  Torres DD. Growth curve analyses of the relationship between early maternal age and children's mathematics and reading performance. Social Science Research. 2015;50:343-66.

365.  Coley RL, Chase-Lansdale PL. Adolescent pregnancy and parenthood: recent evidence and future directions. American Psychologist. 1998;53(2):152.

366.  Fall CHD, Sachdev HS, Osmond C, Restrepo-Mendez MC, Victora C, Martorell R, et al. Association between maternal age at childbirth and child and adult outcomes in the offspring: A prospective study in five low-income and middle-income countries (COHORTS collaboration). The Lancet Global Health. 2015;3(7):e366-e77.

367.  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(2):395-404.

368.  D'Onofrio BM, Goodnight JA, Van Hulle CA, Rodgers JL, Rathouz PJ, Waldman ID, et al. Maternal age at childbirth and offspring disruptive behaviors: testing the causal hypothesis. Journal of child psychology and psychiatry, and allied disciplines. 2009;50(8):1018-28.

369.  Saha S, Barnett AG, Buka SL, McGrath JJ. Maternal age and paternal age are associated with distinct childhood behavioural outcomes in a general population birth cohort. Schizophrenia research. 2009;115(2):130-5.

370.  Weiser M, Reichenberg A, Werbeloff N, Kleinhaus K, Lubin G, Shmushkevitch M, et al. Advanced parental age at birth is associated with poorer social functioning in adolescent males: Shedding light on a core symptom of schizophrenia and autism. Schizophrenia Bulletin. 2008;34(6):1042-6.

371.  Croen LA, Najjar DV, Fireman B, Grether JK. Maternal and paternal age and risk of autism spectrum disorders. Archives of pediatrics & adolescent medicine. 2007;161(4):334-40.

372.  Grether JK, Anderson MC, Croen LA, Smith D, Windham GC. Risk of autism and increasing maternal and paternal age in a large north American population. Am J Epidemiol. 2009;170

1118–26.

373.  Durkin MS, Maenner MJ, Newschaffer CJ, Lee L, Cunniff CM, Daniels JL, et al. Advanced parental age and the risk of autism spectrum disorder. Am J Epidemiol. 2008;168(11):1268-76.

374.  King MD, Fountain C, Dakhlallah D, Bearman PS. Estimated autism risk and older reproductive age. American Journal of Public Health. 2009;99(9):1673.

375.  Parner ET, Baron-Cohen S, Lauritsen MB, Jørgensen M, Schieve LA, Yeargin-Allsopp M, et al. Parental age and autism spectrum disorders. Annals of epidemiology. 2012;22(3):143-50.

376.  Sandin S, Hultman CM, Kolevzon A, Gross R, MacCabe JH, Reichenberg A. Advancing maternal age is associated with increasing risk for autism: a review and meta-analysis. Journal of the American Academy of Child & Adolescent Psychiatry. 2012;51(5):477-86. e1.

377.  Sasanfar R, Haddad SA, Tolouei A, Ghadami M, Yu D, Santangelo SL. Paternal age increases the risk for autism in an Iranian population sample. Molecular autism. 2010;1(1):1-10.

378.  Tsuchiya KJ, Matsumoto K, Miyachi T, Tsujii M, Nakamura K, Takagai S, et al. Paternal age at birth and high-functioning autistic-spectrum disorder in offspring. The British Journal of Psychiatry. 2008;193(4):316-21.

379.  Reichenberg A, Gross R, Weiser M, Bresnahan M, Silverman J, Harlap S, et al. Advancing paternal age and autism. Archives of general psychiatry. 2006;63(9):1026-32.

380.  Anderson JW, Johnstone BM, Remley DT. Breast-feeding and cognitive development: A meta-analysis. American Journal of Clinical Nutrition. 1999;70(4):525-35.

381.  Golding J, Rogers IS, Emmett PM. Association between breast feeding, child development and behaviour. Early Human Development. 1997;49(SUPPL.):S175-S84.

382.  Kramer MS, Aboud F, Mironova E, Vanilovich I, Platt RW, Matush L, et al. Breastfeeding and child cognitive development: new evidence from a large randomized trial. Archives of general psychiatry. 2008;65(5):578.

383.  Fonseca ALM, Albernaz EP, Kaufmann CC, Neves IH, De Figueiredo VLM. Impact of breastfeeding on the intelligence quotient of eight-year-old children. Jornal de Pediatria. 2013;89(4):346-53.

384.  Rodgers B. Feeding in infancy and later ability and attainment: A longitudinal study. Developmental Medicine and Child Neurology. 1978;20(4):421-6.

385.  Horwood LJ, Fergusson DM. Breastfeeding and later cognitive and academic outcomes. Pediatrics. 1998;101(1).

386.  Mortensen EL, Michaelsen KF, Sanders SA, Reinisch JM. The association between duration of breastfeeding and adult intelligence. Journal of the American Medical Association. 2002;287(18):2365-71.

387. Oddy WH, Kendall GE, Blair E, De Klerk NH, Stanley FJ, Landau LI, et al. Breast feeding and cognitive development in childhood: A prospective birth cohort study. Paediatric and Perinatal Epidemiology. 2003;17(1):81-90.

388. Gómez-Sanchiz M, Cañete R, Rodero I, Enrique Baeza J, Avila O. Influence of breast-feeding on mental and psychomotor development. Clinical Pediatrics. 2003;42(1):35-42.

389. Lawlor DA, Najman JM, Batty GD, O'Callaghan MJ, Williams GM, Bor W. Early life predictors of childhood intelligence: Findings from the Mater-University study of pregnancy and its outcomes. Paediatric and Perinatal Epidemiology. 2006;20(2):148-62.

390. Ruiz JMP, León MTM, Herreros JMP, Ibabe MCI. Breastfeeding and its influence into the cognitive process of Spanish school-children (6 years old), measured by the Wechsler intelligence scale. Archivos Latinoamericanos de Nutricion. 2013;63(3):218-23.

391. Slykerman RF, Thompson JMD, Becroft DMO, Robinson E, Pryor JE, Clark PM, et al. Breastfeeding and intelligence of preschool children. Acta Paediatrica, International Journal of Paediatrics. 2005;94(7):832-7.

392. Drane DL, Logemann JA. A critical evaluation of the evidence on the association between type of infant feeding and cognitive development. Paediatric and Perinatal Epidemiology. 2000;14(4):349-56.

393. Jain A, Concato J, Leventhal JM. How good is the evidence linking breastfeeding and intelligence? Pediatrics. 2002;109(6):1044-53.

394. Der G, Batty GD, Deary IJ. Effect of breast feeding on intelligence in children: Prospective study, sibling pairs analysis, and meta-analysis. British Medical Journal. 2006;333(7575):945-8.

395. Walfisch A, Sermer C, Cressman A, Koren G. Breast milk and cognitive development-the role of confounders: A systematic review. BMJ Open. 2013;3(8).

396. Sajjad A, Tharner A, Kiefte-de Jong JC, Jaddoe VVW, Hofman A, Verhulst FC, et al. Breastfeeding duration and non-verbal IQ in children. Journal of Epidemiology and Community Health. 2015;69:775.

397. Jacobson SW, Chiodo LM, Jacobson JL. Breastfeeding effects on intelligence quotient in 4- and 11-year-old children. Pediatrics. 1999;103(5):1-6.

398. Kanazawa S. Breastfeeding Is Positively Associated With Child Intelligence Even Net of Parental IQ. Developmental Psychology. 2015;51(12):1683-9.

399.   Bertini G, Perugi S, Dani C, Pezzati M, Tronchin M, Rubaltelli FF. Maternal education and the incidence and duration of breast feeding: a prospective study. Journal of pediatric gastroenterology and nutrition. 2003;37(4):447-52.

400.   Bartels M, Van Beijsterveldt C, Boomsma D. Breastfeeding, maternal education and cognitive function: a prospective study in twins. Behavior genetics. 2009;39(6):616-22.

401.   Skafida V. The relative importance of social class and maternal education for breast-feeding initiation. Public health nutrition. 2009;12(12):2285-92.

402.   Leventakou V, Roumeliotaki T, Koutra K, Vassilaki M, Mantzouranis E, Bitsios P, et al. Breastfeeding duration and cognitive, language and motor development at 18 months of age: Rhea mother-child cohort in Crete, Greece. Journal of Epidemiology and Community Health. 2015;69(3):232-9.

403.   Dee DL, Li R, Lee LC, Grummer-Strawn LM. Associations between breastfeeding practices and young children's language and motor skill development. Pediatrics. 2007;119(SUPPL. 1):S92-S8.

404.   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(7):635-42.

405.   McCann JC, Ames BN. Is docosahexaenoic acid, an n-3 long-chain polyunsaturated fatty acid, required for development of normal brain function? An overview of evidence from cognitive and behavioral tests in humans and animals. American Journal of Clinical Nutrition. 2005;82(2):281-95.

406.   Herba CM, Roza S, Govaert P, Hofman A, Jaddoe V, Verhulst FC, et al. Breastfeeding and early brain development: The Generation R study. Maternal and Child Nutrition. 2013;9(3):332-49.

407.   Tawia S. Breastfeeding, brain structure and function, cognitive development and educational attainment. Breastfeeding Review. 2013;21(3):15-20.

408.   Ou X, Andres A, Cleves MA, Pivik RT, Snow JH, Ding Z, et al. Sex-specific association between infant diet and white matter integrity in 8-y-old children. Pediatric Research. 2014;76(6):535-43.

409.   Nagel BJ, Bathula D, Herting M, Schmitt C, Kroenke CD, Fair D, et al. Altered white matter microstructure in children with attention-deficit/ hyperactivity disorder. Journal of the American Academy of Child and Adolescent Psychiatry. 2011;50(3):283-92.

410.   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.

411. Hong L, Ziegler J, Rebecca B. Breastfeeding and autism spectrum disorders. Topics in Clinical Nutrition. 2014;29(3):278-85.

412. Schultz ST, Klonoff-Cohen HS, Wingard DL, Akshoomoff NA, Macera CA, Ji M, et al. Breastfeeding, infant formula supplementation, and Autistic Disorder: The results of a parent survey. International Breastfeeding Journal. 2006;16:1-7.

413. Al-Farsi YM, Al-Sharbati MM, Waly MI, Al-Farsi OA, Al-Shafaee MA, Al-Khaduri MM, et al. Effect of suboptimal breast-feeding on occurrence of autism: A case-control study. Nutrition. 2012;28(7-8):e27-e32.

414. Steinman G, Mankuta D. Breastfeeding as a possible deterrent to autism - A clinical perspective. Medical hypotheses. 2013;81(6):999-1001.

415. Hamlyn B, Brooker, S., Oleinikova, K. and Wands, S. Infant Feeding 2000. London: TSO; 2000.

416. Renfrew MJ, Pokhrel, S., Quigley, M. Preventing Disease and Saving Resources: The Potential Contribution of Increasing Breastfeeding Rates in the UK 2012 [Available from: http://www.unicef.org.uk/Documents/Baby_Friendly/Research/Preventing_disease_saving_resources.pdf.

417. Wales NAf. Investing in a Better Start: Promoting Breastfeeding in Wales 2001 [Available from: http://www.wales.nhs.uk/publications/bfeedingstrategy-e.pdf.

418. Bolling KG, C. Hamlyn, B. & Thornton, A. . Infant Feeding Survey 2005. London: The Information Centre for Health and Social Care; 2007.

419. Kamphaus RW, Dresden, J. and Kaufman, A. S. Clinical and psychometric considerations in the assessment of preschool children. In: Willis DJaC, J. L. , editor. Testing young children. Austin, TX: Pro Ed. Allyn & Bacon; 1993.

420. Akhtar N, Menjivar JA. Cognitive and Linguistic Correlates of Early Exposure to More than One Language. Advances in Child Development and Behavior. 2012;42:41-78.

421. Bialystok E, Luk G, Peets KF, Yang S. Receptive vocabulary differences in monolingual and bilingual children. Bilingualism. 2010;13(4):525-31.

422. Oller DK, Eilers RE. Language and Literacy in Bilingual Children. Wei L, editor. Cleavedon, UK: Mulitlingual Matters Ltd; 2002.

423. Bialystok E, Martin MM. Attention and inhibition in bilingual children: Evidence from the dimensional change card sort task. Developmental science. 2004;7(3):325-39.

424. Carlson SM, Meltzoff AN. Paper: Bilingual experience and executive functioning in young children. Developmental science. 2008;11(2):282-98.

425. Martin-Rhee MM, Bialystok E. The development of two types of inhibitory control in monolingual and bilingual children. Bilingualism. 2008;11(1):81-93.

426. Poulin-Dubois D, Blaye A, Coutya J, Bialystok E. The effects of bilingualism on toddlers' executive functioning. Journal of Experimental Child Psychology. 2011;108(3):567-79.

427. Yang S, Yang H, Lust B. Early childhood bilingualism leads to advances in executive attention: Dissociating culture and language. Bilingualism. 2011;14(3):412-22.

428. Bialystok E, Majumder S. The relationship between bilingualism and the development of cognitive processes in problem solving. Applied Psycholinguistics. 1998;19(1):69-85.

429. Lauchlan F, Parisi M, Fadda R. Bilingualism in Sardinia and Scotland: Exploring the cognitive benefits of speaking a 'minority' language. International Journal of Bilingualism. 2013;17(1):43-56.

430. Bialystok E, Craik FIM, Green DW, Gollan TH. Bilingual minds. Psychological Science in the Public Interest, Supplement. 2009;10(3):89-129.

431. Ye Z, Zhou X. Executive control in language processing. Neuroscience and Biobehavioral Reviews. 2009;33(8):1168-77.

432. Kroll JF, Dussias PE, Bogulski CA, Kroff JRV. Juggling two languages in one mind. What bilinguals tell us about language processing and its consequences for cognition. In: Ross BH, editor. Psychology of Learning and Motivation - Advances in Research and Theory. 56. London: Elsevier; 2012. p. 229-62.

433. Calvo A, Bialystok E. Independent effects of bilingualism and socioeconomic status on language ability and executive functioning. Cognition. 2014;130(3):278-88.

434. Bialystok E. Bilingualism: The good, the bad, and the indifferent. Bilingualism. 2009;12(1):3-11.

435. Korkman M, Stenroos M, Mickos A, Westman M, Ekholm P, Byring R. Does simultaneous bilingualism aggravate children's specific language problems? Acta Paediatrica, International Journal of Paediatrics. 2012;101(9):946-52.

436.  Garratt LC, Kelly TP. To what extent does bilingualism affect children's performance on the NEPSY? Child Neuropsychology. 2008;14(1):71-81.

437.  Karlsson LC, Soveri A, Räsänen P, Kärnä A, Delatte S, Lagerström E, et al. Bilingualism and performance on two widely used developmental neuropsychological test batteries. PloS one. 2015;10(4).

438.  Brooks-Gunn J, Duncan GJ. The effects of poverty on children. Future of Children. 1997;7(2):55-71.

439.  Bradley RH, Corwyn RF. Socioeconomic status and child development. Annual Review of Psychology2002. p. 371-99.

440.  Heckman JJ. Skill formation and the economics of investing in disadvantaged children. Science. 2006;312(5782):1900-2.

441.  von Stumm S, Plomin R. Socioeconomic status and the growth of intelligence from infancy through adolescence. Intelligence. 2015;48:30-6.

442.  von Stumm S. You are what you eat? Meal type, socio-economic status and cognitive ability in childhood. Intelligence. 2012;40(6):576-83.

443.  Tucker-Drob EM. Preschools reduce early academic-achievement gaps: A longitudinal twin approach. Psychological Science. 2012;23(3):310-9.

444.  Huston AC. Childcare for low-income families: Problems and promises. In: Crouter ACaB, A., editor. Work-Family Challenges for Low-Income Parents and Their Children. London: Psychology Press; 2004. p. 139-64.

445.  Meyers M, Rosenbaum D, Ruhm C, Waldfogel J. Inequality in early childhood education and care: What do we know? In: Neckerman KM, editor. Social Inequality. New York, US: Russell Sage Foundation; 2004. p. 223-69.

446.  White KR. The relation between socioeconomic status and academic achievement. Psychological Bulletin. 1982;91(3):461-81.

447.  Carneiro P, Heckman JJ. Human capital policy. Inequality in America: What Role for Human Capital Policies? 2003:77-239.

448.  Smith JR, Brooks-Gunn J, Klebanov P. Consequences of Living in Poverty for Young Children's Cognitive and Verbal Ability and Early School Achievement. In: Duncan GJ, Brooks-Gunn, J., editor. Consequences of growing up poor. New York: Russell Sage Foundation; 1997. p. 132-89.

449.  Middeldorp CM, Lamb DJ, Vink JM, Bartels M, Van Beijsterveldt CEM, Boomsma DI. Child care, socio-economic status and problem

behavior: A study of gene-environment interaction in young Dutch twins. Behavior Genetics. 2014;44(4):314-25.

450. 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.

451. Achenbach TM, Bird HR, Canino G, Phares V, Gould MS, Rubio-Stipec M. Epidemiological comparisons of puerto rican and U.S. Mainland children: Parent, teacher, and self-reports. Journal of the American Academy of Child and Adolescent Psychiatry. 1990;29(1):84-93.

452. Duncan GJ, Brooks-Gunn J, Klebanov PK. Economic deprivation and early childhood development. Child development. 1994;65(2 Spec No):296-318.

453. McLeod JD, Shanahan MJ. Poverty, Parenting, and Children's Mental Health. American Sociological Review. 1993;58(3):351-66.

454. Law J, Rush R, Clegg J, Peters T, Roulstone S. The role of pragmatics in mediating the relationship between social disadvantage and adolescent behavior. Journal of Developmental and Behavioral Pediatrics. 2015;36(5):389-98.

455. Quinto Romani A. Parental Socioeconomic Background and Child Behaviour. Social Indicators Research. 2014;116(1):295-306.

456. Schroeders U, Schipolowski S, Wilhelm O. Age-related changes in the mean and covariance structure of fluid and crystallized intelligence in childhood and adolescence. Intelligence. 2015;48:15-29.

457. Deary IJ. The Stability of Intelligence From Childhood to Old Age. Current Directions in Psychological Science. 2014;23(4):239-45.

458. Horn JL, Noll J. Human cognitive capabilities: Gf-Gc theory. In: Flanagan DP, Genshaft, J. L., and Harrison, P. L., editor. Contemporary Intellectual Assessment: Theories, Tests, and Issues. New York, US: Guildford Press; 1997. p. 53-91.

459. Horn JL. Spearman, g, expertise, and the nature of human cognitive capability. In: Kyllonen PC, Roberts, R.D., Stankov, L., editor. Extending intelligence: Enhancement and new constructs. Oxon: Lawrence Erlbaum Associates; 2008. p. 185-230.

460. Baltes PB. On the Incomplete Architecture of Human Ontogeny: Selection, Optimization, and Compensation as Foundation of Developmental Theory. American Psychologist. 1997;52(4):366-80.

461. Salthouse T. Major Issues in Cognitive Aging. New York, US: Oxford University Press; 2010. 1-256 p.

462. Garrett HE. A developmental theory of intelligence. The American psychologist. 1946;1(9):372-8.

463. Tideman E, Gustafsson JE. Age-related differentiation of cognitive abilities in ages 3-7. Personality and Individual Differences. 2004;36(8):1965-74.

464. Hülür G, Wilhelm O, Robitzsch A. Intelligence Differentiation in Early Childhood. Journal of Individual Differences. 2011;32(3):170-9.

465. Gignac GE. Dynamic mutualism versus g factor theory: An empirical test. Intelligence. 2014;42(1):89-97.

466. Juan-Espinosa M, García LF, Colom R, Abad FJ. Testing the age related differentiation hypothesis through the Wechsler's scales. Personality and Individual Differences. 2000;29(6):1069-75.

467. Trahan LH, Stuebing KK, Fletcher JM, Hiscock M. The Flynn effect: a meta-analysis. Psychological bulletin. 2014;140(5):1332-60.

468. Beaujean A, Sheng Y. Assessing the flynn effect in the wechsler scales. Journal of Individual Differences. 2014;35(2):63-78.

469. Dickens WT, Flynn JR. Heritability estimates versus large environmental effects: the IQ paradox resolved. Psychological review. 2001;108(2):346.

470. Teasdale TW, Owen DR. A long-term rise and recent decline in intelligence test performance: The Flynn Effect in reverse. Personality and Individual Differences. 2005;39(4):837-43.

471. Rodgers JL, Wänström L. Identification of a Flynn Effect in the NLSY: Moving from the center to the boundaries. Intelligence. 2007;35(2):187-96.

472. Mingroni MA. Resolving the IQ paradox: heterosis as a cause of the Flynn effect and other trends. Psychological Review. 2007;114(3):806.

473. Woodley MA. Heterosis doesn't cause the Flynn effect: a critical examination of Mingroni (2007). Psychological Review. 2011;118(4):689-93.

474. Teasdale TW, Owen DR. Secular declines in cognitive test scores: A reversal of the Flynn Effect. Intelligence. 2008;36(2):121-6.

475. Dutton E, Lynn R. A negative flynn effect in Finland, 1997-2009. Intelligence. 2014;41(6):817-20.

476. Dutton E, Lynn R. A negative Flynn Effect in France, 1999 to 2008-9. Intelligence. 2015;51.

477. Shayer M, Ginsburg D, Coe R. Thirty years on–a large anti-Flynn effect? The Piagetian test Volume & Heaviness norms 1975–2003. British Journal of Educational Psychology. 2007;77(1):25-41.

478. Reynolds CR, Milam DA. Challenging intellectual testing results. Faust D, editor. New York, NY: . Oxford University

Press; 2012. 311-34 p.

479. Schmidt FL, Le H, Ilies R. Beyond Alpha: An Empirical Examination of the Effects of Different Sources of Measurement Error on Reliability Estimates for Measures of Individual Differences Constructs. Psychological Methods. 2003;8(2):206-24.

480. Watkins MW, Smith LG. Long-term stability of the wechsler intelligence scale for children-fourth edition. Psychological Assessment. 2013;25(2):477-83.

481. Reschly DJ, Bergstrom MK. Response to intervention. The Handbook of School Psychology. 2009:434-60.

482. Reeve CL, Bonaccio S. On the myth and the reality of the temporal validity degradation of general mental ability test scores. Intelligence. 2011;39(5):255-72.

483. Johnson W, Gow AJ, Corley J, Starr JM, Deary IJ. Location in cognitive and residential space at age 70 reflects a lifelong trait over parental and environmental circumstances: The Lothian Birth Cohort 1936. Intelligence. 2010;38(4):402-11.

484. Van Der Maas HLJ, Dolan CV, Grasman RPPP, Wicherts JM, Huizenga HM, Raijmakers MEJ. A dynamical model of general intelligence: The positive manifold of intelligence by mutualism. Psychological Review. 2006;113(4):842-61.

485. Franić S, Dolan CV, Van Beijsterveldt CEM, Pol HEH, Bartels M, Boomsma DI. Genetic and environmental stability of intelligence in childhood and adolescence. Twin Research and Human Genetics. 2014;17(3):151-63.

486. Honzik MP, MacFarlane, J. W., & Allen, L. The stability of mental test performance between two and

eighteen years. Journal of Experimental Education. 1948;17:309-24.

487. Anderson LD. The predictive value of infant tests in relation to intelligence at 5 years. Child Development. 1939;10:202-12.

488. Yang P, Jong YJ, Hsu HY, Lung FW. Role of assessment tests in the stability of intelligence scoring of pre-school children with uneven/delayed cognitive profile. Journal of Intellectual Disability Research. 2011;55(5):453-61.

489. Deary IJ, Whalley LJ, Lemmon H, Crawford JR, Starr JM. The stability of individual differences in mental ability from childhood to old age: Follow-up of the 1932 Scottish mental survey. Intelligence. 2000;28(1):49-55.

490. Deary IJ, Pattie A, Starr JM. The Stability of Intelligence From Age 11 to Age 90 Years: The Lothian Birth Cohort of 1921. Psychological Science. 2013;24(12):2361-8.

491. Mortensen EL, Andresen J, Kruuse E, Sanders SA, Reinisch JM. IQ stability: The relation between child and young adult intelligence test scores in low-birthweight samples. Scandinavian Journal of Psychology. 2003;44(4):395-8.

492. Breslau N, Chilcoat HD, Susser ES, Matte T, Liang KY, Peterson EL. Stability and change in children's intelligence quotient scores: A comparison of two socioeconomically disparate communities. Am J Epidemiol. 2001;154(8):711-7.

493. Canivez GL, Watkins MW. Long-term stability of the wechsler intelligence scale for children-third edition among demographic subgroups: Gender, race/ethnicity, and age. Journal of Psychoeducational Assessment. 1999;17(4):300-13.

494. Cherny SS, Fulker DW, Hewitt JK. Cognitive development from infancy to middle childhood. In: RJ Sternberg EG, editor. Intelligence, heredity and environment. Cambridge: Cambridge University Press; 1997. p. 463-82.

495. Furby L. Interpreting regression toward the mean in developmental research. Developmental Psychology. 1973;8(2):172-9.

496. Basso MR, Bornstein RA, Roper BL, McCoy VL. Limited accuracy of premorbid intelligence estimators: A demonstration of regression to the mean. Clinical Neuropsychologist. 2000;14(3):325-40.

497. Veiela HOF, Koopman RF. The bias in regression-based indices of premorbid IQ. Psychological Assessment. 2001;13(3):356-68.

498. Berg JA, Mayor GH. A study in normal human volunteers to compare the rate and extent of levothyroxine absorption from Synthroid® and Levoxine®. The Journal of Clinical Pharmacology. 1992;32(12):1135-40.

499. Vita R, Saraceno G, Trimarchi F, Benvenga S. A novel formulation of l-thyroxine (l-T4) reduces the problem of l-T4 malabsorption by coffee observed with traditional tablet formulations. Endocrine. 2013;43(1):154-60.

500. Benvenga S, Bartolone L, Pappalardo MA, Russo A, Lapa D, Giorgianni G, et al. Altered intestinal absorption of L-thyroxine caused by coffee. Thyroid : official journal of the American Thyroid Association. 2008;18(3):293-301.

501. Bernal J. Thyroid hormones and brain development. Vitamins & Hormones. 2005;71:95-122.

502. Vose LR, Vinukonda G, Jo S, Miry O, Diamond D, Korumilli R, et al. Treatment with thyroxine restores myelination and clinical recovery after intraventricular hemorrhage. The Journal of Neuroscience. 2013;33(44):17232-46.

# Appendix

# Appendix 1: Timeline of the CATS project

# Timeline of the CATS project



Timeline (2002–2015) with the following events:

1. Women recruited during their pregnancy (mean 12 weeks gestation)

2. Offspring born

3. Cognitive tests (IQ) for offspring (mean age 3 years)

4. CATS I paper published (Lazarus et al NEJM)

5. CATS II recruitment (data collected: DXA scans, DNA, behavioural questionnaires)

6. CATS II cognitive assessments (IQ and additional tests)

KEY
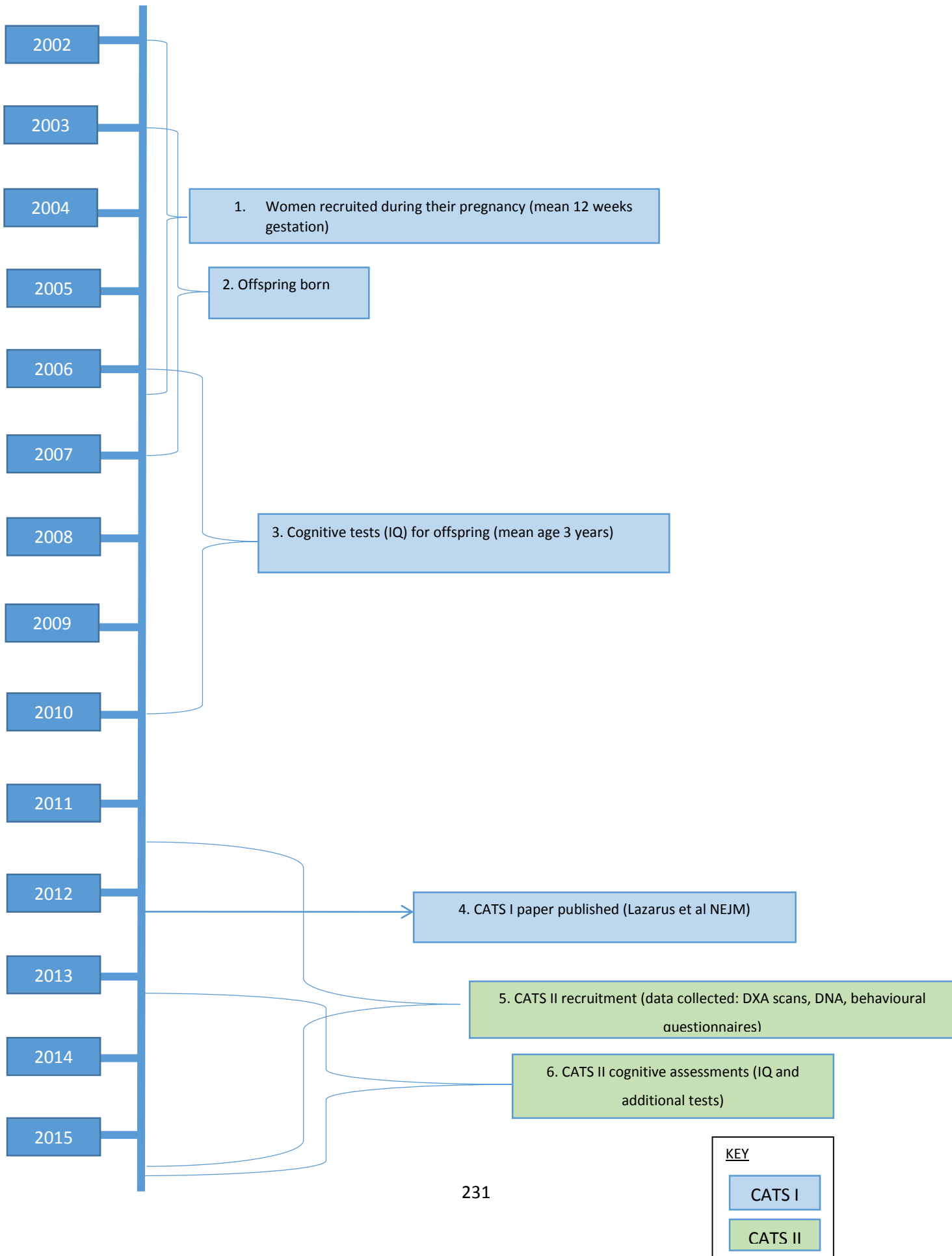
CATS I

CATS II

Appendix 2: The decision process for additional tests for CATS II

**Summary of the decision process for the additional tests to be used in CATS II**

Literature has suggested that as well as IQ impairments for children who have been born to mothers with an underactive thyroid, there may be further deficits that require testing. This document is a summary of how the CATS II research team came to decide which tests to administer, alongside the WISC-IV.

**Memory Tests**

1) Memory for Patterns Delayed
2) Narrative Recall

A memory deficit has been proposed by Willoughby et al. (2013), Gilbert et al. (2012) and Riva and Naranjo (2007). Willoughby et al. identified a difficulty with episodic memory which resulted from hippocampal damage to the fetus from a lack of thyroxin. As episodic memory is one of the long term memory systems, choosing a delayed memory task seemed highly appropriate. 'Narrative Recall' does not investigate possible deficits in long term memory per se, but will examine memory for organised verbal material. Along with the working memory tests from the WISC, this recall task could provide further evidence for any results found.

**Motor Test**

3) Fingertip Tapping

Li et al. (2010) and Haddow et al. (1999) investigated MHT and the effects it may have on the developing child; their results displayed evidence for a fine motor deficit. Li et al. and Zoeller and Rovet (2004) identified a gross motor deficit, however within the CATS II protocol this would be difficult to investigate. The 'Fingertip Tapping' test is quick and easy to administer and was chosen as the fine motor assessment as it could be easily added onto a WISC-IV assessment.

**Deficits that are identified in the literature, but are not to be investigated in CATS II**

- Visual Perception

Visual perception deficits of children who experience MHT are sparsely reported upon. Haddow et al. (2009) did identify this as a problem in such children, and Zoeller and Rovet (2004) have reported upon this difficulty. However, such a deficit has been disproved by Oken et al. (2009) who tested children at age three. As the WISC-IV and additional subtests from the NEPSY could take up to 90 minutes to administer, following through on this line of enquiry was discounted.

- Hearing

The literature has suggested hearing difficulties to be a possible line of enquiry (Haddow et al., 1999; Crofton, 2004). Tests of phonological and auditory awareness were examined but none were suitable enough; age restrictions and time to administer were some of the problems encountered with the proposed hearing tests (see below). On these grounds, investigating hearing as a deficit was cut.

- PHaB – Phonological Assessment Battery
  - 30-40 minutes to administer
- CTOPP – Comprehensive Test of Phonological Processing
  - 30 minutes to administer
- PAT - Phonological Abilities Test
  - 30-40 minutes to administer
- Test of Phonological Awareness
  - Tests only up to age 8
- TAAS – Test of Auditory Analysis Skills
  - Tests only up to age 8
- Language difficulties

Haddow et al. (1999) identified this as a deficit and this will be looked at, but only so far as looking at differences between verbal and perceptual reasoning IQs on the WISC-IV. There is not enough time available to administer separate tests to thoroughly investigate a possible language difficulty.

**References**

Crofton, K. M. (2004). Developmental disruption of thyroid hormone: correlations with hearing dysfunction in rats. *Risk Analysis*, 24 (6), 1665-167.

Gilbert, M. E., Rovet, J., Chen, Z. & Koibuchi, N. (2012). Developmental thyroid hormone disruption: prevalence environmental contaminats and neurodevelopmental consequences. *NeuroToxicology*, 33, 842-852.

Haddow, J. E., Palomaki, G. E., Walkter, C. A., Williams, J. K., Knight, G. J., Gagnon, j., O'Heir, C. E., Mitche;;, M. L., Hermos, R. J., Waisbren, S. E., Faix, J. D. & Klein, K. Z. (1999). Maternal thyroid deficiency during pregnancy and subsequent neuropsychological development of the child. *The New England Journal of Medicine*, 341 (8), 549-555.

Oken, E., Braverman, L. E., Platek, D., Mitche;;, M. L., Lee, S. L. & Pearce, E. N. (2009). Neonatal thyroxine, maternal thyroid function, and child cognition. *Journal of Clinical Endocrinological Metabolism*, 94 (2), 497-503.

Rivas, M. & Naranjo, J. R. (2007). Thyroid hormones, learning and memory. *Genes, Brain and Behaviour*, 6 (1), 40-44.

Willoughby, K. A., McAndrews, M. P. & Rovet, J. (2013). Effects of early thyroid hormone deficiency on children's autobiographical memory performance. *Journal of International Neuropsychological Society*, 19, 1-11.

Li, Y., shaun, Z., Tengt, W., Xiaohui, Y., Lit, Y., Fan, C., Teng, X., Guot, R., Wangt, H., Jai, L., Yanyan, C., WeiWei, W., Masauso, C., Zhang, L., Lang, L., Zhao, Y. & Hua, T. (2010). Abnormalities of maternal thyroid function during pregnancy affect neuropsychological development of their children at 25-30 months. *Clinical Endocrinology*, 72, 825-829.

Zoeller, R. T. & Rovett, J. (2004). Timing of thyroid hormone action in the developing brain: clinical observations and experimental findings. *Journal of Neuroendocrinology*, 16, 809-818.

## Appendix 3: CATS II General Questionnaire

DATE:_____

| MOTHER - GENERAL | | | | | | | |
|---|---|---|---|---|---|---|---|
| Name | | | | | | | |
| DOB | | | | | | | |
| No. of pregnancies | | | | | | | |
| No. of live children (with ages) | | | | | | | |
| Child's Paternal height | | | | | | | |
| Occupation (or details of last job) | | | | | | | |
| Child's Paternal Occupation | | | | | | | |
| Cigarette smoker (per day) pls tick | Never | | Present 0-10 | | 10-20 | | 20+ | |
| Ethnicity pls tick | White | | Asian | | African | | Other | |

| MEDICAL | |
|---|---|
| Previous illnesses/pls give date: | |
| Operations/pls give date: | |
| Family History of Thyroid Disease? | Yes/No  (pls delete as appropriate) |
| Previous Thyroid Hormone Treatment/pls give date(s): | |
| Current Thyroid Hormone Treatment | Yes/No    Dose: |

| Other current drug therapy | Name | Dose | Date started |
|---|---|---|---|
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |

| Other current illnesses | |
|---|---|
| | |
| | |
| | |
| OTHER COMMENTS: | |

| CHILD | | | | | | |
|---|---|---|---|---|---|---|
| DOB | | | M/F | | | |
| Birth Weight | | | | | | |
| Child's Handedness | Right | | Left | | | |
| Language of School | English | | Welsh | | Other | |
| Language Spoken at Home | English | | Welsh | | Other | |
| Gestational age | Full term (37-40 wks | | Preterm (32-37 wks) | | Very Preterm (<32 weeks) | |
| Mode of Delivery | Normal | | Caesarean | | | |
| Medical complications in pregnancy | Diabetes | | | | Yes/No | |
| | High BP | | | | Yes/No | |
| | Pre eclampsia | | | | Yes/No | |
| | Anaemia | | | | Yes/No | |
| | Other: | | | | | |
| Breast fed for more than 1 mth | Yes | | No | | | |

## Appendix 4: Initial contact pack

1. Cover letter
2. Information for the mother
3. Information for the child
4. Response form

1. **Cover letter**

**Dear «returned_forename» «returned_surname»**

**Between 2004 and 2006 when you were attending antenatal clinic you may remember agreeing to take part in a study called CATS (Controlled Antenatal Thyroid Screening) which was a research project looking at the effects of maternal thyroid functioning in pregnancy: At that time you kindly agreed to an extra blood test, then three years later you may have been part of a much smaller group of women who were contacted and their children took part in an assessment at their home.**

**Many thanks to each and every one of you who agreed to take part in CATS which was the largest study of its type in the world at the time, and which is now complete and has been published.**
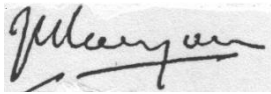
**We would like to invite you to join CATS II, a study led by Professor Marian Ludgate, following up on the progress of the mothers and children from CATS.**

**Enclosed is a Patient Information Leaflet for CATS II which we hope explains the study clearly and in more detail. We have also enclosed a response form for you to complete and return to us in the stamped addressed envelope when you have made your decision. You are certainly not obliged to take part.**

**As you can imagine many people will have changed address since 2004 – we have been working with the Welsh Demographic Service and the Patient Data registry to update our information, as approved by South East Wales Research Ethics Committee and the local Health Boards. We hope we have the correct address for you but if not please let us know.**

**We look forward to receiving your response form and please get in touch if you have any questions.**

**Yours sincerely,**

**Professor John Lazarus.**

## 2. Information for the mother

PART 1

### 1. Title of study

Controlled Antenatal Thyroid Screening Study II (CATS 2)

## 2. Introduction

**We would like to invite you to take part in our research study. Before you decide we would like you to understand why the research is being done and what it would involve for you. Talk to others about the study if you wish.**

**(Part 1 tells you the purpose of this study and what will happen to you if you take part.**

**Part 2 gives you more detailed information about the conduct of the study).**

**Please ask us if there is anything that is not clear.**

*3. What is the purpose of this study?*

Following on from your participation in the original CATS study, scientists and doctors at the University Hospital of Wales are hoping to arrange further studies to investigate how thyroid hormone levels during pregnancy may affect the development of the proportions of fat, muscle and general development in a child. They will assess your child's abilities in a range of areas, such as verbal and non-verbal skills, memory and speed of processing with an IQ test. They will also investigate whether there is any effect on fat and bone development in the mother and/or any influence on their general wellbeing.

*4. Why have I been invited?*

It is only those involved in the previous study that will be able to help us with our ongoing investigations.

We really appreciated your involvement in the CATS 1 study on thyroid hormone levels in pregnancy, and how variations in this might affect your child. The results of this study should be available later this year (we are planning a meeting to let you know the results) and will prove useful in understanding these effects and help to improve the care of pregnant women and their children.

*5. Do I have to take part?*

It is up to you to decide whether or not to take part. If you are interested in continuing your involvement with the CATS study by taking part in CATS II, there are two different levels of involvement detailed below. It is up to you which you would like to be involved in. If you choose not to take part, that is fine by us, you do not have to give us reasons why and this decision will not have any implications for your future medical care.

If you decide to take part you will be asked to sign a consent form. You are still free to withdraw at any time and without giving a reason. A decision to withdraw at any time, or a decision not to take part, will not affect the standard of care you receive.

*6. What will happen to me if I take part?*

The study will involve either:

1) Completing questionnaires about your child's general development, your mood and obtaining saliva or mouth swabs and urine samples from you and/or your child for genetic studies. The saliva or mouth swabs will provide us with a sample of DNA which will be analysed for variations in the genes which influence e.g. control of thyroid function or mood disorders. These will need to be returned to us in an envelope provided, along with the completed consent forms confirming you are happy for us to analyze these questionnaires and samples.

   Or,

2) We would like to invite you to attend an appointment at the University Hospital of Wales for a morning to:

   1. Measure height and weight in you and your child.
   2. Measure blood pressure and assess the blood flow in the blood vessels of you and your child's arm with a probe put on the skin in you and your child.
   3. Complete questionnaires with you about your mood and your child's social and educational development and progress.
   4. Perform a DEXA scan to measure the proportion of fat and muscle in you (providing you are not pregnant at the time) and your child (this would involve lying still on the scan table for a minute or two whilst the scan is performed and does not hurt). DEXA is a simple, rapid and non-invasive technique which is used routinely in clinical practice to measure bone density. This scan does involve exposure to a small amount of radiation (no more than experienced during a day walking around outside in the UK) but the risk is negligible.
   5. Take a blood sample from you and your child (approximately 6 teaspoonfuls), to measure thyroid hormone levels, aspects of the control of bone and fat tissues and variations in the genes

241

which influence the different aspects that we will measure in the study. Some of the tests need for you and your child to have fasted since the night before; we will provide breakfast as soon as possible after you arrive. If you or your child prefers not to have a blood test we would still like to take saliva or mouth swabs for genetic studies.

6. For your child to undertake an intelligence test called Weschler Intelligence Scale for Children (Fourth UK Edition (WISC-IV[uk]). and sections from the Neuropsychological Assessment Second Edition. This usually takes about 1-1½ hours and would generate an IQ score, which would represent your child's intelligence. You can of course be present during the test, but some children do not perform as well if their parents are there as they are distracted. You will be able to have a copy of the report if you wish. If you prefer we could arrange for our research assistant to make an appointment for this test to be done in your home.

7. Take a urine sample from you and your child.

## 7. Expenses and payments

If you wish we will be able to reimburse any travelling expenses / car parking fees incurred while attending for the study visit.

## 8. What do I have to do?

If you are interested in taking part or you would like more information about the study, please complete the response form by putting your initials in the appropriate box and we will be in touch with you very soon.

If you have any hesitation and would like more information, you may contact Professor John Lazarus (phone: 02920 742938 email lazarus@cf.ac.uk). Professor Lazarus set up the original CATS study in which you were involved but is now retired. However, he has kindly agreed to act as an advisor with expert understanding of the issues involved in our proposed study and would be happy to provide you with any advice you might require.

## 9. What are the alternatives for diagnosis or treatment?

At present, the thyroid function of women who are pregnant is not tested at all. Our studies will help to decide whether it is advisable for this to be done, either because of benefits to the child, the mother or both.

## 10. What are the possible disadvantages and risks of taking part?

Other than possible discomfort (temporary pain, swelling, bruising and rarely infection) caused by the collection of blood, no other side effects are anticipated from the study procedures. A DEXA scan does involve exposure to a small amount of radiation but the risk is negligible. It is possible that the blood tests or DEXA scan could by chance pick up an unsuspected abnormality, in which case you will be given an opportunity to discuss these findings further with the doctors.

These investigations will only be performed once we have explained in detail what will be involved, you have given us your consent and your child has agreed to take part. You or your child may decide to undergo some but not all of the tests. It is possible that you decide to participate but your child prefers not too, or vice versa.

*11. What are the side effects of any treatment received when taking part?*

The new study (CATS II) does not involve you taking any medication or receiving any treatment. In CATS I, you may have received tablets containing thyroid hormone which is exactly the same as produced naturally by your thyroid and so does not have any side effects.

*12. Exposure to radiation or ionising radiation*

As mentioned above, you and/or your child may agree to have a DEXA scan. This involves exposure to a small amount of radiation but no more than experienced walking around outside for 1 day in the UK.

*13. Harm to the unborn child*

None of the investigations to be performed have any negative effect on an unborn child EXCEPT the DEXA scan. With your permission, we will perform a pregnancy test and if you it is positive you will not receive the DEXA scan. None of the investigations will affect you or your child if you are breast feeding.

*14. What are the potential benefits of taking part?*

We cannot promise the study will help you but the information we get from this study will help improve the future care of pregnant women with consequent benefit to their children.

*15. What happens when the research study stops?*

Since no treatment is involved, there are no special considerations to take into account. However, with your permission, we will store samples and this is explained in part 2.

*16. What if something goes wrong?*

Any complaint about the way you have been dealt with during the study or any possible harm you might suffer will be addressed. The detailed information on this is given in Part 2.

*17. Will my taking part in this study be kept confidential*

Yes. We will follow ethical and legal practice and all information about you will be handled in confidence. The details are included in Part 2.

**This completes Part 1.**

**If the information in Part 1 has interested you and you are considering participation, please read the additional information in Part 2 before making any decision.**

PART 2

*1. What if relevant new information becomes available?*

Sometimes we get new information about treatments, in this case we are measuring the effects of treatment you may have received when you were pregnant and not during this new study.

*2. What will happen if I don't want to carry on with the study?*

If you withdraw from the study, with your permission, we will keep any samples, and use any data collected up to your withdrawal. A decision to withdraw at any time will not affect the standard of care you receive.

Similarly in the event of your loss of capacity, with your permission, we will keep any samples, and use any data collected prior to this.

*3. What if there is a problem?*

If you have a concern about any aspect of this study, you should ask to speak to the researchers who will do their best to answer your questions (phone 02920 745457 or 02920 745002). This study is being indemnified by Cardiff University. If you are harmed by taking part in this research project, there are no special compensation arrangements. If you are harmed due to someone's negligence, then you may have grounds for a legal action but you may have to pay for it. Regardless of this, if you wish to complain, or have any concerns about any aspect of the way you have been approached or treated during the course of

this study, the normal National Health Service complaints mechanisms should be available to you.

*4. Will my taking part in this study be kept confidential?*

All information which is collected about you during the course of the research will be kept strictly confidential. No personally identifiable information will be held on NHS or University computers although storage of study data on both systems will occur in an anonymised form. These computers are held by Professor Ludgate, Dr Rees and other members of the study team and are stored in locked rooms (rooms 254 & 261, 2nd Floor C block, University Hospital of Wales). The computers are password protected and no identifiable data will be transferred electronically, though final analysis of the anonymised study data may require email transfer to Professor Newcombe in the Department of Primary Care and Public Health who may provide statistical advice.

*5. Involvement of your general practitioner/family doctor*

With your permission your GP will be informed of your participation in this study.

*6. What will happen to any samples I give?*

The blood and urine samples (6 teaspoons) for this study will be collected and stored securely for later analysis in the Centre for Endocrine and Diabetes Sciences at the University Hospital of Wales. Only immediate members of the research team will have access to these samples. With your permission, we also plan to test the samples we collected during CATS I.  It is most likely that the samples will be destroyed (by incineration) at the end of the study, in this case, all identifiable information will be removed. Should we decide to store the samples, we will apply for additional Research Ethics Committee approval before CATS II is complete and request your permission.

*7. Will any genetic tests be done?*

With your permission, we will obtain blood samples, saliva or mouth swabs from you and/or your child to test genes involved in controlling the production of thyroid hormone. Because changes in genes can be inherited, if you are discovered to have such a change it is possible that other members of your family might also be affected. We will therefore ask for your permission to contact your immediate family members (parents, brothers/sisters, children) to see whether they also wish to be tested.

*8. What will happen to the results of the research study?*

The results of the research study will be prepared for publication in appropriate medical journals together with presentation at medical conferences. People participating in the study will be able to obtain a copy of the results after they have been published in the relevant journal(s). Participants will not be identified in any report/publication.

We will also organise a meeting at the end of the study to let you know about the results.

*9. Who is organising and funding the research?*

The study is being organised by Professor Marian Ludgate, Dr Aled Rees, Professor John Gregory and Professor John Lazarus from the Centre for Endocrine and Diabetes Sciences at the University Hospital of Wales. Funding for the study is provided by the Charles Wolfson Charitable Trust. The doctors conducting the research are not being paid for including and looking after women and children in the study.

*10. Who has reviewed the study?*

The Cardiff & Vale University Health Board Research & Development Office and by the South East Wales Research Ethics Committee.

*11. Future Studies*

It is possible that further research may be carried out related to the CATS studies. If this happens, we may contact you again at some time in the future to ask if you would be prepared to be involved in future studies.

*12. Further information and contact details*

Should you have any further queries regarding this research study, then please do not hesitate to contact us on 02920 745457 or 02920 5002. You can also contact us via e-mail on [ludgate@cf.ac.uk](mailto:ludgate@cf.ac.uk) or [reesda@cf.ac.uk](mailto:reesda@cf.ac.uk)

**Thank you for considering taking part in this study.**

## 3. Information for the child

**Child Information Sheet,  V4, 1 April 2014 (Initial)**

**Study title**

Testing how the thyroid gland affects the growth and health of mothers and their children.

**What is research? Why is this project being done?**

Research is a way we try to find out the answers to questions. When your mother was expecting you she took part in some tests and may have taken some medicine until you were born. You might remember that when you were 3, you did some puzzles for us in your house. Now that you are growing up, we would like you to help us find out more. We want to look at how much fat and muscle there is in your body, and how a gland in your neck called the thyroid gland can affect this.

**Why have I been asked to take part?**

There is nothing wrong with you, and everybody has this gland, but you have been specially chosen because your mother was tested before.

**Did anyone else check the study is OK to do?**

Before any research is allowed to happen, it has to be checked by a group of people called a Research Ethics Committee. They make sure that the research is fair. Your project has been checked by the South East Wales Research Ethics Committee.

**Do I have to take part?**

You do not have to take part, it is completely your choice. Even if you do decide to join in you do not have to have all of the tests if you prefer not to.

**What will happen to me if I take part in the research?**

You and your mother might prefer to help us from home. In this case, your mother will complete some forms explaining how you are getting on in school and with your friends. We will also ask you to scrape the inside of your cheek with a special stick (it won't hurt) or spit into a tube. We will use this to find out whether you have inherited anything from your mum which alters how your thyroid works.

If you agree to come to the hospital with your mother to have some tests, it will take all morning and you may miss school. On the day that you come for the tests, you should not have breakfast because we will provide that for you. You will have the same tests as your mother.

We will see how much you weigh, check how tall you are, and measure your heart beat.  We will ask you for a urine sample.  We will also ask your mother some questions about you. We will also take a picture of you with a special

247

camera, to see how much bone and muscle you have and give you a copy to show your friends.

Also, if you will let us, and your parents say it is OK, we will take a small amount of blood (6 teaspoonfuls) to see how the gland in your neck is working. If you don't want a blood test we will ask for a scrape from inside your cheek (it won't hurt) or spit into a tube. We will use this to find out whether you have inherited anything from your mother which alters how your thyroid works.

**We would also like for you to take part in a test that will take about 1-1½ hours, which will check your learning and concentration.  The test will involve puzzles, answering questions and remembering certain things.**

**Might anything about the research upset me?**

Most of the tests will not hurt at all. If you agree for us to take some blood, to prevent this from hurting, we can put some special cream on your skin before the blood test which helps to stop you from feeling it.

**Will joining in help me?**

We cannot promise the study will help you but the information we get might help treat mothers who have a problem with their thyroid, especially when they are pregnant, and their children.

**What happens when the research stops?**

The study does not involve you taking any medicine so there will be no change for you.

**What if something goes wrong during the project?**

This is very unlikely since we are doing tests to find out if medicine your mother may have taken before is having an effect.

**Will my medical details be kept private if I take part? Will anyone else know I'm doing this?**

Only the doctors involved in the study will know about you taking part.

**What if I don't want to do the research anymore?**

If at any time you don't want to do any of the tests, just tell your parents, doctor or nurse. They will not be cross with you.

### 4. Response form

# CATS II

Response form V, V3, 1 April 2014

Please complete and return in the envelope provided.

Name:………………………………………………………………….………..…………

(Previous name if recently married):
……………………..…………………………………

Address:………………………………………………………………….…………………

Post Code: ………………………….......................DOB: …………………………

Contact No.:………………………………………………………………….

Email Address: ………………….………………………………………

(Please place a star next to desired mode of contact)

Child's Name:………………………………. DOB:…………….………………

Please initial to indicate your preference.

□ Before deciding to participate or not, I wish to obtain further information and would like to talk to the patient adviser (Professor John Lazarus). I give my permission for you to contact me by telephone.

□ I am interested in participating by providing a DNA and a urine sample from myself and my child and completing some questionnaires at home. I give my permission for you to provide me with the required information, kits and questionnaires by post.

□ I am interested in participating by having a home visit in which my child will be evaluated using tests of cognition, by providing a DNA and a urine sample from myself (using the provided kits) and my child and my completing some questionnaires.

□ I am interested in participating and am prepared to attend at the University Hospital of Wales, Heath Park, Cardiff, with my child, for a half day for a range of non-invasive tests. I give my permission for you to contact me by telephone to organise a convenient time and date to attend.

□ I give my permission for you to inform my GP about the study.

Thank you for your time.

## Appendix 5: Information sheets from the post pack

1. Information for the mother
2. Information for the child

1. Information for the mother

V4(ii), 6 February 2012

PART 1

**2. Title of study**
Controlled Antenatal Thyroid Screening Study II (CATS 2)
3. Introduction

Thank you for taking the time to read this leaflet. You have indicated that you are prepared to take part in the CATS 2 study. Before you confirm that you wish to become involved it is important that you understand why the research is being done and what it will involve. Please take time to read the following information carefully and discuss it with others if you wish

Part 1 reminds you of the purpose of this study and what will happen if you take part.

Part 2 gives you more detailed information about the conduct of the study.

Please do not hesitate to ask us if there is anything that is not clear or if you would like more information.

Take time to consider whether or not you still wish to take part.

*4. What is the purpose of this study?*

Following on from your participation in the original CATS study, scientists and doctors at the University Hospital of Wales are hoping to arrange further studies to investigate how thyroid hormone levels during pregnancy may affect the development of the proportions of fat, muscle and general development in a child. They will also investigate whether there is any effect on fat and bone development in the mother and/or any influence on their general wellbeing. It is only those involved in the previous study that will be able to help us with our ongoing investigations.

*5. Why have I been chosen?*

It is only those involved in the previous study that will be able to help us with our ongoing investigations.

We really appreciated your involvement in the CATS 1 study on thyroid hormone levels in pregnancy, and how variations in this might affect your child. The results of this study should be available later this year and will prove useful in understanding these effects and help to improve the care of pregnant women and their children.

*6. Do I have to take part?*

It is up to you to decide whether or not to take part. If you choose not to take part, that is fine by us, you do not have to give us reasons why and this decision will not have any implications for your future medical care. If you decide to take part you will be asked to sign a consent form. You are still free to withdraw at any time and without giving a reason. A decision to withdraw at any time, or a decision not to take part, will not affect the standard of care you receive.

*7. What will happen to me if I take part?*

You and/or your child have opted to participate from home, this will involve completing some questionnaires and obtaining mouth swabs or saliva samples from you and your child for genetic studies. These will provide us with a sample of DNA which will be analyzed for variations in genes. These will need to be returned to us, along with the completed consent forms confirming you are happy for us to analyse these questionnaires and samples.

*8. Expenses and payments*

There should not be any expense incurred by you since we have provided envelopes with the postage paid for you to return the samples and questionnaires.

*9. What do I have to do?*

   1.      Complete the questionnaires in your own time. There are 4 of these labelled:

ADHD, this provides information about your child's social behaviour.

SCQ, this provides information about your child's activity level.

SDQ, this indicates how your child is progressing.

General, this contains information which is useful to us about you and your child.

2. Use the kits provided to obtain saliva or a sample from the inside of your and your child's mouth. There are instructions in the box but please contact us on 02920 745457, if you need more information. We will use the samples to obtain DNA to look for variations in the genes which influence the different aspects that we will measure in the study. If you need more information, you may contact Dr Marian Ludgate (02920 745457) or Dr Aled Rees (02920 745002).

3. Please return all completed questionnaires, samples and signed consent forms in the stamped addressed containers provided.

*10. What are the alternatives for diagnosis or treatment?*

At present, the thyroid function of women who are pregnant is not tested at all. Our studies will help to decide whether it is advisable for this to be done, either because of benefits to the child, the mother or both.

*11. What are the possible disadvantages and risks of taking part?*

No side effects are anticipated from the study procedures.

*12. What are the side effects of any treatment received when taking part?*

The new study (CATS II) does not involve you taking any medication or receiving any treatment. In CATS I, you may have received tablets containing thyroid hormone which is exactly the same as produced naturally by your thyroid and so does not have any side effects.

*13. Exposure to radiation or ionising radiation*

None.

*14. Harm to the unborn child*

None of the investigations to be performed have any negative effect on an unborn child or will affect you or your child if you are breast feeding.

*15. What are the potential benefits of taking part?*

We cannot promise the study will help you but the information we get from this study will help improve the future care of pregnant women with consequent benefit to their children.

*16. What happens when the research study stops?*

Since no treatment is involved, there are no special considerations to take into account. However, with your permission, we will store samples and this is explained in part 2.

*17. What if something goes wrong?*

Any complaint about the way you have been dealt with during the study or any possible harm you might suffer will be addressed. The detailed information on this is given in Part 2.

*18. Will my taking part in this study be kept confidential?*

Yes. We will follow ethical and legal practice and all information about you will be handled in confidence. The details are included in Part 2.

**This completes Part 1.**

**If the information in Part 1 has interested you and you are considering participation, please read the additional information in Part 2 before making any decision.**

PART 2

*1. What if relevant new information becomes available?*

Sometimes we get new information about treatments, in this case we are measuring the effects of treatment you may have received when you were pregnant and not during this new study.

*2. What will happen if I don't want to carry on with the study?*

If you withdraw from the study, with your permission, we will keep any samples, and use any data collected up to your withdrawal. A decision to withdraw at any time will not affect the standard of care you receive.

Similarly in the event of your loss of capacity, with your permission, we will keep any samples, and use any data collected prior to this.

*3. What if there is a problem?*

If you have a concern about any aspect of this study, you should ask to speak to the researchers who will do their best to answer your questions (phone 02920 745457 or 02920 745002). This study is being indemnified by Cardiff University. If you are harmed by taking part in this research project, there are no special compensation arrangements. If you are harmed due to someone's negligence, then you may have grounds for a legal action but you may have to pay for it. Regardless of this, if you wish to complain, or have any concerns about any aspect of the way you have been approached or treated during the course of this study, the normal National Health Service complaints mechanisms should be available to you.

*4. Will my taking part in this study be kept confidential?*

All information which is collected about you during the course of the research will be kept strictly confidential. No personally identifiable information will be held on NHS or University computers although storage of study data on both systems will occur in an anonymised form. These computers are held by Drs Ludgate, Rees and other members of the study team and are stored in locked rooms (rooms 254 & 261, 2nd Floor C block, University Hospital of Wales). The computers are password protected and no identifiable data will be transferred electronically, though final analysis of the anonymised study data may require email transfer to Professor Newcombe in the Department of Primary Care and Public Health who may provide statistical advice.

*5. Involvement of your general practitioner/family doctor*

With your permission your GP will be informed of your participation in this study.

*6. What will happen to any samples I give?*

The DNA samples for this study will be collected and stored securely for later analysis in the Centre for Endocrine and Diabetes Sciences at the University Hospital of Wales. Only immediate members of the research team will have access to these samples.

With your permission, we also plan to test the samples we collected during CATS 1.

It is most likely that the samples will be destroyed (by incineration) at the end of the study, in this case, all identifiable information will be removed. Should we decide to store the samples, we will apply for additional LREC approval before CATS II is complete and request your permission.

*7. Will any genetic tests be done?*

With your permission, we will obtain mouth swabs from you and/or your child to test genes involved in controlling the production of thyroid hormone. Because changes in genes can be inherited, if you are discovered to have such a change it is possible that other members of your family might also be

affected. We will therefore ask for your permission to contact your immediate family members (parents, brothers/sisters, children) to see whether they also wish to be tested.

*8. What will happen to the results of the research study?*

The results of the research study will be prepared for publication in appropriate medical journals together with presentation at medical conferences. People participating in the study will be able to obtain a copy of the results after they have been published in the relevant journal(s). Participants will not be identified in any report/publication.

We will also organise a meeting at the end of the study to let you know about the results.

*9. Who is organising and funding the research?*

The study is being organised by Dr Marian Ludgate, Dr Aled Rees, Professor John Gregory and Professor John Lazarus from the Centre for Endocrine and Diabetes Sciences at the University Hospital of Wales. Funding for the study is provided by the Charles Wolfson Charitable Trust. The doctors conducting the research are not being paid for including and looking after women and children in the study.

*10. Who has reviewed the study?*

The Cardiff & Vale University Health Board Research & Development Office and by the South Wales Research Ethics Committee.

*11. Future studies*

It is possible that further research may be carried out related to the CATS studies. If this happens, we may contact you again at some time in the future to ask if you would be prepared to be involved in future studies.

*12. Further information and contact details*

Should you have any further queries regarding this research study, then please do not hesitate to contact us on 02920 745457 or 02920 5002. You can also contact us via e-mail on **ludgate@cf.ac.uk** or **reesda@cf.ac.uk**

Thank you for considering taking part in this study.

## 2. Information for the child

**Child Information Sheet, V4(i), 6 February 2012  (Remote )**

**Study title**

Testing how the thyroid gland affects the growth and health of <mark>mothers</mark> and their children.

**What is research? Why is this project being done?**

Research is a way we try to find out the answers to questions. When your mother was expecting you she took part in some tests and may have taken some medicine until you were born. You might remember that when you were 3, you did some puzzles for us in your house. Now that you are growing up, we would like you to help us find out more. We want to look at how much fat and muscle there is in your body, and how a gland in your neck called the thyroid gland can affect this.

**Why have I been asked to take part?**

There is nothing wrong with you, and everybody has this gland, but you have been specially chosen because your mother was tested before.

**Did anyone else check the study is OK to do?**

Before any research is allowed to happen, it has to be checked by a group of people called a Research Ethics Committee. They make sure that the research is fair. Your project has been checked by the South East Wales Research Ethics Committee.

**Do I have to take part?**

You do not have to take part, it is completely your choice. Even if you do decide to join in you do not have to have all of the tests if you prefer not to.

**What will happen to me if I take part in the research?**

You and your mother prefer to help us from home. Your mother will complete some forms explaining how you are getting on in school and with your friends. We will also ask you to scrape the inside of your cheek with a special stick (it won't hurt) or spit into a tube. We will use this to find out whether you have inherited anything from your mum which alters how your thyroid works.

**Might anything about the research upset me?**

None of the tests will hurt at all.

**Will joining in help me?**

We cannot promise the study will help you but the information we get might help treat mothers who have a problem with their thyroid, especially when they are pregnant, and their children.

**What happens when the research stops?**

The study does not involve you taking any medicine so there will be no change for you.

**What if something goes wrong during the project?**

This is very unlikely since we are doing tests to find out if medicine your mother may have taken before is having an effect.

**Will my medical details be kept private if I take part? Will anyone else know I'm doing this?**

Only the doctors involved in the study will know about you taking part.

**What if I don't want to do the research anymore?**

If at any time you don't want to do any of the tests, just tell your parents, doctor or nurse. They will not be cross with you.

## Appendix 6: Appointment letters

1. Remote/home visit
   a. Cover letter
   b. Information for mother
   c. Information for child
   d. Copy of consent form
2. Research centre visit
   a. Cover letter
   b. Information for mother
   c. Information for child
   d. Copy of consent form

1. Remote/home visit
   a. Cover letter

>>ADDRESS<<

Dear >>NAME<<,

Further to your conversation with my colleague on the >>DATE<<, I have enclosed some information letters and a copy of the consent form for my visit at your home.

I look forward to meeting you and >>CHILD NAME<< on >>APPOINTMENT DATE AND TIME<<. If you have any problems with the appointment, please ring or text: 07908 243 142.

Yours sincerely,

Charlotte Hales

b. Information for mother

V6, 1 April 2014

PART 1

**2. Title of study**
Controlled Antenatal Thyroid Screening Study II (CATS 2)
3. Introduction

Thank you for taking the time to read this leaflet. You have indicated that you are prepared to take part in the CATS 2 study. Before you confirm that you wish to become involved it is important that you understand why the research is being done and what it will involve. Please take time to read the following information carefully and discuss it with others if you wish.

Part 1 reminds you of the purpose of this study and what will happen if you take part.

Part 2 gives you more detailed information about the conduct of the study.

Please do not hesitate to ask us if there is anything that is not clear or if you would like more information.

Take time to consider whether or not you still wish to take part.

*4. What is the purpose of this study?*

Following on from your participation in the original CATS study, scientists and doctors at the University Hospital of Wales are hoping to arrange further studies to investigate how thyroid hormone levels during pregnancy may affect the development of the proportions of fat, muscle and general development in a child. They will also investigate whether there is any effect on fat and bone development in the mother and/or any influence on their general wellbeing. It is only those involved in the previous study that will be able to help us with our ongoing investigations.

*5. Why have I been chosen?*

It is only those involved in the previous study that will be able to help us with our ongoing investigations.

We really appreciated your involvement in the CATS 1 study on thyroid hormone levels in pregnancy, and how variations in this might affect your child. The results of this study should be available later this year and will prove useful in understanding these effects and help to improve the care of pregnant women and their children.

*6. Do I have to take part?*

It is up to you to decide whether or not to take part. If you choose not to take part, that is fine by us, you do not have to give us reasons why and this decision will not have any implications for your future medical care. If you decide to take part you will be asked to sign a consent form. You are still free to withdraw at any time and without giving a reason. A decision to withdraw at any time, or a decision not to take part, will not affect the standard of care you receive.

*7. What will happen to me if I take part?*

You and/or your child have opted to participate from home, this will involve completing some questionnaires, providing a urine sample from you and your child and obtaining mouth swabs or saliva samples from you and your child for genetic studies. These will provide us with a sample of DNA which will be analyzed for variations in genes. These will need to be returned to us, along with the completed consent forms confirming you are happy for us to analyse these questionnaires and samples.

If you and your child agree, your child would also undertake an intelligence test called Weschler Intelligence Scale for Children (Fourth UK Edition (WISC-IV[uk]) . and sections from the Neuropsychological Assessment Second Edition. Our research assistant will arrange an appointment to visit your home. The test usually takes about 1-1½ hours and would generate an IQ score, which would represent your child's intelligence. You can of course be present during the test, but some children do not perform as well if their parents are there as they are distracted. You will be able to have a copy of the report if you wish.

*8. Expenses and payments*

There should not be any expense incurred by you since we have provided envelopes with the postage paid for you to return the samples and questionnaires.

*9. What do I have to do?*

1. Complete the questionnaires in your own time. There are 4 of these labelled:

ADHD, this provides information about your child's social behaviour.

SCQ, this provides information about your child's activity level.

SDQ, this indicates how your child is progressing.

General, this contains information which is useful to us about you and your child.

3. Use the kits provided to obtain urine sample and saliva or a sample from the inside of your and your child's mouth. There are instructions in the box but please contact us on 02920 745457, if you need more information. We will use the samples to obtain DNA to look for variations in the genes which influence the different aspects that we will measure in the study. If you need more information, you may contact Professor Marian Ludgate (02920 745457) or Dr Aled Rees (02920 745002).

4. Please return all completed questionnaires, samples and signed consent forms in the stamped addressed containers provided. Or if you have agreed, give these to the research assistant who visits your home to conduct the IQ test on your child.
5. If you and your child have agreed to the IQ test, if possible please make a quiet space available with a table, but don't worry if this is difficult.

*10. What are the alternatives for diagnosis or treatment?*

At present, the thyroid function of women who are pregnant is not tested at all. Our studies will help to decide whether it is advisable for this to be done, either because of benefits to the child, the mother or both.

*11. What are the possible disadvantages and risks of taking part?*

No side effects are anticipated from the study procedures.

*12. What are the side effects of any treatment received when taking part?*

The new study (CATS II) does not involve you taking any medication or receiving any treatment. In CATS I, you may have received tablets containing thyroid hormone which is exactly the same as produced naturally by your thyroid and so does not have any side effects.

*13. Exposure to radiation or ionising radiation*

None.

*14. Harm to the unborn child*

None of the investigations to be performed have any negative effect on an unborn child or will affect you or your child if you are breast feeding.

*15. What are the potential benefits of taking part?*

We cannot promise the study will help you but the information we get from this study will help improve the future care of pregnant women with consequent benefit to their children.

*16. What happens when the research study stops?*

Since no treatment is involved, there are no special considerations to take into account. However, with your permission, we will store samples and this is explained in part 2.

*17. What if something goes wrong?*

Any complaint about the way you have been dealt with during the study or any possible harm you might suffer will be addressed. The detailed information on this is given in Part 2.

*18. Will my taking part in this study be kept confidential?*

Yes. We will follow ethical and legal practice and all information about you will be handled in confidence. The details are included in Part 2.


**This completes Part 1.**

**If the information in Part 1 has interested you and you are considering participation, please read the additional information in Part 2 before making any decision.**


PART 2

*1. What if relevant new information becomes available?*

Sometimes we get new information about treatments, in this case we are measuring the effects of treatment you may have received when you were pregnant and not during this new study.

*2. What will happen if I don't want to carry on with the study?*

If you withdraw from the study, with your permission, we will keep any samples, and use any data collected up to your withdrawal. A decision to withdraw at any time will not affect the standard of care you receive.

Similarly in the event of your loss of capacity, with your permission, we will keep any samples, and use any data collected prior to this.

*3. What if there is a problem?*

If you have a concern about any aspect of this study, you should ask to speak to the researchers who will do their best to answer your questions (phone 02920 745457 or 02920 745002). This study is being indemnified by Cardiff University. If you are harmed by taking part in this research project, there are no special compensation arrangements. If you are harmed due to someone's negligence, then you may have grounds for a legal action but you may have to pay for it. Regardless of this, if you wish to complain, or have any concerns about any aspect of the way you have been approached or treated during the course of this study, the normal National Health Service complaints mechanisms should be available to you.

*4. Will my taking part in this study be kept confidential?*

All information which is collected about you during the course of the research will be kept strictly confidential. No personally identifiable information will be held on NHS or University computers although storage of study data on both systems will occur in an anonymised form. These computers are held by Professor Ludgate, Dr Rees and other members of the study team and are stored in locked rooms (rooms 254 & 261, 2nd Floor C block, University Hospital of Wales). The computers are password protected and no identifiable data will be transferred electronically, though final analysis of the anonymised study data may require email transfer to Professor Newcombe in the Department of Primary Care and Public Health who may provide statistical advice.

*5. Involvement of your general practitioner/family doctor*

With your permission your GP will be informed of your participation in this study.

*6. What will happen to any samples I give?*

The urine and DNA samples for this study will be collected and stored securely for later analysis in the Centre for Endocrine and Diabetes Sciences at the University Hospital of Wales. Only immediate members of the research team will have access to these samples.

With your permission, we also plan to test the samples we collected during CATS 1.

It is most likely that the samples will be destroyed (by incineration) at the end of the study, in this case, all identifiable information will be removed. Should we decide to store the

samples, we will apply for additional Research Ethics Committee approval before CATS II is complete and request your permission.

*7. Will any genetic tests be done?*

With your permission, we will obtain mouth swabs from you and/or your child to test genes involved in controlling the production of thyroid hormone. Because changes in genes can be inherited, if you are discovered to have such a change it is possible that other members of your family might also be affected. We will therefore ask for your permission to contact your immediate family members (parents, brothers/sisters, and children) to see whether they also wish to be tested.

*8. What will happen to the results of the research study?*

The results of the research study will be prepared for publication in appropriate medical journals together with presentation at medical conferences. People participating in the study will be able to obtain a copy of the results after they have been published in the relevant journal(s). Participants will not be identified in any report/publication.

We will also organise a meeting at the end of the study to let you know about the results.

*9. Who is organising and funding the research?*

The study is being organised by Professor Marian Ludgate, Dr Aled Rees, Professor John Gregory and Professor John Lazarus from the Centre for Endocrine and Diabetes Sciences at the University Hospital of Wales. Funding for the study is provided by the Charles Wolfson Charitable Trust. The doctors conducting the research are not being paid for including and looking after women and children in the study.

*10. Who has reviewed the study?*

The Cardiff & Vale University Health Board Research & Development Office and by the South East Wales Research Ethics Committee

*11. Future studies*

It is possible that further research may be carried out related to the CATS studies. If this happens, we may contact you again at some time in the future to ask if you would be prepared to be involved in future studies.

*12. Further information and contact details*

Should you have any further queries regarding this research study, then please do not hesitate to contact us on 02920 745457 or 02920 5002. You can also contact us via e-mail on [ludgate@cf.ac.uk](mailto:ludgate@cf.ac.uk) or [reesda@cf.ac.uk](mailto:reesda@cf.ac.uk)

**Thank you for considering taking part in this study.**

**Child Information Sheet,  V6, 1 April 2014 (Remote )**

**Study title**

Testing how the thyroid gland affects the growth and health of mothers and their children.

**What is research? Why is this project being done?**

Research is a way we try to find out the answers to questions. When your mother was expecting you she took part in some tests and may have taken some medicine until you were born. You might remember that when you were 3, you did some puzzles for us in your house. Now that you are growing up, we would like you to help us find out more. We want to look at how much fat and muscle there is in your body, and how a gland in your neck called the thyroid gland can affect this.

**Why have I been asked to take part?**

There is nothing wrong with you, and everybody has this gland, but you have been specially chosen because your mother was tested before.

**Did anyone else check the study is OK to do?**

Before any research is allowed to happen, it has to be checked by a group of people called a Research Ethics Committee. They make sure that the research is fair. Your project has been checked by the South East Wales Research Ethics Committee.

**Do I have to take part?**

You do not have to take part, it is completely your choice. Even if you do decide to join in you do not have to have all of the tests if you prefer not to.

**What will happen to me if I take part in the research?**

You and your mother prefer to help us from home. Your mother will complete some forms explaining how you are getting on in school and with your friends. We will also ask you to scrape the inside of your cheek with a special stick (it won't hurt) or spit into a tube. We will use this to find out whether you have inherited anything from your mum which alters how your thyroid works.  We will also ask you for a urine sample.

If you and your mother agree, we would also like for you to take part in a test that will take about 1-1½ hours, which will check your learning and concentration.  The test will involve puzzles, answering questions and remembering certain things.

**Might anything about the research upset me?**

None of the tests will hurt at all.

**Will joining in help me?**

We cannot promise the study will help you but the information we get might help treat mothers who have a problem with their thyroid, especially when they are pregnant, and their children.

**What happens when the research stops?**

The study does not involve you taking any medicine so there will be no change for you.

**What if something goes wrong during the project?**

This is very unlikely since we are doing tests to find out if medicine your mother may have taken before is having an effect.

**Will my medical details be kept private if I take part? Will anyone else know I'm doing this?**

Only the doctors involved in the study will know about you taking part.

**What if I don't want to do the research anymore?**

If at any time you don't want to do any of the tests, just tell your parents, doctor or nurse. They will not be cross with you.

d. Copy of consent form

**Participant Consent Form: Involvement in Controlled Antenatal Thyroid Screening II (CATS 2) Study,  V7, 1 April 2014 (Remote participants)**

**Name of Study: CATS 2**

**Researchers: Dr Marian Ludgate, Dr Aled Rees, Professor John Gregory, Professor John Lazarus**

I, (full name)………………………………………………….  Date of Birth………………….…..

(address)……………………………………………………………………………………………………….

……………………………………………………………………………………………………………………

as the mother of……………………………………………………(full name of child)

agree to myself ☐ and/or my child ☐ being involved in the above study.

I have read the accompanying information leaflet and understand that involvement in the study is voluntary, and that non-involvement in the study will not affect the medical treatment of me or my child in any way, and that I will be free to withdraw at any time, without giving any reason, without my medical care or legal rights being affected.

The data will be analysed in accordance with the study protocol and remain confidential. I understand that if any abnormality is discovered in the tests carried out that I will be informed of this and I will be directed to the appropriate clinical services.

The aspects of the study that I specifically agree to participate in are (please initial):

☐    collection of *my child 's* DNA from a mouth swab or from saliva sample to test for variations within genes that affect the different aspects that we will measure in the study.  After testing the sample will probably be destroyed but it is possible that they may be used in a related project.

☐    collection of *my* DNA from a mouth swab or saliva sample to test for variations within genes that affect the different aspects that we will measure in the study.  After testing the sample will probably be destroyed but it is possible that they may be used in a related project.

☐    Collection of my urine sample.

☐    Collection of a urine sample from my child.

☐    I give permission for the study team to contact my immediate family members (parents, brothers/sisters, children) to see whether they also

wish to have mouth testing carried out to test genes involved in controlling the production of thyroid hormone.

☐   My completion of the Strengths and Difficulties, ADHD & SCQ Questionnaires with respect to my child and the Edinburgh Postnatal Depression Scale with respect to myself;

☐   I give my permission for my child to take part in the Weschler Intelligence Scale for children (this will involve 1-1½ hour test) and sections from the Neuropsychological Assessment Second Edition.

☐   In the event of my loss of capacity, I give you my consent to retain and use the samples and data you have collected from me.

☐   I give my consent for you to inform my GP of my participation in the study.

☐   I am happy for my contact details to be retained by Cardiff University Research Team to  consider whether I would be suitable to take part in any future related studies and am  happy to be contacted about such studies.

_____    _____    _____

Name of participant                    Date                Signature

_____    _____    _____

Researcher                            Date                Signature

_____    _____    _____

Name of person taking consent       Date                Signature

(if different from researcher)

2. Research centre visit
   a. Cover letter

<div align="center">

Department of Medicine

</div>
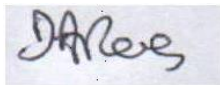
Dear **>>NAME<<**,

Further to your telephone conversation with my colleague, thank you for agreeing to take part in the CATS II Study.  We would be grateful if you and your **>>SON/DAUGHTER<<** could attend the Clinical Research Facility at the University Hospital of Wales on >>DATE AND TIME OF APPOINTMENT<<, a map with directions to the Clinical Research Facility is enclosed.

Please could we remind you that it is important that you attend having fasted since the night before.  You can drink as much water as you like and we will provide breakfast as soon as possible after you arrive.

If you have any queries, please do not hesitate to contact us one of our dedicated CATS mobiles: 07908 243142 or  07866 980039.

Yours sincerely,

Dr Aled Rees

Senior Lecturer and Honorary Consultant

Centre for Endocrine & Diabetes Sciences

# Directions to the Clinical Research Facility (CRF) from the Multi-Storey Car Park

N

**KEY**

→ Direction to follow from Multi-Storey Car Park

B Bus Stops

Residential Road

Allensbank Rd Entrance

B

B

Sports Centre

Traffic Lights

Covered walkway

Lake

Car Park for Disabled Badge Holders

MULTI STOREY

To A48

Gateway Circus

RJP 08.11

Central Way

B

Concourse

Stairs

Lifts

Pharmacy

X-Ray Dept

Outside Area

Emergency Unit

GET LIFT TO UPPER GROUND

Lifts

Lifts

H Helipad

The Gateway

TWO STOREY CAR PARK

Slip Road from Gabalfa Interchange

Heath Park Way

Upper Ground Floor Corridor

Lifts

Lifts

Road Closed

CRF

Covered Bridge over road

B

Staff Entrance

Academic Avenue

Rhydelig Avenue Entrance

Heart Research Centre

Tenovus

Dental Hospital

      b.   Information for the mother

**PARTICIPANT INFORMATION LEAFLET (UHW attenders)**

V5, 1 April 2014

PART 1

## 1. Title of study
Controlled Antenatal Thyroid Screening Study II (CATS 2)

## 2. Introduction

Thank you for taking the time to read this leaflet. You have indicated that you are prepared to take part in the CATS 2 study. Before you confirm that you wish to become involved it is important that you understand why the research is being done and what it will involve. One of our team will go through the information sheet with you and answer any questions you have, this should take about 5 minutes. Please take time to read the following information carefully and discuss it with others if you wish.

Part 1 reminds you of the purpose of this study and what will happen if you take part.

Part 2 gives you more detailed information about the conduct of the study.

Please do not hesitate to ask us if there is anything that is not clear or if you would like more information.

Take time to consider whether or not you still wish to take part.

*3. What is the purpose of this study?*

Following on from your participation in the original CATS study, scientists and doctors at the University Hospital of Wales are hoping to arrange further studies to investigate how thyroid hormone levels during pregnancy may affect the development of the proportions of fat, muscle and general development in a child.  They will assess your child's abilities in a range of areas, such as verbal and non-verbal skills, memory and speed of processing with an IQ test.  They will also investigate whether there is any effect on fat and bone development in the mother and/or any influence on their general wellbeing.

*4. Why have I been chosen?*

It is only those involved in the previous study that will be able to help us with our ongoing investigations.  We really appreciated your involvement in the CATS 1 study on thyroid hormone levels in pregnancy, and how variations in this might affect your child. The results of this study should be available later this year and will prove useful in understanding these effects and help to improve the care of pregnant women and their children.

*5. Do I have to take part?*

It is up to you to decide whether or not to take part. If you choose not to take part, that is fine by us, you do not have to give us reasons why and this decision will not have any implications for your future medical care. If you decide to take part you will be asked to sign a consent form. You are still free to withdraw at any time and without giving a reason. A decision to withdraw at any time, or a decision not to take part, will not affect the standard of care you receive.

*6. What will happen to me if I take part?*

As you and your child have opted to attend the University Hospital of Wales (UHW) this will involve some or all (depending on your wishes) of the following tests and could take most of the morning:

1) Measure height and weight in you and your child.
2) Measure blood pressure and assess the blood flow in the blood vessels of you and your child's arm with a probe put on the skin in you and your child.
3) Complete questionnaires with you about your mood and your child's social and educational development and progress.
4) Perform a DEXA scan to measure the proportion of fat and muscle in you (providing you are not pregnant at the time) and your child (this would involve lying still on the scan table for a minute or two whilst the scan is performed and does not hurt).  DEXA is a simple, rapid and non-invasive technique which is used routinely in clinical practice to measure bone density.  This scan does involve exposure to a small amount of radiation (no more than experienced during a day walking around outside in the UK) but the risk is negligible.
5) Take a blood sample from you and your child (about 6 teaspoons), to measure thyroid hormone levels, aspects of the control of bone and fat tissues and variations in the genes which influence the different aspects that we will measure in the study.  Some of the tests need for you and your child to have fasted since the night before; we will provide breakfast as soon as possible after you arrive.  If you or your child prefers not to have a blood test we would still like to take saliva samples or mouth swabs for genetic studies.
6. For your child to undertake an intelligence test called Weschler  Intelligence Scale for Children (Fourth UK Edition (WISC-IV[uk]) and sections from the

Neuropsychological Assessment Second Edition..  This usually takes about 1-1½ hours and would generate an IQ score, which would represent your child's intelligence. You can of course be present during the test, but some children do not perform as well if their parents are there as they are distracted.  You will be able to have a copy of the report if you wish.

7.    Ask you and your child to provide a urine sample.

## 7. Expenses and payments

If you wish, we will be able to reimburse any traveling expenses/car parking fees incurred while attending for the study visit.

## 8. What do I have to do?

Please decide which of the tests (if any) you are willing to undertake. Please also discuss with your child which tests (if any) he/she is willing to undertake. Please feel free to ask for more information from the staff in CRF before signing the consent form indicating which tests you and your child will undertake.

## 9. What are the alternatives for diagnosis or treatment?

At present, the thyroid function of women who are pregnant is not tested at all. Our studies will help to decide whether it is advisable for this to be done, either because of benefits to the child, the mother or both.

## 10.  What are the possible disadvantages and risks of taking part?

Other than possible discomfort (temporary pain, swelling, bruising and rarely infection) caused by the collection of blood, no other side effects are anticipated from the study procedures. A DEXA scan does involve exposure to a small amount of radiation but the risk is negligible. It is possible that the blood tests or DEXA scan could by chance pick up an unsuspected abnormality, in which case you will be given an opportunity to discuss these findings further with the doctors.

These investigations will only be performed once we have explained in detail what will be involved, you have given us your consent and your child has agreed to take part. You or your child may decide to undergo some but not all of the tests. It is possible that you decide to participate but your child prefers not too, or vice versa.

## 11. What are the side effects of any treatment received when taking part?

The new study (CATS II) does not involve you taking any medication or receiving any treatment. In CATS I, you may have received tablets containing thyroid hormone which is exactly the same as produced naturally by your thyroid and so does not have any side effects.

## 12. Exposure to radiation or ionising radiation

As mentioned above, you and/or your child may agree to have a DEXA scan. This involves exposure to a small amount of radiation but no more than experienced walking around outside for 1 day in the UK.

*13. Harm to the unborn child*

None of the investigations to be performed have any negative effect on an unborn child EXCEPT the DEXA scan. With your permission, we will perform a pregnancy test and if you it is positive you will not receive the DEXA scan. None of the investigations will affect you or your child if you are breast feeding.

*14. What are the potential benefits of taking part?*

We cannot promise the study will help you but the information we get from this study will help improve the future care of pregnant women with consequent benefit to their children.

*15. What happens when the research study stops?*

Since no treatment is involved, there are no special considerations to take into account. However, with your permission, we will store samples and this is explained in part 2.

*16. What if something goes wrong?*

Any complaint about the way you have been dealt with during the study or any possible harm you might suffer will be addressed. The detailed information on this is given in Part 2.

*17. Will my taking part in this study be kept confidential?*

Yes. We will follow ethical and legal practice and all information about you will be handled in confidence. The details are included in Part 2.

**This completes Part 1.**

**If the information in Part 1 has interested you and you are considering participation, please read the additional information in Part 2 before making any decision.**

PART 2

*1. What if relevant new information becomes available?*

Sometimes we get new information about treatments, in this case we are measuring the effects of treatment you may have received when you were pregnant and not during this new study.

*2. What will happen if I don't want to carry on with the study?*

If you withdraw from the study, with your permission, we will keep any samples, and use any data collected up to your withdrawal. A decision to withdraw at any time will not affect the standard of care you receive.

Similarly in the event of your loss of capacity, with your permission, we will keep any samples, and use any data collected prior to this.

*3. What if there is a problem?*

If you have a concern about any aspect of this study, you should ask to speak to the researchers who will do their best to answer your questions (phone 02920 745457 or 02920 745002). This study is being indemnified by Cardiff University. If you are harmed by taking part in this research project, there are no special compensation arrangements. If you are harmed due to someone's negligence, then you may have grounds for a legal action but you may have to pay for it. Regardless of this, if you wish to complain, or have any concerns about any aspect of the way you have been approached or treated during the course of this study, the normal National Health Service complaints mechanisms should be available to you.

*4. Will my taking part in this study be kept confidential?*

All information which is collected about you during the course of the research will be kept strictly confidential. No personally identifiable information will be held on NHS or University computers although storage of study data on both systems will occur in an anonymised form. These computers are held by Professor Ludgate, Dr Rees and other members of the study team and are stored in locked rooms (rooms 254 & 261, 2nd Floor C block, University Hospital of Wales). The computers are password protected and no identifiable data will be transferred electronically, though final analysis of the anonymised study data may require email transfer to Professor Newcombe in the Department of Primary Care and Public Health who may provide statistical advice.

*5. Involvement of your general practitioner/family doctor*

With your permission your GP will be informed of your participation in this study.

*6. What will happen to any samples I give?*

The blood and urine samples (6 teaspoons) for this study will be collected and stored securely for later analysis in the Centre for Endocrine and Diabetes Sciences at the University Hospital of Wales. Only immediate members of the research team will have access to these samples.

With your permission, we also plan to test the samples we collected during CATS I.

It is most likely that the samples will be destroyed (by incineration) at the end of the study, in this case, all identifiable information will be removed. Should we decide to store the samples, we will apply for additional Research Ethics Committee approval before CATS II is complete and request your permission.
*7. Will any genetic tests be done?*

With your permission, we will obtain blood samples, saliva or mouth swabs from you and/or your child to test genes involved in controlling the production of thyroid hormone. Because changes in genes can be inherited, if you are discovered to have such a change it is possible that other members of your family might also be affected. We will therefore ask for your permission to contact your immediate family members (parents, brothers/sisters, children) to see whether they also wish to be tested.

*8. What will happen to the results of the research study?*

The results of the research study will be prepared for publication in appropriate medical journals together with presentation at medical conferences. People participating in the study will be able to obtain a copy of the results after they have been published in the relevant journal(s). Participants will not be identified in any report/publication.
We will also organise a meeting at the end of the study to let you know about the results.

*9. Who is organising and funding the research?*

The study is being organised by Professor Marian Ludgate, Dr Aled Rees, Professor John Gregory and Professor John Lazarus from the Centre for Endocrine and Diabetes Sciences at the University Hospital of Wales. Funding for the study is provided by the Charles Wolfson Charitable Trust. The doctors conducting the research are not being paid for including and looking after women and children in the study.
*10. Who has reviewed the study?*

The Cardiff & Vale University Health Board Research & Development Office and by the South East Wales Research Ethics Committee.
*11. Future Studies*

It is possible that further research may be carried out related to the CATS studies. If this happens, we may contact you again at some time in the future to ask if you would be prepared to be involved in future studies.

*12. Further information and contact details*

Should you have any further queries regarding this research study, then please do not hesitate to contact us on 02920 745457 or 02920 5002. You can also contact us via e-mail on [ludgate@cf.ac.uk](mailto:ludgate@cf.ac.uk) or [reesda@cf.ac.uk](mailto:reesda@cf.ac.uk)

**Thank you for considering taking part in this study.**

c. Information for the child

**Department of Medicine**

**Child Information Sheet, V5, 1 April 2014 (UHW attenders )**

**Study title**
Testing how the thyroid gland affects the growth and health of mothers and their children.

**What is research? Why is this project being done?**
Research is a way we try to find out the answers to questions. When your mother was expecting you she took part in some tests and may have taken some medicine until you were born. You might remember that when you were 3, you did some puzzles for us in your house. Now that you are growing up, we would like you to help us find out more. We want to look at how much fat and muscle there is in your body, and how a gland in your neck called the thyroid gland can affect this.

**Why have I been asked to take part?**
There is nothing wrong with you, and everybody has this gland, but you have been specially chosen because your mother was tested before.

**Did anyone else check the study is OK to do?**
Before any research is allowed to happen, it has to be checked by a group of people called a Research Ethics Committee. They make sure that the research is fair. Your project has been checked by the South East Wales Research Ethics Committee.

**Do I have to take part?**
You do not have to take part, it is completely your choice. Even if you do decide to join in you do not have to have all of the tests if you prefer not to.

**What will happen to me if I take part in the research?**
If you agree to come to the hospital with your mother to have some tests, it will take all morning and you may miss school. On the day that you come for the tests, you should not have breakfast because we will provide that for you. You will have the same tests as your mother. We will see how much you weigh, check how tall you are, and measure your heart beat.  We will ask you for a urine sample.  We will also ask your mother some questions about you. We will also take a picture of you with a special camera, to see how much bone and muscle you have and give you a copy to show your friends. Also, if you will let us, and your parents say it is OK, we will take a small amount of blood (about 6 teaspoonfuls) to see how the gland in your neck is working. If you don't want a blood test we will ask for a scrape from inside your cheek (it won't hurt) or ask you to spit into a tube. We will use this to find out whether you have inherited anything from your mum which alters how your thyroid works.

**We would also like for you to take part in a test that will take about 1-**$1\frac{1}{2}$ hours, which will check your learning and concentration.  The test will involve puzzles, answering questions and remembering certain things.

**Might anything about the research upset me?**
Most of the tests will not hurt at all. If you agree for us to take some blood, to prevent this from hurting, we can put some special cream on your skin before the blood test which helps to stop you from feeling it.

**Will joining in help me?**
We cannot promise the study will help you but the information we get might help treat mothers who have a problem with their thyroid, especially when they are pregnant, and their children.

**What happens when the research stops?**
The study does not involve you taking any medicine so there will be no change for you.

**What if something goes wrong during the project?**
This is very unlikely since we are doing tests to find out if medicine your mother may have taken before is having an effect.

**Will my medical details be kept private if I take part? Will anyone else know I'm doing this?**
Only the doctors involved in the study will know about you taking part.

**What if I don't want to do the research anymore?**
**If at any time you don't want to do any of the tests, just tell your parents, doctor**

**or nurse.  They will not be cross with you.**

**Participant Consent Form: Involvement in Controlled Antenatal Thyroid Screening II (CATS 2) Study. V6, 1 April 2014, (UHW attenders)**
**Name of Study: CATS 2**

**Researchers: Dr Marian Ludgate, Dr Aled Rees, Professor John Gregory, Professor John Lazarus**

I, (full name)……………………………………………Date of Birth…….…………..

(address)…………………………………………………………………………………

as the mother of………………………………………………(full name of child)

agree to myself ☐ and/or my child ☐ being involved in the above study.

I have read the accompanying information leaflet and understand that involvement in the study is voluntary, and that non-involvement in the study will not affect the medical treatment of me or my child in any way, and that I will be free to withdraw at any time, without giving any reason, without my medical care or legal rights being affected.

The data will be analysed in accordance with the study protocol and remain confidential. I understand that if any abnormality is discovered in the tests carried out that I will be informed of this and I will be directed to the appropriate clinical services.

The aspects of the study that I specifically agree to participate in are (please initial):

☐ measurement of *my child's* height, weight and hip/waist circumference;

☐ measurement of *my* height, weight and waist/hip circumference;

☐ measurement of *my child's* blood pressure, heart rate and blood flow (involves being attached to a machine but does not hurt);

☐ measurement of *my* blood pressure, heart rate and blood flow;

☐ measurement of *my child's* proportion of fat, bone and muscle by DEXA scan (Involves lying flat on a scan table for a few minutes and does not hurt);

☐ measurement of *my* proportion of fat, bone and muscle by DEXA scan

(providing I am not pregnant);

☐     pregnancy test;

☐     for me to provide a urine sample

☐     for my child to provide a urine sample

☐     blood sampling (6 teaspoons) for collection of *my child's* DNA to test for genes, and variations within genes that affect the different aspects that we will measure in the study. After testing the sample will probably be destroyed but it is possible that they may be used in a related project. I also agree that the blood may be tested for the functioning of the thyroid gland, for salts in the blood, and for lipids and cholesterol;

☐     blood sampling (6 teaspoons) for collection of *my* DNA to test for genes and variations within genes that affect the different aspects that we will measure in the study, and nothing else. After testing the sample will be destroyed. I also agree that the blood may be tested for the functioning of the thyroid gland, for salts in the blood, and for lipids and cholesterol;

☐     collection of *my* DNA from a mouth swab or saliva sample to test for genes and variations within genes that affect the development of the thyroid gland and the production of thyroid hormones, and nothing else. After testing the sample will be destroyed;

☐     I give permission for the study team to contact my immediate family members (parents, brothers/sisters, children) to see whether they also wish to have mouth testing carried out to test genes involved in controlling the production of thyroid hormone.

☐     collection of *my* child's DNA from a mouth swab or saliva sample to test for genes and variations within genes that affect the development of the thyroid gland and the production of thyroid hormones, and nothing else. After testing the sample will be destroyed;

☐     *my* completion of the Strengths and Difficulties, ADHD & ADI Questionnaires

with respect to my child and the Edinburgh Postnatal Depression Scale with respect to myself;

☐     I give permission for my child to take part in the Weschler Intelligence Scale for

Children and sections from the Neuropsychological Assessment Second Edition.

(this will involve 1-1½ hour test)

☐    I give my consent for you to inform my GP of my participation in the study.

☐    In the event of my loss of capacity, I give you my consent to retain and use the samples and data you have collected from me.

☐    I am happy for my contact details to be retained by the Cardiff University Research Team to consider whether I would be suitable to take part in any future related studies and am happy to be contacted about such studies

_____     _____     _____

Name of participant         Date              Signature

_____     _____     _____

Researcher                 Date              Signature

_____     _____     _____

Name of person taking consent     Date              Signature

(if different from researcher)

## Appendix 7: Behavioural questionnaires, additional regression models

1. Strengths and difficulties, emotion
2. ADHD, overactivity
3. ADHD, impulsivity
4. Social communication questionnaire

Tables are only displayed for significant interactions.

1. Strength and difficulties, emotion

Table A1
*Linear Regression Model Fitting Information for the Strengths and Difficulties Questionnaire Emotion Above the 'High' Classification*

| Model | Model Fitting Criteria | Likelihood Ratio Tests | | |
|---|---|---|---|---|
| | -2 Log Likelihood | Chi-Square | *df* | Sig. |
| Intercept Only | 397.687 | | | |
| Final | 381.540 | 16.147 | 7 | .024 |

Note. See improved figure of -2 log likehood. Df=degrees of freedom.

Table A2
*Main Output from Multinomial Logistic Regression, Strengths and Difficulties Questionnaire (SDQ) Emotion Above the 'High' Classification*

| SDQ emotion ≤ High | *B* | Std. Error | Wald | *df* | Sig. | Exp(*B*) | 95% Confidence Interval for Exp(*B*) | |
|---|---|---|---|---|---|---|---|---|
| | | | | | | | Lower Bound | Upper Bound |
| Intercept | 1.806 | 1.760 | 1.053 | 1 | .305 | | | |
| Child Gender | -.416 | .264 | 2.478 | 1 | .115 | .660 | .393 | 1.107 |
| Child Age | -.037 | .174 | .044 | 1 | .833 | .964 | .685 | 1.357 |
| Breast Fed | .273 | .273 | 1.007 | 1 | .316 | 1.315 | .771 | 2.243 |
| Mother Age | .087 | .128 | .460 | 1 | .498 | 1.090 | .849 | 1.400 |
| Social Deprivation | .259 | .092 | 7.904 | 1 | .005* | 1.296 | 1.082 | 1.553 |
| [Normal GTF] | -.465 | .354 | 1.727 | 1 | .189 | .628 | .314 | 1.257 |
| [Treated SGTF] | -.459 | .398 | 1.325 | 1 | .250 | .632 | .290 | 1.380 |
| [Untreated SGTF] | 0 | . | . | 0 | . | . | . | . |

Note. The reference category was scores ≥ High. *Significance < .05. B=beta, df=degrees of freedom, SGTF=suboptimal gestational thyroid function.

Table B1

*Linear Regression Model Fitting Information for the Strengths and Difficulties Questionnaire Emotion Above the 'High' Classification*

| Model | Model Fitting Criteria | Likelihood Ratio Tests | | |
|---|---|---|---|---|
| | -2 Log Likelihood | Chi-Square | *df* | Sig. |
| Intercept Only | 318.075 | | | |
| Final | 298.795 | 19.280 | 6 | .004 |

Note. See improved figure of -2 log likehood. Df=degrees of freedom.

Table B2

*Main Output from Multinomial Logistic Regression, Strengths and Difficulties Questionnaire (SDQ) Emotion Above the 'High' Classification*

| SDQ emotion ≤ High | *B* | Std. Error | Wald | *df* | Sig. | Exp(*B*) | 95% Confidence Interval for Exp(*B*) | |
|---|---|---|---|---|---|---|---|---|
| | | | | | | | Lower Bound | Upper Bound |
| Intercept | 2.953 | 2.016 | 2.145 | 1 | .143 | | | |
| Child Gender | -.700 | .303 | 5.333 | 1 | .021 | .497 | .274 | .899 |
| Child Age | -.180 | .204 | .778 | 1 | .378 | .836 | .561 | 1.245 |
| Breast Fed | .287 | .304 | .890 | 1 | .345 | 1.332 | .734 | 2.416 |
| Mother Age | .058 | .142 | .166 | 1 | .683 | 1.060 | .802 | 1.400 |
| Social Deprivation | .330 | .104 | 10.055 | 1 | .002* | 1.390 | 1.134 | 1.705 |
| [Normal GTF] | .050 | .318 | .025 | 1 | .874 | 1.052 | .563 | 1.963 |
| [Treated SGTF] | 0 | . | . | 0 | . | . | . | . |

Note. The reference category was scores ≥ High. *Significance < .05. B=beta, df=degrees of freedom, SGTF=suboptimal gestational thyroid function.

2. ADHD, overactivity

Table C1
*Linear Regression Model Fitting Information for the Child Attention Deficit Hyperactivity Questionnaire, Overactivity Domain above one Standard Deviation*

| Model | Model Fitting Criteria | Likelihood Ratio Tests | | |
|---|---|---|---|---|
| | -2 Log Likelihood | Chi-Square | *df* | Sig. |
| Intercept Only | 332.589 | | | |
| Final | 305.344 | 27.245 | 6 | .000 |

Note. See improved figure of -2 log likehood. Df=degrees of freedom.

Table C2
*Main Output from Multinomial Logistic Regression, Child Attention Deficit Hyperactivity (ADHD) Questionnaire, Overactivity Domain above one Standard Deviation (SD)*

| ADHD Overactivity ≤ 1 SD | *B* | Std. Error | Wald | *df* | Sig. | Exp(*B*) | 95% Confidence Interval for Exp(*B*) | |
|---|---|---|---|---|---|---|---|---|
| | | | | | | | Lower Bound | Upper Bound |
| Intercept | -5.257 | 1.902 | 7.637 | 1 | .006 | | | |
| Child Gender | .774 | .301 | 6.621 | 1 | .010* | 2.168 | 1.203 | 3.910 |
| Child Age | .457 | .190 | 5.793 | 1 | .016 | 1.579 | 1.088 | 2.289 |
| Breast Fed | .296 | .300 | .974 | 1 | .324 | 1.345 | .747 | 2.422 |
| Mother Age | .100 | .142 | .494 | 1 | .482 | 1.105 | .837 | 1.458 |
| Social Deprivation | .144 | .103 | 1.966 | 1 | .161 | 1.155 | .944 | 1.412 |
| [Normal GTF] | .707 | .295 | 5.737 | 1 | .017* | 2.027 | 1.137 | 3.615 |
| [Treated SGTF] | 0 | . | . | 0 | . | . | . | . |

Note. The reference category was scores ≥ 1 SD. *Significance < .05. B=beta, df=degrees of freedom, SGTF=suboptimal gestational thyroid function.

Table D1

*Linear Regression Model Fitting Information for the Child*
*Attention Deficit Hyperactivity Questionnaire, Overactivity*
*Domain above one Standard Deviation*

| Model | Model Fitting Criteria | Likelihood Ratio Tests | | |
|---|---|---|---|---|
| | -2 Log Likelihood | Chi-Square | *df* | Sig. |
| Intercept Only | 448.405 | | | |
| Final | 413.029 | 35.377 | 7 | .000 |

Note. See improved figure of -2 log likehood. Df=degrees of freedom.

Table D2

*Main Output from Multinomial Logistic Regression, Child Attention Deficit Hyperactivity*
*(ADHD) Questionnaire, Overactivity Domain above one Standard Deviation (SD)*

| ADHD Overactivity ≤ 1 SD | *B* | Std. Error | Wald | *df* | Sig. | Exp(*B*) | 95% Confidence Interval for Exp(*B*) | |
|---|---|---|---|---|---|---|---|---|
| | | | | | | | Lower Bound | Upper Bound |
| Intercept | -4.512 | 1.614 | 7.809 | 1 | .005 | | | |
| Child Gender | .755 | .251 | 9.073 | 1 | .003* | 2.128 | 1.302 | 3.478 |
| Child Age | .379 | .157 | 5.810 | 1 | .016* | 1.461 | 1.073 | 1.990 |
| Breast Fed | .374 | .256 | 2.141 | 1 | .143 | 1.454 | .881 | 2.401 |
| Mother Age | .111 | .121 | .841 | 1 | .359 | 1.117 | .882 | 1.416 |
| Social Deprivation | .168 | .087 | 3.775 | 1 | .052 | 1.184 | .999 | 1.403 |
| [Normal GTF] | .567 | .306 | 3.418 | 1 | .064 | 1.762 | .966 | 3.213 |
| [Treated SGTF] | -.155 | .323 | .230 | 1 | .632 | .856 | .454 | 1.614 |
| [Untreated SGTF] | 0 | . | . | 0 | . | . | . | . |

Note. The reference category was scores ≥ 1 SD. *Significance < .05. B=beta, df=degrees of freedom, SGTF=suboptimal gestational thyroid function.

3. ADHD, Impulsivity

Table E1

*Linear Regression Model Fitting Information for the Child Attention Deficit Hyperactivity Questionnaire, Impulsivity Domain above one Standard Deviation*

| Model | Model Fitting Criteria | Likelihood Ratio Tests | | |
|---|---|---|---|---|
| | -2 Log Likelihood | Chi-Square | *df* | Sig. |
| Intercept Only | 319.575 | | | |
| Final | 296.949 | 22.626 | 7 | .002 |

Note. See improved figure of -2 log likehood. Df=degrees of freedom.

Table E2

*Main Output from Multinomial Logistic Regression, Child Attention Deficit Hyperactivity (ADHD) Questionnaire, Impulsivity Domain above one Standard Deviation (SD)*

| ADHD Impulsivity ≤ 1 (SD) | *B* | Std. Error | Wald | *df* | Sig. | Exp(*B*) | 95% Confidence Interval for Exp(*B*) | |
|---|---|---|---|---|---|---|---|---|
| | | | | | | | Lower Bound | Upper Bound |
| Intercept | -1.484 | 1.964 | .571 | 1 | .450 | | | |
| Child Gender | .258 | .305 | .713 | 1 | .398 | 1.294 | .712 | 2.353 |
| Child Age | .223 | .194 | 1.321 | 1 | .250 | 1.250 | .854 | 1.828 |
| Breast Fed | .718 | .315 | 5.195 | 1 | .023* | 2.051 | 1.106 | 3.804 |
| Mother Age | .302 | .154 | 3.851 | 1 | .050* | 1.353 | 1.000 | 1.830 |
| Social Deprivation | .053 | .106 | .248 | 1 | .619 | 1.054 | .856 | 1.299 |
| [Normal GTF] | .094 | .407 | .053 | 1 | .817 | 1.098 | .495 | 2.438 |
| [Treated SGTF] | -.652 | .412 | 2.505 | 1 | .113 | .521 | .232 | 1.168 |
| [Untreated SGTF] | 0 | . | . | 0 | . | . | . | . |

Note. The reference category was scores ≥ 1 SD. *Significance < .05. B=beta, df=degrees of freedom, SGTF=suboptimal gestational thyroid function.

Table F1

*Linear Regression Model Fitting Information for the Child
Attention Deficit Hyperactivity Questionnaire, Impulsivity
Domain above one Standard Deviation*

| Model | Model Fitting Criteria | Likelihood Ratio Tests | | |
|---|---|---|---|---|
| | -2 Log Likelihood | Chi-Square | *df* | Sig. |
| Intercept Only | 250.467 | | | |
| Final | 229.956 | 20.512 | 6 | .002 |

Note. See improved figure of -2 log likehood. Df=degrees of
freedom.

Table F2

*Main Output from Multinomial Logistic Regression, Child Attention Deficit Hyperactivity
(ADHD) Questionnaire, Impulsivity Domain above one Standard Deviation (SD)*

| ADHD Impulsivity ≤ 1 (SD) | *B* | Std. Error | Wald | *df* | Sig. | Exp(*B*) | 95% Confidence Interval for Exp(*B*) | |
|---|---|---|---|---|---|---|---|---|
| | | | | | | | Lower Bound | Upper Bound |
| Intercept | -2.305 | 2.262 | 1.039 | 1 | .308 | | | |
| Child Gender | .186 | .350 | .283 | 1 | .595 | 1.204 | .607 | 2.391 |
| Child Age | .255 | .228 | 1.247 | 1 | .264 | 1.291 | .825 | 2.019 |
| Breast Fed | .761 | .354 | 4.625 | 1 | .032* | 2.140 | 1.070 | 4.282 |
| Mother Age | .369 | .175 | 4.440 | 1 | .035* | 1.447 | 1.026 | 2.040 |
| Social Deprivation | .003 | .122 | .000 | 1 | .982 | 1.003 | .789 | 1.275 |
| [Normal GTF] | .723 | .348 | 4.311 | 1 | .038* | 2.060 | 1.041 | 4.075 |
| [Treated SGTF] | 0 | . | . | 0 | . | . | . | . |

Note. The reference category was scores ≥ 1 SD. *Significance < .05. B=beta, df=degrees
of freedom, SGTF=suboptimal gestational thyroid function.

4. Social communication questionnaire

Table G1
*Linear Regression Model Fitting Information for the Social Communication Questionnaires, scores above cut off ≥15*

| Model | Model Fitting Criteria | Likelihood Ratio Tests | | |
|---|---|---|---|---|
| | -2 Log Likelihood | Chi-Square | *df* | Sig. |
| Intercept Only | 94.879 | | | |
| Final | 86.702 | 8.177 | 6 | .225 |

Note. See improved figure of -2 log likehood. Df=degrees of freedom.

Table G2
*Main Output from Multinomial Logistic Regression, the Social Communication Questionnaires (SCQ), scores above cut off ≥15*

| SCQ ≤ 15 | *B* | Std. Error | Wald | *df* | Sig. | Exp(*B*) | 95% Confidence Interval for Exp(*B*) | |
|---|---|---|---|---|---|---|---|---|
| | | | | | | | Lower Bound | Upper Bound |
| Intercept | .425 | 4.105 | .011 | 1 | .917 | | | |
| Child Gender | .517 | .661 | .611 | 1 | .434 | 1.677 | .459 | 6.125 |
| Child Age | .106 | .412 | .067 | 1 | .796 | 1.112 | .496 | 2.491 |
| Breast Fed | -.307 | .676 | .207 | 1 | .650 | .736 | .196 | 2.766 |
| Mother Age | -.220 | .308 | .513 | 1 | .474 | .802 | .439 | 1.466 |
| Social Deprivation | .368 | .213 | 2.979 | 1 | .084 | 1.446 | .951 | 2.196 |
| [Normal GTF] | 1.419 | .657 | 4.665 | 1 | .031* | 4.132 | 1.140 | 14.974 |
| [Treated SGTF] | 0 | . | . | 0 | . | . | . | . |

Note. The reference category was scores ≥ 15. *Significance < .05. B=beta, df=degrees of freedom, SGTF=suboptimal gestational thyroid function.

Table H1
*Linear Regression Model Fitting Information for the Social
Communication Questionnaires, scores above cut off ≥15*

| Model | Model Fitting Criteria | Likelihood Ratio Tests | | |
|---|---|---|---|---|
| | -2 Log Likelihood | Chi-Square | *df* | Sig. |
| Intercept Only | 107.875 | | | |
| Final | 96.758 | 11.117 | 7 | .134 |

Note. See improved figure of -2 log likehood. Df=degrees of
freedom.

Table H2
*Main Output from Multinomial Logistic Regression, the Social Communication
Questionnaires (SCQ), scores above cut off ≥15*

| SCQ ≤ 15 | *B* | Std. Error | Wald | *df* | Sig. | Exp(*B*) | 95% Confidence Interval for Exp(*B*) | |
|---|---|---|---|---|---|---|---|---|
| | | | | | | | Lower Bound | Upper Bound |
| Intercept | 2.917 | 4.201 | .482 | 1 | .488 | | | |
| Child Gender | .608 | .637 | .910 | 1 | .340 | 1.837 | .527 | 6.405 |
| Child Age | .007 | .408 | .000 | 1 | .986 | 1.007 | .453 | 2.239 |
| Breast Fed | -.147 | .633 | .054 | 1 | .816 | .863 | .250 | 2.984 |
| Mother Age | -.143 | .296 | .234 | 1 | .629 | .867 | .486 | 1.547 |
| Social Deprivation | .397 | .205 | 3.764 | 1 | .052 | 1.488 | .996 | 2.222 |
| [Normal GTF] | -.632 | 1.130 | .313 | 1 | .576 | .532 | .058 | 4.873 |
| [Treated SGTF] | -2.032 | 1.090 | 3.475 | 1 | .062 | .131 | .015 | 1.110 |
| [Untreated SGTF] | 0 | . | . | 0 | . | . | . | . |

Note. The reference category was scores ≥ 15. *Significance < .05. B=beta, df=degrees of
freedom, SGTF=suboptimal gestational thyroid function.

# Appendix 8: Revised CATS general questionnaire

# CATS GENERAL QUESTIONNAIRE

Please answer-
Section A: MOTHER-     General information about you.
Section B: MEDICAL-     Medical history, please answer if you HAVE NOT attended the Heath Hospital for a morning.
Section C: CHILD-     General information about your child in the study.

| Section A: MOTHER | | | | | | |
|---|---|---|---|---|---|---|
| A1. Name | | | | | | |
| A2. DOB | | | | | | |
| A3. No. of pregnancies | | | | | | |
| A4. No. of live children (with ages) | | | | | | |
| A5. Education; pls tick all that apply | None | | GCSE or equivalent | | A Level or equivalent | |
| | Degree | | Other (please specify) | | | |
| A6. Occupation (or details of last job) | | | | | | |
| A7. Cigarette smoker (per day) pls tick | Never | | Present 0-10 | 10-20 | | 20+ |
| A8. Ethnicity pls tick | White | | Asian | African | | Other |

| Section B: MEDICAL (please answer if you HAVE NOT attended the Heath Hospital for a morning) | |
|---|---|
| B1. Previous illnesses/pls give date: | |
| B2. Operations/pls give date: | |
| B3. Family history of thyroid disease? | Yes/No  (pls delete as appropriate) |
| B4. Previous thyroid hormone treatment/pls give date(s): | |
| B5. Current thyroid hormone treatment | Yes/No    Dose: |

| B6. Other current drug therapy | Name | Dose | Date started |
|---|---|---|---|
| | | | |
| | | | |
| | | | |
| | | | |
| | | | |

| B7. Other current illnesses | |
|---|---|
| B8. Other comments | |

**Please turn over**

| Section C: CHILD | | | | | | | |
|---|---|---|---|---|---|---|---|
| C1. Name | | | | | | | |
| C2. DOB | | | | | | | |
| C3. Sex | | | | | | | |
| C4. Birth weight | | | | | | | |
| C5. Gestational age | Full term (37-40 wks | | Preterm (32-36 wks) | | Very Preterm (<32 weeks) | | |
| C6. Mode of delivery | Normal | | | | Caesarean | | |
| C7. Medical complications in pregnancy | Diabetes | | | | | | Yes/No |
| | High BP | | | | | | Yes/No |
| | Pre-eclampsia | | | | | | Yes/No |
| | Anaemia | | | | | | Yes/No |
| | Other: | | | | | | |
| C8. Breast fed for more than 1 mth | Yes | | | No | | | |
| C9. Child's handedness | Right | | | Left | | | |
| C10. Language of school | English | | | Welsh | | | |
| C11. Language spoken at home | English | | Welsh | | Other | | |
| C12. Child's biological father's occupation (if known) | | | | | | | |
| C13. Child's biological father's height (if known) | | | | | | | |

## Appendix 9: Supplementary CATS I IQ statistics

For the following statistics, with a 5% two-sided significance level and 80% power, a sample of 50 from both the treated and untreated SGTF groups allowed a detection of a difference of 7.5 IQ points in mean IQ; and with a Cohen's d of 0.5 for the effect size.

The following analyses address different ranges of T4 and TSH. All analyses were of continuous data and were multivariate with unadjusted and adjusted means presented; controlled for child gender, mother age at time of consent into CATS I, and social deprivation.

### Overt Hypothyroidism

This analysis explored IQ results of children born to women who had abnormal T4 and TSH between those who were treated and untreated during their pregnancies; this was classified as any maternal T4 <2.5th percentile, and also having TSH >97.5th percentile. There were 37 participants in this category (18 from the untreated SGTF group). More recent guidelines indicate that overt hypothyroidism should also be diagnosed if an individual presents a TSH > 10 mIU/L with normal T4 levels (1). Within the CATS I cohort, this occurred in five individuals, two from the untreated SGTF group.

In total, with the above overt hypothyroidism classification, there were 22 from the treated SGTF group, and 20 from the untreated SGTF group populating this subgroup. CATS I IQs were compared to explore whether treatment for overt hypothyroidism had any impact on childhood cognition at age 3. Descriptive statistics are presented firstly (table 1) and means were compared for verbal, performance and full scale IQ by a MANOVA and MANCOVA to take account adjustments.

Table 1
*Intelligent Quotient (IQ) Means for Wechsler Preschool and Primary Scale of Intelligence (WPPSI)-III, UK at Age 3 in the those classified as having Overt Hypothyroidism*

|  | CATS GROUP | Adjusted Data* | |
|---|---|---|---|
|  |  | N | Mean |
| WPPSI Verbal IQ | Treated SGTF | 20 | 106.23 (11.75) |
|  | Untreated SGTF | 22 | 108.10 (14.43) |
| WPPSI Performance IQ | Treated SGTF | 20 | 105.64 (12.44) |
|  | Untreated SGTF | 22 | 108.25 (13.39) |
| WPPSI Full scale IQ | Treated SGTF | 20 | 106.91 (11.96) |
|  | Untreated SGTF | 22 | 109.45 (12.73) |

The unadjusted MANOVA identified no significant results between the groups; $\Lambda_{ROY}$ = .012, $F$ (3, 38) = .157, $p$ = .924, $\eta_p^2$ = .012. The adjusted MANCOVA also identified no significant IQ results between the groups; $\Lambda_{ROY}$ = .015, $F$ (3, 35) = .178, $p$ = .910, $\eta_p^2$ = .015.

From the results in this supplementary analysis, treatment of overt hypothyroidism was of no benefit to offspring IQ at age 3. Rates of miscarriage were explored in the treated and untreated SGTF groups, none were recorded; although 49 occurred in those with normal thyroid function during pregnancy.

*Subclinical hypothyroidism*

Within the CATS I UK cohort, there were a total of 267 mothers with subclinical hypothyroidism; T4 measurements at consent > 2.5th percentile and with a TSH > 97.5th percentile (untreated SGTF = 117, treated SGTF = 150). See table 2 for further descriptive details.

Table 2
*Intelligent Quotient (IQ) Means for Wechsler Preschool and Primary Scale of Intelligence (WPPSI)-III, UK at Age 3 in the those classified as having Subclinical Hypothyroidism*

|  | CATS GROUP | Adjusted Data* | |
|---|---|---|---|
|  |  | N | Mean |
| WPPSI Verbal IQ | Treated SGTF | 150 | 110.87 (10.81) |
|  | Untreated SGTF | 116 | 108.57 (12.63) |
| WPPSI Performance IQ | Treated SGTF | 150 | 107.75 (14.33) |
|  | Untreated SGTF | 116 | 106.96 (14.44) |
| WPPSI Full scale IQ | Treated SGTF | 150 | 110.70 (12.17) |
|  | Untreated SGTF | 116 | 109.00 (13.54) |

The unadjusted MANOVA identified a non-significant difference between the treated and untreated SGTF subclinical hypothyroid group; $\Lambda$ROY = .018, $F$ (3, 263) = 1.570, $p$ = .197, $\eta_p^2$

= .018. The adjusted MANCOVA was also non-significant; $\Lambda$ROY = .019, $F$ (3, 259) = 1.624, $p$ = .184, $\eta_p^2$ = .018.

Within this subclinical hypothyroid group, there were some participants with T4 in the lower ranges (2.5-10th percentiles), of which 25 were from the untreated SGTF and 30 from the treated SGTF group. See table 3 for further descriptive statistics.

Table 3
*Intelligent Quotient (IQ) Means for Wechsler Preschool and Primary Scale of Intelligence (WPPSI)-III, UK at Age 3 in the those classified as having Subclinical Hypothyroidism with low T4*

|  | CATS GROUP | Adjusted Data* | |
|  |  | N | Mean |
| WPPSI Verbal IQ | Treated SGTF | 30 | 112.57 (11.49) |
|  | Untreated SGTF | 25 | 107.20 (13.60) |
| WPPSI Performance IQ | Treated SGTF | 30 | 109.47 (15.98) |
|  | Untreated SGTF | 25 | 109.88 (12.54) |
| WPPSI Full scale IQ | Treated SGTF | 30 | 112.83 (13.61) |
|  | Untreated SGTF | 25 | 109.88 (12.65) |

Note. *Adjusted for child gender, mother age at time of consent into CATS I, and social deprivation quintile. Standard deviations appear in parentheses below means.

The unadjusted MANOVA identified a non-significant difference between the treated and untreated SGTF subclinical hypothyroid, low T4, group; $\Lambda$ROY = .064, $F$ (3, 51) = 1.091, $p$ = .362, $\eta_p^2$ = .060. The adjusted MANCOVA was also non-significant; $\Lambda$ROY = .074, $F$ (3, 48) = 1.181, $p$ = .327, $\eta_p^2$ = .069.

These exploratory statistics revealed, that within the CATS I UK cohort, treatment made no benefit to IQ scores at age 3 of offspring born to mothers with subclinical hypothyroidism.

*Hypothyroxinemia*

Hypothyroxinemia was also explored in the CATS I cohort. When exploring T4 < 10th percentile with TSH < 97.5[th] percentile, 259 participants populated this group (untreated SGTF = 125, treated SGTF = 134). See table 4 for further descriptive statistics.

Table 4

*Intelligent Quotient (IQ) Means for Wechsler Preschool and Primary Scale of Intelligence (WPPSI)-III, UK at Age 3 in the those classified as having Hypothyroxinemia*

| | CATS GROUP | Adjusted Data* | |
|---|---|---|---|
| | | N | Mean |
| WPPSI Verbal IQ | Treated SGTF | 134 | 106.62 (10.69) |
| | Untreated SGTF | 125 | 106.20 (12.65) |
| WPPSI Performance IQ | Treated SGTF | 134 | 104.76 (12.73) |
| | Untreated SGTF | 125 | 104.39 (13.94) |
| WPPSI Full scale IQ | Treated SGTF | 134 | 106.55 (11.09) |
| | Untreated SGTF | 125 | 106.10 (13.01) |

Note. *Adjusted for child gender, mother age at time of consent into CATS I, and social deprivation quintile. Standard deviations appear in parentheses below means.

The unadjusted MANOVA identified a non-significant difference between the treated and untreated SGTF hypothyroxinemic group (T4 <10th percentile); $\Lambda$ROY = .001, $F$ (3, 255) = .034, $p$ = .992, $\eta_p^2$ = .001. The adjusted MANCOVA was also non-significant; $\Lambda$ROY = .001, $F$ (3, 252) = .060, $p$ = .981, $\eta_p^2$ = .001.

T4 < 2.5th percentile in the hypothyroxinemic group identified a reduced dataset of 237 (untreated SGTF 116, treated SGTF = 121). See table 5 for further descriptive details.

Table 5
*Intelligent Quotient (IQ) Means for Wechsler Preschool and Primary Scale of Intelligence (WPPSI)-III, UK at Age 3 in the those classified as having Hypothyroxinemia with T4 < 2.5th percentile*

| | CATS GROUP | Adjusted Data* | |
|---|---|---|---|
| | | N | Mean |
| WPPSI Verbal IQ | Treated SGTF | 121 | 106.92 (10.63) |
| | Untreated SGTF | 116 | 105.97 (12.78) |
| WPPSI Performance IQ | Treated SGTF | 121 | 104.93 (12.73) |
| | Untreated SGTF | 116 | 104.17 (14.19) |
| WPPSI Full scale IQ | Treated SGTF | 121 | 106.81 (11.32) |

| | | | |
|---|---|---|---|
| Untreated SGTF | 116 | 105.86 (13.24) | |

Note. *Adjusted for child gender, mother age at time of consent into CATS I, and social deprivation quintile. Standard deviations appear in parentheses below means.

The unadjusted MANOVA identified a non-significant difference between the treated and untreated SGTF hypothyroxinemic group (T4 <2.5$^{th}$ percentile); $\Lambda$ROY = .002, $F$ (3, 233) = .186, $p$ = .906, $\eta_p^2$ = .002. The adjusted MANCOVA was also non-significant; $\Lambda$ROY = .003, $F$ (3, 230) = .218, $p$ = .884, $\eta_p^2$ = .003.

These exploratory statistics revealed, that within the CATS I UK cohort, treatment made no benefit to IQ scores at age 3 of offspring born to mothers with hypothyroxinemia.

*Sub-group of CATS I; subclinical hypothyroidism and hypothyroxinemia*

A large sub-group comprised of hypothyroxinemia (T4 < 10$^{th}$ percentile and TSH < 97.5$^{th}$ percentile) and subclinical hypothyroidism (T4 2.5-10$^{th}$ percentile and TSH > 97.5$^{th}$ percentile) was populated by 314 participants (untreated SGTF = 150, treated SGTF = 164). This sub-group had age 3 IQs compared by unadjusted (MANOVA) and adjusted (MANCOVA) multivariate statistics; adjustments were made for child gender, mother age at time of consent into CATS I, and social deprivation. Descriptive statistics were presented firstly in table 6.

Table 6

*Intelligent Quotient (IQ) Means for Wechsler Preschool and Primary Scale of Intelligence (WPPSI)-III, UK at Age 3 in the sub-group in CATS I UK data*

| | CATS GROUP | Adjusted Data* | |
|---|---|---|---|
| | | N | Mean |
| WPPSI Verbal IQ | Treated SGTF | 164 | 107.71 (11.04) |
| | Untreated SGTF | 150 | 106.37 (12.77) |
| WPPSI Performance IQ | Treated SGTF | 164 | 105.62 (13.45) |
| | Untreated SGTF | 150 | 105.31 (13.83) |
| WPPSI Full scale IQ | Treated SGTF | 164 | 107.70 (11.80) |
| | Untreated SGTF | 150 | 106.73 (12.99) |

Note. *Adjusted for child gender, mother age at time of consent into CATS I, and social deprivation quintile. Standard deviations appear in parentheses below means.

The unadjusted MANOVA identified a non-significant difference between the treated and untreated SGTF sub-group; $\Lambda$ROY = .003, $F$ (3, 310) = .359, $p$ = .782, $\eta_p^2$ = .003. The adjusted MANCOVA was also non-significant; $\Lambda$ROY = .005, $F$ (3, 307) = .466, $p$ = .706, $\eta_p^2$ = .005.

*Conclusion*

The exploratory analyses of difference reference ranges of T4 and TSH did not reveal any statistically significant results between those that were treated or untreated for underactive thyroid function during pregnancy. As no control group of normal thyroid function during pregnancy was tested at age 3, I am unable to conclude whether overt hypothyroidism, subclinical hypothyroidism or hypothyroxinemia had an effect on offspring IQ in the CATS I UK cohort. Haddow et al. (2) included a sub-group of women who were treated for gestational subclinical hypothyroidism, and compared to those that were untreated, there were no cognitive difference between the offspring. Willoughby et al. (3, 4) also included participants who were treated for gestational subclinical hypothyroidism, but these individuals were compared to participants with normal thyroid function; therefore, I am unable to compare results here.

These findings support the CATS I publication (5) and the continuous mean analyses in chapter 1.2., as further exploration into the SGTF groups revealed no significant differences for differing reference ranges of T4 and TSH.

*References*

1.      Garber JR, Cobin RH, Gharib H, Hennessey JV, Klein I, Mechanick JI, et al. Clinical practice guidelines for hypothyroidism in adults: cosponsored by the American Association of Clinical Endocrinologists and the American Thyroid Association. Thyroid : official journal of the American Thyroid Association. 2012;22(12):1200-35.
2.      Haddow JE, Palomaki GE, Allan WC, Williams JR, Knight GJ, Gagnon J, et al. Maternal thyroid deficiency during pregnancy and subsequent neuropsychological development of the child. The New England journal of medicine. 1999;341(8):549-55.
3.      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.
4.      Willoughby KA, McAndrews MP, Rovet JF. Effects of maternal hypothyroidism on offspring hippocampus and memory. Thyroid :

official journal of the American Thyroid Association. 2014;24(3):576-84.

5.      Lazarus JH, Bestwick JP, Channon S, Paradice R, Maina A, Rees R, et al. Antenatal thyroid screening and childhood cognitive function. The New England journal of medicine. 2012;366(6):493-501.

## Appendix 10: Supplementary CATS II behavioural questionnaire statistics

This supplementary analysis explored the effect of treatment during gestation in the treated SGTF group who participated in CATS II. Firstly, I analysed how the T4 and TSH changed over the three time points; consent into CATS I, 6 weeks post consent, and 30 weeks gestation. Descriptive statistics for T4 and TSH were presented firstly, followed by a repeated measures ANOVA to identify any differences between the three time points. Women from the treated SGTF group were identified as over-treated during their pregnancies if their T4 values were > 97.5th of the entire CATS cohort (T4 > 17.7 pmol/L) following initiation of levothyroxine therapy; i.e. at either 6 weeks post consent, or at 30 weeks gestation. These over-treated individuals were compared to the rest of the CATS II cohort for behavioural questionnaire scores.

### Thyroid function results during pregnancy

Table 1 displays the current reference ranges for thyroid function per trimester. CATS I took blood samples at the end of the first trimester, six weeks post this point (arguably in the second trimester), and at 30 weeks consent (third trimester). There is no data available for reference ranges at this time, however, as CATS was such a large cohort study, any reference ranges available would have been based on smaller cohorts (around 100 women).

Table 43

*Current Reference Ranges*

| Thyroid | Pregnancy | Normal range |
|---------|-----------|--------------|
| T4 | 1st trimester | 10.5-18.3 |
| | 2nd trimester | 9.5-15.9 |
| | 3rd trimester | 8.6-13.7 |
| TSH | 1st trimester | 0.09-2.84 |
| | 2nd trimester | 0.18-2.81 |
| | 3rd trimester | 0.30-2.92 |

### T4

Table 2 displays the descriptive statistics of T4 during pregnancy for the treated SGTF group. Table 3 summarises the results from the repeated measure ANOVA and identified that all three time points were significantly different from one another. Figure 1 is a boxplot of the mean T4 values at the three time points, with the current T4 trimester normal references

304

ranges in red. This boxplot highlights how at the second and third CATS I time points, women were potentially over-treated with levothyroxine.

Table 2

*Descriptive Statistics of T4 Throughout the Treated SGTF Mothers' Pregnancies*

| Time point | Mean | Std. Deviation | N |
|---|---|---|---|
| T4 measurement taken at consent into CATS I (1) | 11.9243 | 1.93103 | 115 |
| T4 measurement taken 6 weeks following consent (2) | 16.2800 | 2.94631 | 115 |
| T4 measurement taken at 30 weeks gestation (3) | 15.4800 | 2.33024 | 115 |

Table 3

*Pairwise Comparisons of T4 at the Three Time Points Throughout the Treated SGTF Mothers' Pregnancies*

| (I) factor1 | (J) factor1 | Mean Difference (I-J) | Std. Error | Sig.[**] | 95% Confidence Interval for Difference | |
|---|---|---|---|---|---|---|
| | | | | | Lower Bound | Upper Bound |
| 1 | 2 | -4.356[*] | .297 | .001 | -5.077 | -3.635 |
| | 3 | -3.556[*] | .258 | .001 | -4.182 | -2.929 |
| 2 | 1 | 4.356[*] | .297 | .001 | 3.635 | 5.077 |
| | 3 | .800[*] | .236 | .003 | .227 | 1.373 |
| 3 | 1 | 3.556[*] | .258 | .001 | 2.929 | 4.182 |
| | 2 | -.800[*] | .236 | .003 | -1.373 | -.227 |

Note. Based on estimated marginal means. *The mean difference is significant at the .05 level. **Adjustment for multiple comparisons: Bonferroni. 1= T4 measurement taken at consent into CATS I. 2= T4 measurement taken 6 weeks following consent. 3= T4 measurement taken at 30 weeks gestation.

Figure 1

*Boxplot of T4 Values during Pregnancy of the Treated SGTF Group*
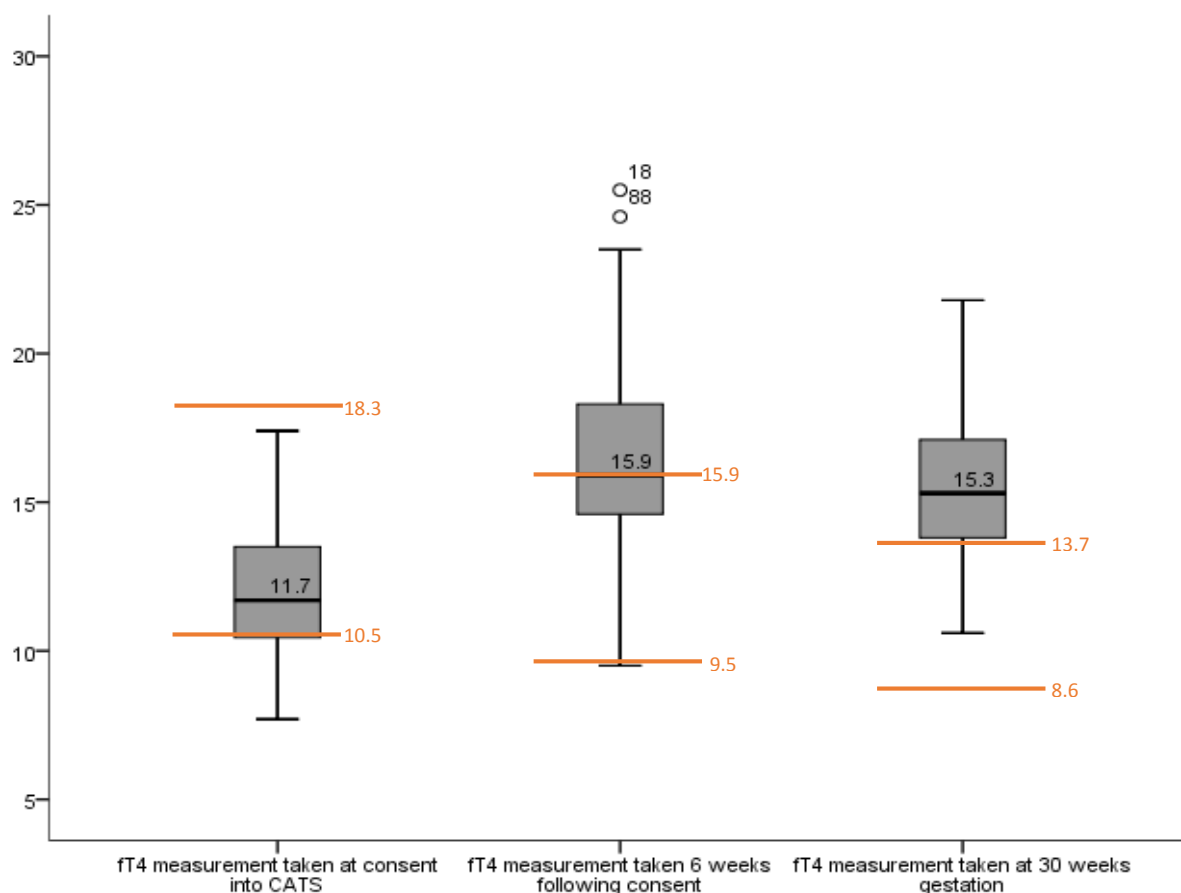


Note. Red bars display the current reference ranges per trimester for T4 values.

*TSH*

Table 4 displays the descriptive statistics of TSH during pregnancy for the treated SGTF group. Table 5 summarises the results from the repeated measure ANOVA and identified that only TSH at time of consent into CATS I was significantly different to time point 2 and 3; this indicates that TSH was corrected and reduced with levothyroxine therapy. As time points 2 and 3 were not significantly different, this displays how any adjustments to levothyroxine therapy at this time did not have a substantial effect on TSH; however, T4 did significantly change. Figure 2 displays the means of TSH presented by a box plot with normal current TSH reference ranges plotted; time points 2 and 3 in CATS I indicate women to be in the lower portion of the reference range. This suggests that dosage of levothyroxine was centred on TSH levels, rather than T4.

Table 4

*Descriptive Statistics of TSH Throughout the Treated SGTF Mothers' Pregnancies*

| Time point | Mean | Std. Deviation | N |
|---|---|---|---|
| TSH measurement taken at consent into CATS (1) | 4.218522 | 3.9876989 | 115 |
| TSH measurement taken 6 weeks following consent (2) | .693478 | .8730035 | 115 |
| TSH measurement taken at 30 weeks gestation (3) | .615565 | .8741725 | 115 |

Table 5

*Pairwise Comparisons of TSH at the Three Time Points Throughout the Treated SGTF Mothers' Pregnancies*

| (I) factor1 | (J) factor1 | Mean Difference (I-J) | Std. Error | Sig.** | 95% Confidence Interval for Difference | |
|---|---|---|---|---|---|---|
| | | | | | Lower Bound | Upper Bound |
| 1 | 2 | 3.525* | .342 | .001 | 2.694 | 4.356 |
| | 3 | 3.603* | .344 | .001 | 2.767 | 4.439 |
| 2 | 1 | -3.525* | .342 | .001 | -4.356 | -2.694 |
| | 3 | .078 | .066 | .714 | -.082 | .238 |
| 3 | 1 | -3.603* | .344 | .001 | -4.439 | -2.767 |
| | 2 | -.078 | .066 | .714 | -.238 | .082 |

Note. Based on estimated marginal means. *The mean difference is significant at the .05 level. **Adjustment for multiple comparisons: Bonferroni. 1= T4 measurement taken at consent into CATS I. 2= T4 measurement taken 6 weeks following consent. 3= T4 measurement taken at 30 weeks gestation.

Figure 2

*Boxplot of TSH Values during Pregnancy of the Treated SGTF Group*
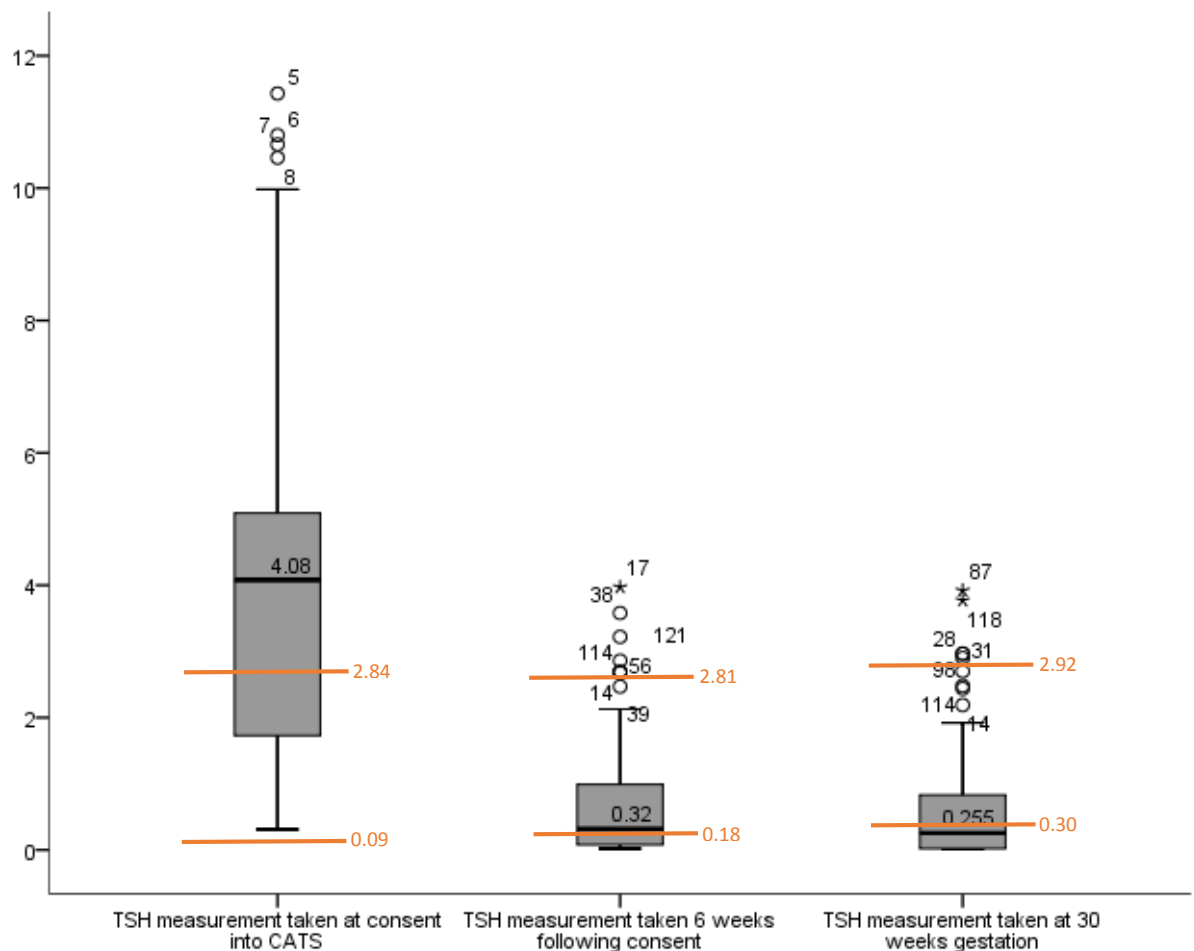


Note. As shown in the repeated measures ANOVA, there is no sig. difference between 6wks and 30wks. Red bars display the current reference ranges per trimester for T4 values.

*Over-treatment in CATS*

Those who were over-treated during their pregnancies (T4 > 17.7 pmol/L at 6 weeks post consent and 30 weeks gestation) were compared by a MANCOVA to the rest of the CATS II cohort and secondly to the rest of the treated SGTF group; adjustments were made for child gender, mother age at time of consent into CATS I, whether the mother breastfed over one month, and social deprivation.

There were 33 (28%) of the treated SGTF group that were over-treated during their pregnancies at 6 weeks post consent. Those offspring with maternal T4 > 17.7 pmol/L were compared to the rest of the CATS II cohort and were found to have significantly higher ADHD Overactivity scores ($p$ = .008 [95% CI: 0.322, 2.103]), and SCQ total score was not significantly

different (*p* = .269) (see table 6 for adjusted means and SDs). When comparing maternal T4 > 17.7 pmol/L within the treated SGTF group for ADHD Overactivity and SCQ total, the MANCOVA was not significant (Roy's largest root = .023, *F* (2,112) = 1.280, *p* = .282, $\eta_p^2$ = .022); see table 7 for adjusted means and SDs.

Table 6

*Mean scores of selected CATS II questionnaires; maternal T4 > 17.7 pmol/L at 6 weeks post initiation of levothyroxine therapy compared to the rest of the CATS II cohort*

|  | T4 > 17.7 pmol/L* | Mean | N |
|---|---|---|---|
| SCQ Mean | no** | 4.42 (3.74) | 432 |
|  | yes | 5.09 (4.86) | 33 |
|  | Total | 4.47 (3.83) | 465 |
| ADHD Overactivity Mean | no | 2.29 (2.52) | 432 |
|  | yes | 3.50 (3.50) | 33 |
|  | Total | 2.38 (2.62) | 465 |

Note. *At six weeks post initiation of levothyroxine therapy. **No=rest of CATS II study group. Standard deviations appear in parenthesis below means. SCQ= Social Communication Questionnaire.

Table 7

*Mean scores of selected CATS II questionnaires; maternal T4 > 17.7 pmol/L at 6 weeks post initiation of levothyroxine therapy within in the treated SGTF group*

|  | T4 > 17.7 pmol/L* | Mean | N |
|---|---|---|---|
| SCQ Mean | no** | 5.15 (4.63) | 86 |
|  | yes | 5.09 (4.86) | 33 |
|  | Total | 5.13 (4.68) | 119 |
| ADHD Overactivity Mean | no | 2.68 (2.86) | 86 |
|  | yes | 3.50 (3.50) | 33 |

| | | 2.90 | 119 |
|---|---|---|---|
| Total | | (3.06) | |

Note. *At six weeks post initiation of levothyroxine therapy. **No=rest of CATS II study group. Standard deviations appear in parenthesis below means. SCQ= Social Communication Questionnaire.

At 30 weeks gestation, 21 (18%) of the treated SGTF group were over-treated; 13 of which were also over-treated from 6 weeks post consent. ADHD Overactivity was significantly higher in the over-treated group compared to the rest of the CATS II cohort ($p$ = .004, [95% CI: 0.542, 2.746]), SCQ total score was not ($p$ = .241); see table 8 for means and SDs. At the multivariate level, SCQ total score and ADHD Overactivity were not significantly different between those who were over-treated during pregnancy at 30 weeks gestation, and the rest of the treated SGTF group (Roy's largest root = .033, $F$ (2,112) = 1.866, p = .160, $\eta_p^2$ = .032); see table 9 for means and SDs.

Table 8

*Mean scores of selected CATS II questionnaires; maternal T4 > 17.7 pmol/L at 30 weeks gestation compared to the rest of the CATS II cohort*

| | T4 > 17.7 pmol/L* | Mean | N |
|---|---|---|---|
| SCQ Mean | no** | 4.43 (3.74) | 444 |
| | yes | 5.31 (5.40) | 21 |
| | Total | 4.47 (3.83) | 465 |
| ADHD Overactivity Mean | no | 2.30 (2.54) | 444 |
| | yes | 4.00 (3.67) | 21 |
| | Total | 2.38 (2.62) | 465 |

Note. *At 30 weeks post initiation of levothyroxine therapy. **No=rest of CATS II study group. Standard deviations appear in parenthesis below means. SCQ= Social Communication Questionnaire.

Table 9

*Mean scores of selected CATS II questionnaires; maternal T4 > 17.7 pmol/L at 30 weeks post initiation of levothyroxine therapy within in the treated SGTF group*

| | T4 > 17.7 pmol/L* | Mean | N |
|---|---|---|---|
| SCQ Mean | no** | 5.05 (4.53) | 99 |
| | yes | 5.53 (5.44) | 20 |
| | Total | 5.13 (4.68) | 119 |
| ADHD Overactivitity Mean | no | 2.65 (2.87) | 99 |
| | yes | 4.15 (3.70) | 20 |
| | Total | 2.90 (3.06) | 119 |

Note. *At 30 weeks post initiation of levothyroxine therapy. **No=rest of CATS II study group. Standard deviations appear in parenthesis below means. SCQ= Social Communication Questionnaire.

### ADHD Overactivity prevalence

The prevalence of ADHD Overactivity was explored in this section. This included exploring scores > 1 SD and 2 SDs in participants from the treated SGTF group with maternal T4 > 17.7, compared to the untreated SGTF group, and selected individuals from the normal GTF group with T4 and TSH within the 10-90th percentiles.

Scores between these three groups were explored firstly by a chi-square, and significant differences were analysed by a logistic regression; this also allowed adjustment for the four covariates listed above. Following these analyses, mean scores of these three identified groups were explored by a ANOVA and ANCOVA; any significant difference identified were explored by a Bonferroni correction.

### ADHD Overactivity; prevalence >1SD

Table 10 displays the chi-square results of offspring who scored > 1 SD for ADHD overactivity in CATS II. An unadjusted significant difference ($p$ = .030) was identified between the three groups. This was explored further by regressions. Table 11 displays the model fitting data for the sub-group of the normal GTF compared to the T4 > 17.7 treated SGTF group; regression results following table 12, there was no significant difference between the groups. Table 13 displays the model fitting data for the T4 > 17.7 treated SGTF group compared to the untreated SGTF group; the regression results in table 14 identified a non-significant difference for ADHD Overactivity scores > 1 SD. A possible reason that there was a significant

difference for the chi-square, and not the regressions, was that the regressions took account of adjustments.

Table 10
*ADHD Overactivity scores > 1 SD per study group*

|  | ADHD Overactivity > 1 SD (%) |
|---|---|
| Treated SGTF* | 9 |
| (n =33) | (27%) |
| Untreated SGTF | 26 |
| (n = 106) | (24%) |
| Normal GTF** | 26 |
| (n = 189) | (14%) |
| Pearson Chi-Square | *p*=.030*** |

Note. Scores were from the Child ADHD Questionnaire. Percentages of scores per group are appear in parentheses below totals. *only those with T4 > 17.7. **only those with T4 and TSH between the 10-90th percentile. ***Significance < .05. SGTF=suboptimal gestational thyroid function.

Table 11

Table Displaying the Regression Model's Fit for the Data

| Model | Model Fitting Criteria | Likelihood Ratio Tests | | |
|---|---|---|---|---|
|  | -2 Log Likelihood | Chi-Square | *df* | Sig. |
| Intercept Only | 119.517 | | | |
| Final | 110.761 | 8.756 | 5 | .119 |

Note. See improved figure for -2 Log Likelihood. Df=degrees of freedom.

Table 12

*Main Output from Multinomial Logistic Regression, ADHD Overactivity > 1 SD; sub-group of the normal GTF compared to the T4 > 17.7 pmol/L treated SGTF*

| ADHD Overactivity > 1 SD | *B* | Std. Error | Wald | *df* | *Sig.* | Exp(*B*) | 95% Confidence Interval for Exp(*B*) | |
|---|---|---|---|---|---|---|---|---|
|  |  |  |  |  |  |  | Lower | Upper |
| Gender | -.626 | .391 | 2.565 | 1 | .109 | .535 | .249 | 1.150 |

| | B | Std. Error | Wald | df | Sig. | Exp(B) | Lower | Upper |
|---|---|---|---|---|---|---|---|---|
| Breastfed >1mns | -.427 | .393 | 1.185 | 1 | .276 | .652 | .302 | 1.408 |
| Mother age | .050 | .185 | .074 | 1 | .785 | 1.052 | .732 | 1.510 |
| Social deprivation | -.132 | .143 | .853 | 1 | .356 | .876 | .662 | 1.160 |
| [Normal GTF*] | -.879 | .462 | 3.616 | 1 | .057 | .415 | .168 | 1.027 |
| [Treated SGTF**] | 0 | . | . | 0 | . | . | . | . |

Note. The reference category was ADHD Overactivity < 1 SD. *sub-group of the normal GTF (10-90th percentile). **only those with T4 > 17.7 pmol/L. SGTF=suboptimal gestational thyroid function, B=beta, df=degrees of freedom.

Table 13

Table Displaying the Regression Model's Fit for the Data

| Model | Model Fitting Criteria | Likelihood Ratio Tests | | |
|---|---|---|---|---|
| | -2 Log Likelihood | Chi-Square | df | Sig. |
| Intercept Only | 121.058 | | | |
| Final | 111.314 | 9.744 | 5 | .083 |

Note. See improved figure for -2 Log Likelihood. Df=degrees of freedom.

Table 14

*Main Output from Multinomial Logistic Regression, ADHD Overactivity > 1 SD; T4 > 17.7 pmol/L treated SGTF compared to the untreated SGTF*

| ADHD Overactivity > 1 SD | B | Std. Error | Wald | df | Sig. | Exp(B) | 95% Confidence Interval for Exp(B) | |
|---|---|---|---|---|---|---|---|---|
| | | | | | | | Lower | Upper |
| Gender | -.617 | .401 | 2.362 | 1 | .124 | .540 | .246 | 1.185 |

| | B | SE | Wald | df | Sig | Exp(B) | Lower | Upper |
|---|---|---|---|---|---|---|---|---|
| Breastfed >1mns | -.749 | .440 | 2.896 | 1 | .089 | .473 | .200 | 1.120 |
| Mother age | -.035 | .205 | .029 | 1 | .864 | .966 | .647 | 1.442 |
| Social deprivation | -.215 | .148 | 2.118 | 1 | .146 | .807 | .604 | 1.077 |
| [Treated SGTF*] | .244 | .480 | .259 | 1 | .611 | 1.277 | .498 | 3.272 |
| [Untreated SGTF] | 0 | . | . | 0 | . | . | . | . |

Note. The reference category was ADHD Overactivity < 1 SD. *only those with T4 > 17.7 pmol/L. ***Significance < .05. SGTF=suboptimal gestational thyroid function, B=beta, df=degrees of freedom.

Those from the treated SGTF group with T4 > 17.7 pmol/L at 6 weeks post initiation of levothyroxine therapy had the highest percentage of offspring scoring > 1 SD for ADHD Overactivity. Scores were significantly different between groups however, following adjustments there was no difference.

*ADHD Overactivity; prevalence >2SD*

Table 15 displays the chi-square results of offspring who scored > 2 SD for ADHD overactivity in CATS II. An unadjusted significant difference ($p$ = .023) was identified between the three groups. This was explored further by regressions. Table 16 and 17 display the model fitting data and regression for the sub-group of the normal GTF compared to the T4 > 17.7 treated SGTF group. The regression revealed that the T4 > 17.7 treated SGTF group were 1.59 times more likely to score > 2 SD compared to the sub-group normal GTF group ($p$ = .015). Table 18 and 19 display the model fitting data for the T4 > 17.7 treated SGTF group compared to the untreated SGTF group and the regression results; which were non-significant.

Table 15
*ADHD Overactivity scores > 2 SD per study group*

| | ADHD Overactivity > 2 SD (%) |
|---|---|
| Treated SGTF* (n =33) | 5 (15%) |
| Untreated SGTF (n = 106) | 5 (5%) |
| Normal GTF** (n = 189) | 7 (4%) |
| Pearson Chi-Square | $p$=.023*** |

Note. Scores were from the Child ADHD Questionnaire. Percentages of scores per group are appear in parentheses below totals. *only those with T4 > 17.7 pmol/L. **only those with T4 and TSH between the 10-90th percentile. ***Significance < .05.  SGTF=suboptimal gestational thyroid function.

Table 16

Table Displaying the Regression Model's Fit for the Data

| Model | Model Fitting Criteria | Likelihood Ratio Tests | | |
|---|---|---|---|---|
| | -2 Log Likelihood | Chi-Square | *df* | Sig. |
| Intercept Only | 68.624 | | | |
| Final | 58.877 | 9.747 | 5 | .083 |

Note. See improved figure for -2 Log Likelihood. Df=degrees of freedom.

Table 17

*Main Output from Multinomial Logistic Regression, ADHD Overactivity > 2 SD; sub-group of the normal GTF compared to the T4 > 17.7 pmol/L treated SGTF*

| ADHD Overactivity > 2 SD | *B* | Std. Error | Wald | *df* | *Sig.* | Exp(*B*) | 95% Confidence Interval for Exp(*B*) | |
|---|---|---|---|---|---|---|---|---|
| | | | | | | | Lower | Upper |
| Gender | 1.198 | .709 | 2.855 | 1 | .091 | 3.314 | .826 | 13.298 |
| Breastfed >1mns | .287 | .647 | .197 | 1 | .657 | 1.332 | .375 | 4.737 |
| Mother age | .018 | .310 | .003 | 1 | .954 | 1.018 | .554 | 1.870 |
| Social deprivation | .101 | .232 | .191 | 1 | .662 | 1.106 | .702 | 1.743 |
| [Normal GTF*] | 1.588 | .656 | 5.866 | 1 | .015*** | 4.896 | 1.354 | 17.703 |
| [Treated SGTF**] | 0 | . | . | 0 | . | . | . | . |

Note. The reference category was ADHD Overactivity < 2 SD. *sub-group of the normal GTF (10-90th percentile). **only those with T4 > 17.7 pmol/L. ***Significance < 0.05. SGTF=suboptimal gestational thyroid function, B=beta, df=degrees of freedom.

Table 18

Table Displaying the Regression Model's Fit for the Data

| Model | Model Fitting Criteria | Likelihood Ratio Tests | | |
|---|---|---|---|---|
| | -2 Log Likelihood | Chi-Square | *df* | Sig. |
| Intercept Only | 59.849 | | | |
| Final | 50.221 | 9.629 | 5 | .086 |

Note. See improved figure for -2 Log Likelihood. Df=degrees of freedom.

Table 19

*Main Output from Multinomial Logistic Regression, ADHD Overactivity > 2 SD; T4 > 17.7 pmol/L treated SGTF compared to the untreated SGTF*

| ADHD Overactivity > 2 SD | *B* | Std. Error | Wald | *df* | *Sig.* | Exp(*B*) | 95% Confidence Interval for Exp(*B*) | |
|---|---|---|---|---|---|---|---|---|
| | | | | | | | Lower | Upper |
| Gender | .472 | .686 | .472 | 1 | .492 | 1.603 | .418 | 6.151 |
| Breastfed >1mns | .757 | .789 | .920 | 1 | .338 | 2.132 | .454 | 10.017 |
| Mother age | .638 | .425 | 2.253 | 1 | .133 | 1.893 | .823 | 4.356 |
| Social deprivation | .024 | .252 | .009 | 1 | .924 | 1.024 | .625 | 1.679 |
| [Treated SGTF*] | -1.199 | .702 | 2.915 | 1 | .088 | .302 | .076 | 1.194 |
| [Untreated SGTF] | 0 | . | . | 0 | . | . | . | . |

Note. The reference category was ADHD Overactivity < 2 SD. *only those with T4 > 17.7 pmol/L. ***Significance < .05. SGTF=suboptimal gestational thyroid function, B=beta, df=degrees of freedom.

Those from the treated SGTF group with T4 > 17.7 pmol/L at 6 weeks post initiation of levothyroxine therapy had the highest percentage of offspring scoring > 2 SD for ADHD Overactivity; around three times more participants meet the 2 SD threshold compared to the sub-group of the normal GTF and the untreated SGTF groups. The regressions identified that the T4 > 17.7 pmol/L treated SGTF group were significantly more likely to meet the threshold compared to the sub-group normal GTF; interactions to the untreated SGTF group were non-significant.

### Over-treated SGTF group mean ADHD Overactivity scores

As discussed in the main body of the thesis, binary outcomes are susceptible to type-1 errors, therefore I also explored continuous mean scores for significant differences between the T4 > 17.7 pmol/L treated SGTF, untreated SGTF and also the sub-group normal GTF group. Descriptive statistics of means and SDs are presented firstly (table 20), followed by the unadjusted ANOVA, and adjusted ANCOVA for ADHD Overactivity scores.

Table 20
*Mean scores for ADHD Overactivity in CATS II*

|  | CATS GROUP | N | Mean |
| --- | --- | --- | --- |
| ADHD Overactivity Mean | Treated SGTF* | 33 | 3.50 (3.50) |
|  | Untreated SGTF | 105 | 2.48 (2.69) |
|  | Normal GTF** | 189 | 2.12 (2.38) |

Note. *Only including those with T4 > 17.7 pmol/L at 6 weeks post initiation of levothyroxine therapy. **sub-group of the normal GTF; only including those with T4 and TSH between the 10-90th percentiles. Standard deviations appear in parentheses below means.

The unadjusted ANOVA revealed a significant difference between the groups for mean ADHD Overactivity score; $F (2, 325) = 4.061$, $p = .018$, $\eta_p^2 = .024$. Post hoc analysis (Bonferroni corrected) identified that the treated SGTF (only those with T4 > 17.7 pmol/L at 6 weeks post initiation of levothyroxine therapy) was significantly higher than the normal GTF sub-group; $p = .017$ (95% CI [.184, 2.231]); all other group interactions were $p > .05$. The adjusted ANCOVA was also significant; $F (2, 320) = 4.100$, $p = .017$, $\eta_p^2 = .025$. Similar to the ANOVA,

post hoc analysis revealed it was the treated SGTF to normal GTF sub-groups that had significantly different results; with the treated SGTF scoring higher ($p$ = .014, 95% CI [.214, 2.538]), see table 21 for further details.

Table 21

*Pairwise comparisons of adjusted ANCOVA exploring ADHD Overactivty mean scores between groups.*

| | | Mean Difference | Std. Error | Sig.** | 95% Confidence Interval for Difference | |
|---|---|---|---|---|---|---|
| | | | | | Lower Bound | Upper Bound |
| normal GTF | treated SGTF | -1.376 | .483 | .014* | -2.538 | -.214 |
| | untreated SGTF | -.297 | .312 | 1.000 | -1.048 | .453 |
| treated SGTF | normal GTF | 1.376 | .483 | .014* | .214 | 2.538 |
| | untreated SGTF | 1.078 | .509 | .105 | -.147 | 2.304 |
| untreated SGTF | normal GTF | .297 | .312 | 1.000 | -.453 | 1.048 |
| | treated SGTF | -1.078 | .509 | .105 | -2.304 | .147 |

Note. *The mean difference is significant at the .05 level. SGTF=suboptimal gestational thyroid function. **Adjustment for multiple comparisons: Bonferroni.

These statistics support the logistic regression exploring ADHD Overactivity scores > 2 SDs, as significant differences persistently appeared between the T4 > 17.7 pmol/L treated SGTF and the normal GTF sub-groups.

### Conclusion

Overall, we have identified within the CATS II cohort that treatment for underactive thyroid function during pregnancy has resulted in offspring having more ADHD Overactivity and autism-type behaviours, though not clinically significant. One of the reasons could be that just under a third of the mothers in CATS II were over-treated with levothyroxine during their pregnancies, which identifies a need for clinicians to closely monitor dosage levels during the gestation period.