# Thyroid Function in Early Pregnancy, Child IQ, and Autistic Traits: A Meta-Analysis of Individual Participant Data.

**Levie D**, Korevaar TIM, Bath SC, Dalmau-Bueno A, Murcia M, Espada M,

Dineva M, Ibarluzea JM, Sunyer J, Tiemeier H, Rebagliato M, Rayman MP,

Peeters RP*, Guxens M.*

* Denotes equal contribution

*J Clin Endocrinol Metab. 2018;103(8):2967-2979.*

# Abstract

**Context:** Low maternal free thyroxine (FT4) has been associated with poor child neurodevelopment in some single-center studies. Evidence remains scarce for the potential adverse effects of high FT4 and whether associations differ in countries with different iodine status.

**Objective:** To assess the association of maternal thyroid function in early pregnancy with child neurodevelopment in countries with a different iodine status.

**Design, Setting and Participants:** Meta-analysis of individual participant data from 9,036 mother-child pairs from three prospective population-based birth cohorts: INMA [INfancia y Medio Ambiente (Environment and Childhood project) (Spain)], Generation R (Netherlands) and ALSPAC (Avon Longitudinal Study of Parents and Children, United Kingdom). The exclusion criteria were multiple pregnancies, fertility treatments, thyroid interfering medication usage, and known thyroid disease.

**Main outcomes:** Child non-verbal IQ at 5 to 8 years of age, verbal IQ at 1.5 to 8 years of age, and autistic traits within the clinical range at 5 to 8 years of age.

**Results:** FT4 <2.5$^{th}$ percentile was associated with a 3.9-point (95% CI -5.7 to -2.2) lower non-verbal IQ and a 2.1-point (95% CI -4.0 to -0.1) lower verbal IQ. A suggestive association of hypothyroxinemia with a greater risk of autistic traits was observed. FT4 >97.5$^{th}$ percentile was associated with a 1.9-fold (95% CI 1.0 to 3.4) greater risk of autistic traits. No independent associations were found with TSH.

**Conclusions:** Low maternal FT4 was consistently associated with lower IQ across the cohorts. Further studies are needed to replicate the findings of autistic traits and investigate the potential modifying role of maternal iodine status. FT4 seems a reliable marker of fetal thyroid state in early pregnancy, regardless of the type of immunoassay.

# Introduction

Thyroid hormone regulates crucial processes of brain development, including the proliferation, migration, and differentiation of neuronal cells, as shown in animal studies [1,2]. Because the fetal thyroid gland is not functionally mature until approximately week 18 of pregnancy [3], the fetus is dependent on placental transfer of maternal thyroid hormone during this period. Adequate maternal thyroid hormone concentrations during early pregnancy are therefore essential for optimal fetal brain development.

Previous studies focused mainly on the possible adverse effects of low maternal hormone availability on fetal brain development. In several studies, either overt hypothyroidism or low maternal free T4 (FT4) was associated with lower child IQ [4–8] and lower gray-matter volume [4], a greater risk of autistic traits [9], impaired psychomotor function [10], and schizophrenia [11]. Although the association of high maternal FT4 on child neurodevelopment has been less well studied, experimental evidence from rodents has indicated that high hormone availability may also have adverse effects [12–18]. A recent study from the Netherlands has shown that high maternal FT4 is associated with lower IQ and gray matter volumes in the child [4]. However, it is unclear whether these findings from an iodine-sufficient population in the Netherlands [19] can be extrapolated to other countries with a different iodine status and whether high maternal FT4 is also associated with other adverse neurodevelopmental outcomes other than IQ.

Neither of the two randomized controlled trials that studied the effect of levothyroxine treatment in women with subclinical hypothyroidism or hypothyroxinemia on child IQ showed any benefit of treatment [20,21]. However, these negative results could be ascribable to a relatively late start of treatment in both trials (13 weeks and 16 to 18 weeks, respectively), a relatively high dose was given that might have led to potential overtreatment [20], or a lack of power to detect the expected 3- to 4-point difference in IQ [21,22]. Therefore, further studies are required to better quantify and replicate the potential effects of both low and high maternal thyroid hormone availability on fetal neurodevelopment. These studies can help improve the design of future controlled trials.

The aim of the present study was to investigate the association of maternal thyroid function in early pregnancy across the full range of FT4 and thyrotropin (TSH) concentrations with child's IQ and autistic traits in three prospective birth cohorts.

# Material and Methods

## Study design and populations

For the present study, we used individual participant data from three prospective population-based birth cohorts: INfancia y Medio Ambiente [INMA (Environment and Childhood project), Spain, three regions] [23], Generation R (the Netherlands) [24], and the Avon Longitu-

dinal Study of Parents and Children (ALSPAC, United Kingdom) [25]. In INMA, the eligible study participants were pregnant women with a singleton pregnancy residing in the regions of Valencia, Sabadell, and Gipuzkoa from November 2003 to January 2008. In Generation R, the eligible study participants were pregnant women living in the Rotterdam area with an expected delivery date from April 2002 to January 2006. In ALSPAC, the eligible study participants resided in a defined area in the southwest of England, with an expected date of delivery from April 1991 to December 1992 (the study website of ALSPAC contains details of all the data available through a fully searchable data dictionary [26]). For the present study, eligible women were enrolled in the three cohorts during the first half of pregnancy (≤18th week of gestation). Women with multiple pregnancies or fertility treatment and/or using medication affecting the thyroid or having a known thyroid disease were excluded (Fig. 1). Ethical approval was obtained from the local ethical committees at the time of study enrollment; all parents or guardians of the children provided informed consent.

## Thyroid function

Thyroid function was measured in serum samples stored at -80°C (INMA and Generation R) or -20°C (ALSPAC). The samples were obtained at early pregnancy [ (mean ± SD) gestational age: INMA, 13.1 ± 1.3 weeks; Generation R, 13.4 ± 2.0 weeks; ALSPAC, 11.0 ± 3.2 weeks] (Table 1). Different assays were used to measure FT4 and TSH (Supplemental Table 1). Although thyroid peroxidase antibody (TPOAb) was not measured in INMA, TPOAb measurements were available from Generation R and ALSPAC. The FT4 and TSH concentrations were logarithmically transformed, and cohort-specific SD scores were calculated with a mean of 0 and a SD of 1 based on the data of TPOAb-negative women, as advocated by the guidelines when defining population-based reference ranges [27].

Hypothyroxinemia [normal (2.5th-97.5$^{th}$ percentile) TSH, low (<2.5$^{th}$ percentile) FT4], subclinical hypothyroidism [high (>97.5$^{th}$ percentile) TSH, normal FT4], and subclinical hyperthyroidism (low TSH, normal FT4) were defined according to the 2.5th and the 97.5th population-based percentiles of the whole study population in INMA, because TPOAb measurements were not available. Thyroid disease entities were defined using the same population-based percentiles in Generation R and ALSPAC. However, in these cohorts, the population-based percentiles were based on the results from TPOAb-negative women. The reference group consisted of euthyroid women (TSH and FT4 between the 2.5th and 97.5th percentiles). Additionally, to improve the statistical power, we identified thyroid disease entities using the 5th and the 95th population-based percentiles. The untransformed 2.5$^{th}$ and 97.5$^{th}$ population-based percentiles. The untransformed 2.5th and 97.5th population-based percentiles based on TPOAb-negative women when possible were 0.14 and 3.86, 0.05 and 4.13, and 0.07 and 2.58 mIU/L for TSH and 8.4 and 14.0, 10.4 and 22.1, and 12.6 and 22.5 pmol/L for FT4 in INMA, Generation R, and ALSPAC, respectively.

## Non-verbal and verbal IQ

In INMA, non-verbal and verbal IQ were assessed by a psychologist at a median age of 4.6 years using the McCarthy Scales of Children´s Abilities [28]. In Generation R, non-verbal IQ was assessed by trained staff at a median age of 6.0 years using a subset of the Snijders Oomen Nonverbal Intelligence Test (2.5-7-Revised) [29], and verbal IQ was estimated by the parent-reported short form of the McArthur Communicative Development Inventory [30] obtained at a median age of 1.5 years. In ALSPAC, non-verbal and verbal IQ were assessed by trained staff at a median age of 8.6 years using the Wechsler Intelligence Scale For Children, third UK edition [31]. To homogenize the different scores, raw cohort-specific scores were standardized to a mean of 100 and a SD of 15 (new score = 100 + 15 x SD).

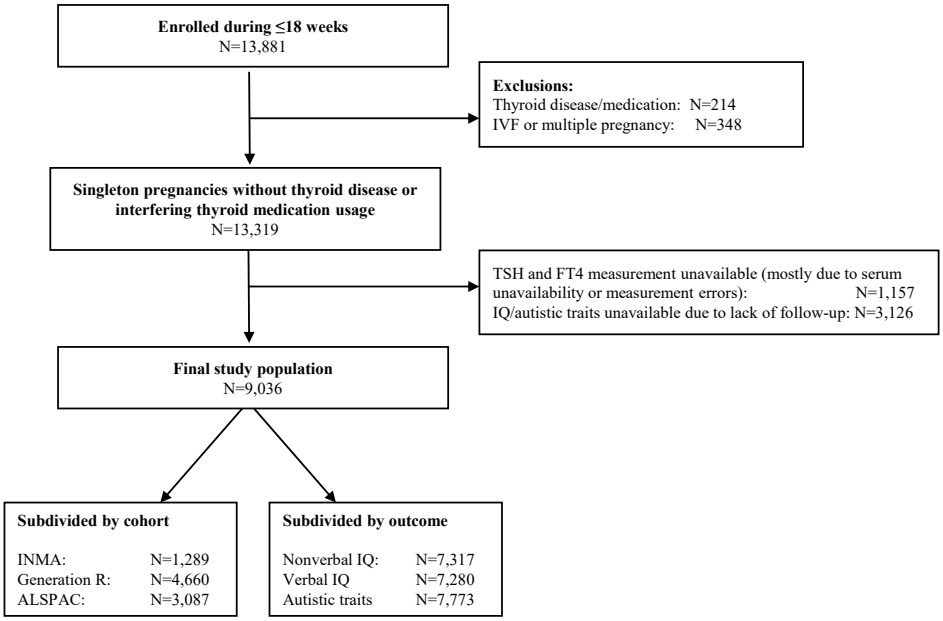



**Figure 1** Flowchart for the selection of the final study population. IVF, *in vitro* fertilization.

## Autistic traits within the clinical range

Autistic traits are symptoms that represent subclinical deficits in social behavior, communication, and or restricted, repetitive patterns of behavior common to ASD but that do not meet clinical ASD diagnosis [32]. Autistic traits within the clinical range were defined as the presence of an autistic traits score greater than the specific cutoff for each assessment tool, which had been previously validated in other studies to detect children at risk of ASD. In INMA, autistic traits were assessed with the Childhood Autism Spectrum Test (CAST) by a psychologist at a median age of 4.6 years, with a cutoff of ≥ 15 points to define autistic traits within the clinical range [33]. In Generation R, autistic traits were assessed using the Pervasive Developmental

Problems subscale of the Child Behavior Checklist for Toddlers (CBCL1½–5) by the parents at a median age of 5.9 years, with a cutoff of ≥98$^{th}$ percentile to define autistic traits within the clinical range [34]. In ALSPAC, autistic traits were assessed with the Social Communication Disorder Checklist (SCDC) by the parents at a median age of 7.6 years of age, with a cutoff of nine or more points to define autistic traits within the clinical range [35].

## Potential confounding variables

A direct acyclic graph [36] facilitated decision making regarding which covariates should be adjusted for in the analysis. Information on maternal variables [age, educational level (low, medium, high), ethnicity (cohort-specific categories), parity (zero, one, two or more), pre-pregnancy body mass index (BMI), and smoking during pregnancy (never smoked, smoked in the beginning or until pregnancy confirmed, continued smoking)] was collected during pregnancy using questionnaires. Gestational age at blood sampling was defined using ultrasonography or the last menstrual period. Child sex and age at IQ or autistic trait ascertainment were obtained during the study visits.

## Statistical analyses

We used linear regression models to study the association of maternal FT4, TSH, and thyroid disease entities with child non-verbal or verbal IQ. We used logistic regression models to study the association of maternal FT4, TSH, and thyroid disease entities with child autistic traits within the clinical range.

We studied these associations using a one-step and a two-step approach. In the one-step approach, we assessed non-linearity between FT4 and TSH and each outcome using restricted cubic splines with three to five knots. An ANOVA test was used to report an overall *P* value for the null hypothesis that the mean IQ or probability of autistic traits within the clinical range was similar across the whole distribution of TSH or FT4. In the two-step approach, we combined cohort-specific effect estimates of the association between FT4, TSH, and thyroid disease entities and each outcome using random-effects meta-analyses. For this analysis, FT4 and TSH concentrations were categorized as <2.5th, <5th, >95th, or >97.5th percentiles using women with values within the interquartile range (within the 25th and 75th percentile range) as the reference group. Compared with the one-step approach, the two-step approach allows for differences in participant characteristics between cohorts, and heterogeneity between cohorts can be calculated [37]. Heterogeneity was assessed using the Cochran Q test and the $I^2$ statistic [38]. All models were adjusted for maternal age, educational level, ethnicity, parity, pre-pregnancy BMI, smoking, gestational age at blood sampling, and child sex. Because one-step approach models could not be adjusted for age at IQ or autistic trait ascertainment, cohort, and ethnicity at the same time owing to collinearity, we adjusted them only for ethnicity. The two-step approach models could be adjusted for these variables because the effect-estimates were calculated separately by cohort.

**Table 1** Distribution of maternal and child characteristics.

| Variable | INMA (n=1,289) | Generation R (n=4,660) | ALSPAC (n=3,087) |
|---|---|---|---|
| Maternal TSH, median (IQR), mIU/L | 1.24 (0.84-1.81) | 1.36 (0.85-2.03) | 1.00 (0.64-1.46) |
| Maternal FT4, median (IQR), pmol/L | 10.6 (9.7-11.6) | 14.8 (13.2-16.7) | 16.2 (14.8-17.7) |
| Thyroid disease entities,[a] n (%) | | | |
| Hypothyroxinemia | 32 (2.5) | 111 (2.4) | 61 (2.0) |
| Subclinical hypothyroidism | 31 (2.4) | 140 (3.0) | 110 (3.6) |
| Subclinical hyperthyroidism | 20 (1.6) | 69 (1.5) | 34 (1.1) |
| TPOAb positivity, n (%) | NA | 254 (5.8) | 392 (12.8) |
| Gestational age at blood sampling, mean ± SD, wk | 13.1 (1.3) | 13.4 (2.0) | 11.0 (3.2) |
| Maternal educational level, n (%) | | | |
| Low | 281 (21.9) | 353 (8.0) | 736 (24.7) |
| Medium | 537 (41.8) | 1,904 (42.9) | 1,828 (61.3) |
| High | 468 (36.4) | 2,179 (49.1) | 416 (14.0) |
| Maternal ethnicity, n (%) | | | |
| Spanish | 1,202 (93.4) | NA | NA |
| Latin-American | 60 (4.7) | NA | NA |
| European/other | 25 (1.9) | NA | NA |
| Dutch | NA | 2,606 (56.7) | NA |
| Indonesian | NA | 150 (3.3) | NA |
| Cape Verdean | NA | 170 (3.7) | NA |
| Moroccan | NA | 225 (4.9) | NA |
| Dutch Antilles | NA | 104 (2.3) | NA |
| Surinamese | NA | 351 (7.6) | NA |
| Turkish | NA | 356 (7.8) | NA |
| Asian | NA | 51 (1.1) | NA |
| Other, non-Western | NA | 162 (3.5) | NA |
| Other, Western | NA | 418 (9.1) | NA |
| White | NA | NA | 2,924 (98.6) |
| Non-white | NA | NA | 42 (1.4) |
| Maternal age, mean ± SD, y | 31.5 (4.0) | 30.3 (4.8) | 28.0 (4.6) |
| Parity, n (%) | | | |
| 0 | 731 (56.8) | 2,721 (58.4) | 1,410 (47.2) |
| 1 | 472 (36.7) | 1,386 (29.7) | 1,033 (34.6) |
| ≥2 | 84 (6.5) | 553 (11.9) | 543 (18.2) |
| Maternal smoking, n (%) | | | |
| Never smoked | 883 (69.4) | 3,085 (73.5) | 2,391 (79.2) |
| Smoked at the beginning of pregnancy | 174 (13.7) | 396 (9.4) | 142 (4.7) |
| Continued smoking | 216 (17.0) | 719 (17.1) | 486 (16.1) |
| Pre-pregnancy BMI, median (IQR), kg/m$^2$ | 22.5 (20.8-25.1) | 22.6 (20.7-25.2) | 22.1 (20.5-24.2) |
| Child female sex, n (%) | 635 (49.3) | 2,313 (49.6) | 1,500 (48.6) |
| Child autistic traits within the clinical range, n (%) | 16 (1.4) | 117 (3.1) | 206 (7.5) |

Data might not sum to 100 because of rounding.

Abbreviations: BMI, body mass index; IQR, interquartile range; NA, not available.

[a] Based on the 2.5th and 97.5th population-based percentiles.

As a sensitivity analysis, we adjusted the analyses of autistic traits for non-verbal IQ, a language- and culture-free measure of cognitive ability. Additionally, when we observed associations between maternal TSH and child IQ or autistic traits, we repeated the analysis stratifying by low-, mid-, and high-normal FT4. Finally, all analyses were repeated in the TPOAb-negative women only.

We applied inverse probability weighting to correct for potential differential loss to follow-up [39]. We performed multiple imputation using chained equations to account for missing values for the potential confounding variables [40]. A total of 25 datasets were generated and analyzed using standard procedures for multiple imputation. Statistical analyses were performed in STATA, version 14.0 (StataCorp, College Station, TX) and R statistical software, version 3.3.2, package rms and lme4 (R Foundation, Vienna, Austria).

# Results

After exclusions, the final study population included 9036 mother-child pairs (Fig. 1), the characteristics of which are shown in Table 1. The mean maternal age varied across cohorts: 31.5 years in INMA, 30.3 years in Generation R, and 28.0 years in ALSPAC. The percentage of mothers who continued smoking during pregnancy was similar among the cohorts (~ 16% to 17%). Autistic traits within the clinical range occurred in 1.4% of the children in INMA, 3.1% in Generation R, and 7.5% in ALSPAC. The two most prevalent thyroid disease entities were hypothyroxinemia (2.0% to 2.5% across the cohorts) and subclinical hypothyroidism (2.4% to 3.6% across the cohorts). Compared with the final study population, the women not included in the analysis had a lower level of education, were less often native or white, and were younger in all three cohorts (Supplemental Table 2).

## Non-verbal IQ

We observed a statistically significant nonlinear association between maternal FT4 and mean non-verbal IQ (Fig. 2). FT4 $\leq 2.5^{th}$ percentile was associated with a 3.9-point (95% CI -5.7 to -2.3; $P<0.001$) lower non-verbal IQ. Similar results were observed when using the fifth percentile cut-off. A high FT4 was not associated with the non-verbal IQ. TSH $\geq 97.5^{th}$ and $\geq 95^{th}$ percentile was associated with a statistically non-significant slightly greater non-verbal IQ (1.5 points; 95% CI, -0.3 to 3.3; $P=0.100$; and 1.2 points, 95% CI, -0.1 to 2.5; $P=0.063$, respectively; Supplemental Fig.1). However, the sensitivity analysis showed that this association was driven by women with a FT4 concentration in the mid- or high-normal range (Supplemental Table 3). No heterogeneity was observed among the cohorts. Results remained similar after excluding TPOAb-positive women.

## Verbal IQ

A statistically non-significant linear association was found between maternal FT4 and mean verbal IQ (Fig. 3). FT4 ≤2.5$^{th}$ percentile was associated with a 2.1-point (95% CI, -4.0 to -0.1; *P*=0.039) lower verbal IQ. In contrast, the association of FT4 at the fifth percentile or less was associated with a statistically non-significant slightly lower verbal IQ (-1.4 points; 95% CI -2.9 to 0.2; *P*=0.078). A high FT4 was not associated with verbal IQ. A low TSH was also not associated with verbal IQ (Supplemental Fig. 2). TSH ≥97.5$^{th}$ percentile was associated with a greater verbal IQ (1.9 points; 95% CI, 0.1 to 3.7; *P*=0.039). However, no association was found for TSH ≥95$^{th}$ percentile. The sensitivity analysis showed that the positive association of a high TSH ≥97.5$^{th}$ percentile with verbal IQ was driven by women with a FT4 concentration in the mid- or high-normal range (Supplemental Table 4). No heterogeneity was observed among the cohorts. The results remained similar after excluding TPOAb-positive women.

## Autistic traits

No continuous association was found for maternal FT4 with child autistic traits (Fig.4). FT4 ≤2.5th percentile was not associated with autistic traits, but FT4 ≤5th percentile was associated with a statistically non-significant slightly higher risk of autistic traits [odds ratio (OR), 1.5; 95% CI, 1.0 to 2.3; *P*=0.080). FT4 ≥97.5th percentile was associated with a 1.9-fold (95% CI, 1.0 to 3.4; *P*=0.043) greater risk of autistic traits. A similar association was found after adjusting for non-verbal IQ (data not shown). FT4 ≥95$^{th}$ percentile was not associated with autistic traits. TSH was not associated with autistic traits (Supplemental Fig. 3). No heterogeneity was observed among the cohorts. The results remained similar after excluding TPOAb-positive women.

## Clinical disease entities

Highly similar results were obtained when FT4 and TSH were combined into clinical disease entities. Hypothyroxinemia, based on the 2.5th and 97.5th population-based percentiles, was associated with a 3.8-point (95% CI, -5.7 to -2.0; *P*<0.001) lower non-verbal IQ and a 2.8-point (95% CI, -4.8 to -0.7; *P*=0.007) lower verbal IQ (Supplemental Fig. 4) but was not associated with autistic traits. For hypothyroxinemia, based on the 5$^{th}$ and 95$^{th}$ population-based percentiles, similar results were found with non-verbal and verbal IQ, with a 1.8-fold (95% CI, 1.1 to 2.8; *P*=0.011) greater risk was found with autistic traits (Supplemental Fig. 4), which remained after adjusting for non-verbal IQ (data not shown).

Subclinical hypothyroidism, based on the 2.5th and 97.5th population-based percentiles, was associated with a 1.9-point (95% CI, 0.1 to 3.6; *P*=0.037) greater non-verbal IQ but not with verbal IQ or autistic traits (Supplemental Fig. 5). When defining subclinical hypothyroidism using the 5th and 95th population-based percentiles, the association with non-verbal IQ became statistically non-significant (1.3 points; 95% CI -0.2 to 2.9; *P*=0.096). Subclinical hyperthyroidism was not associated with non-verbal IQ, verbal IQ or autistic traits (Supplemental Fig. 6).

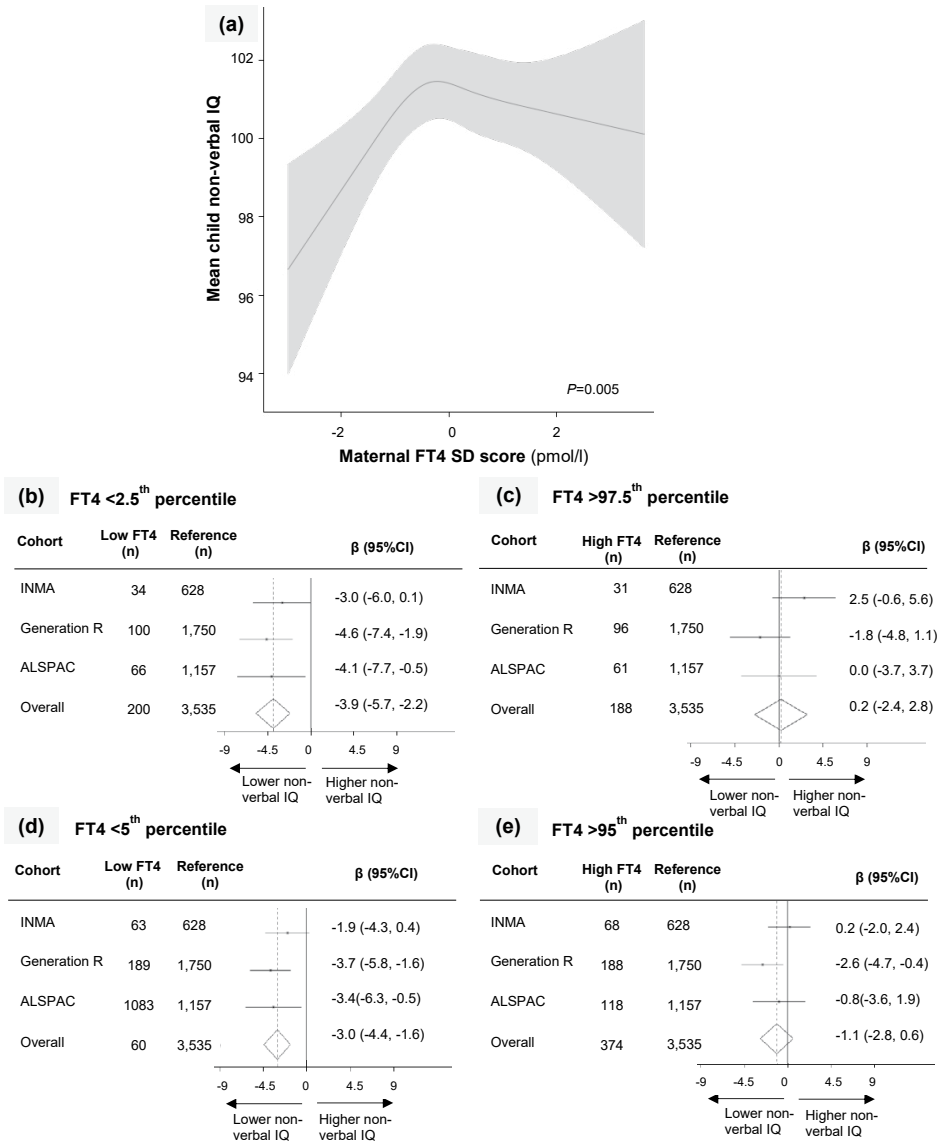



**Figure 2** Association of maternal FT4 during early pregnancy with child non-verbal IQ.
Association shown as a) a continuous association depicted as the mean child non-verbal IQ (black line) with 95% CI (gray area) and by cohort-specific maternal FT4 concentrations in the (b) <2.5th percentile, (c) >97.5th percentile, (d), <5th percentile, and (e) >95$^{th}$ percentile compared with interquartile range (between 25th and 75th percentiles), depicted as effect estimate (dot) with the 95% CI per cohort and overall as estimated by random-effects meta-analysis (diamond). The $I^2$ for each model is as follows: for FT4<2.5$^{th}$ percentile, $I^2$=0.0%; for FT4>97.5$^{th}$ percentile, $I^2$=48.5%; for FT4<5$^{th}$ percentile, $I^2$=0.0%; for FT4>95$^{th}$ percentile, $I^2$=37.8%.

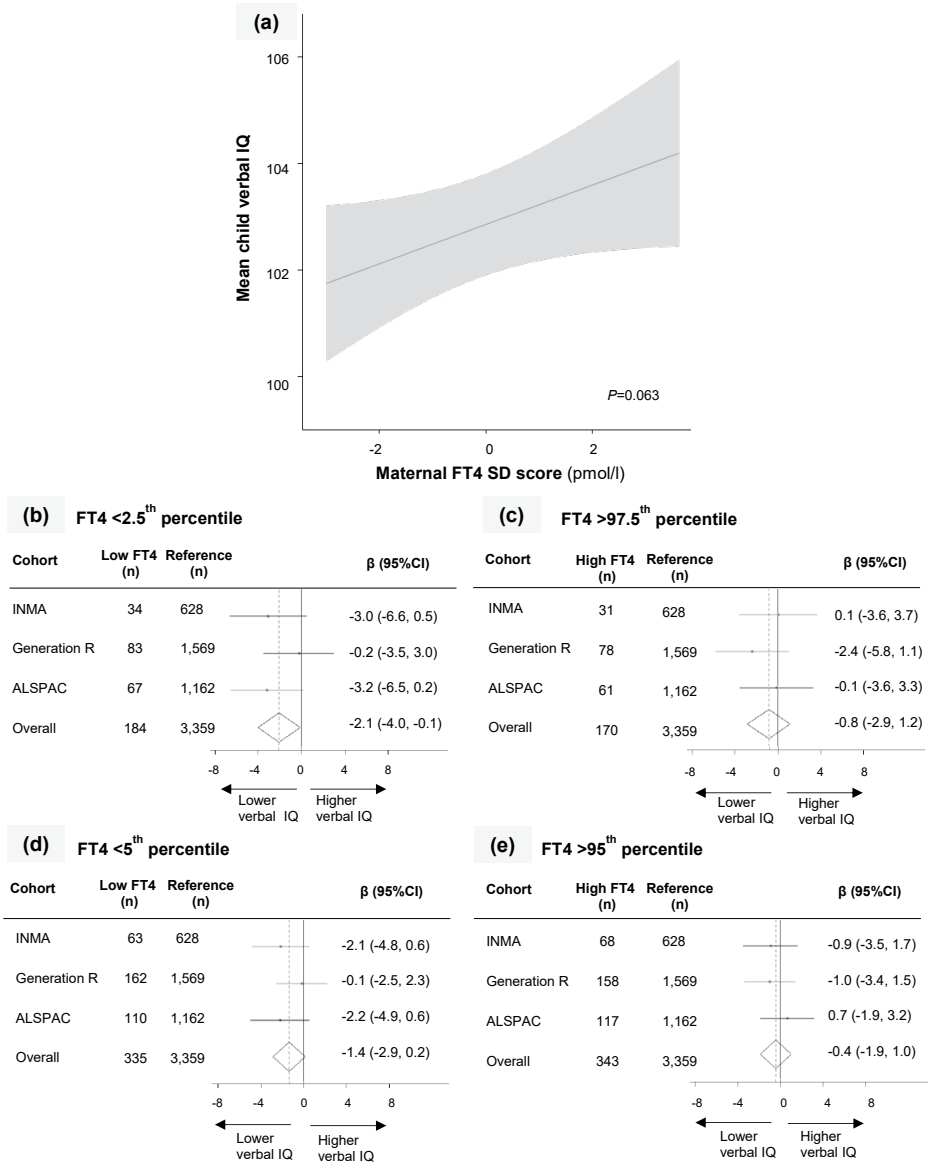



**Figure 3** Association of maternal FT4 during early pregnancy with child verbal IQ.
Association shown as (a) a continuous association depicted as the mean child verbal IQ (black line) with 95% CI (gray area) and by cohort-specific maternal FT4 concentrations in the (b) <2.5th percentile, (c) >97.5th percentile, (d) <5th percentile, and (e) >95th percentile compared with interquartile range (between 25th and 75th percentiles), depicted as effect estimate (dot) with the 95% CI per cohort and overall as estimated by random-effects meta-analysis (diamond). The $I^2$ for each model is as follows: for FT4<2.5$^{th}$ percentile, $I^2$=0.0%; for FT4>97.5$^{th}$ percentile, $I^2$=0.0%; for FT4<5$^{th}$ percentile, $I^2$=0.0%; for FT4>95$^{th}$ percentile, $I^2$=0.0%.

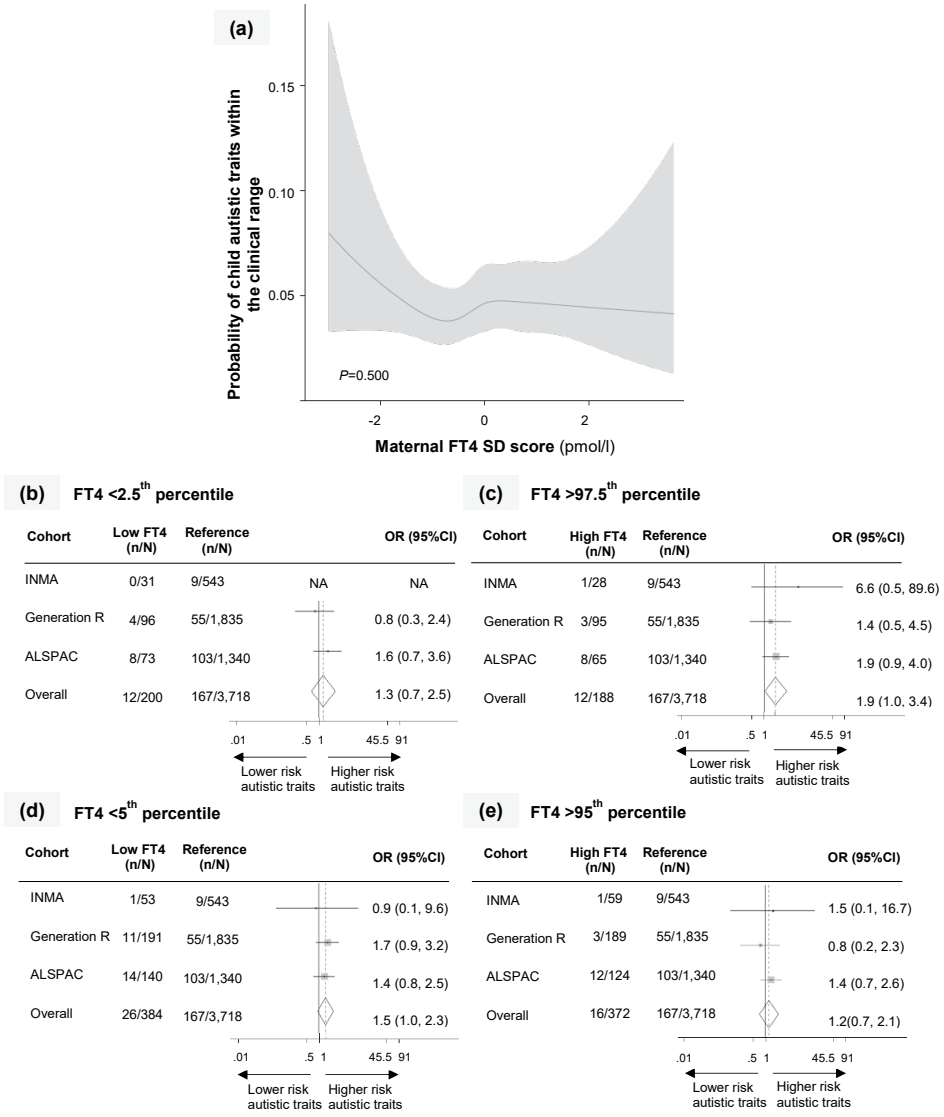



**Figure 4** Association of maternal FT4 during early pregnancy with child autistic traits within the clinical range. Association shown as (a) a continuous association depicted as the mean risk of child autistic traits within the clinical range (black line) with 95% CI (gray area) and by cohort-specific maternal FT4 concentrations in the (b) <2.5th percentile (c) >97.5th percentile, (d) <5th percentile, and (e) >95th percentile compared with interquartile range (between 25th and 75th percentile), depicted as effect estimate (dot) with 95% CI per cohort and overall as estimated by random-effects meta-analysis (diamond). The $I^2$ for each model is as follows: for FT4<2.5$^{th}$ percentile, $I^2$=7.4%; for FT4>97.5$^{th}$ percentile, $I^2$=0.0%; for FT4<5$^{th}$ percentile, $I^2$=0.0%; for FT4>95$^{th}$ percentile, $I^2$=0.0%.

# Discussion

To the best of our knowledge, the present study is the first individual participant data meta-analysis. We have demonstrated that low maternal FT4 in early pregnancy is associated with lower non-verbal and verbal child IQs. We also found a suggestive association between maternal hypothyroxinemia and high FT4 with a greater risk of autistic traits within the clinical range. In contrast to FT4, maternal TSH was not independently associated with non-verbal, verbal IQ, or autistic traits within the clinical range.

The association between low maternal FT4 and child IQ, specifically non-verbal IQ, was highly similar among the three cohorts, convincingly replicating the results of previous observational studies [4–9]. A recent randomized controlled trial studied the effects of levothyroxine treatment for women with subclinical hypothyroidism or hypothyroxinemia on child full IQ [21]. Although levothyroxine treatment of hypothyroxinemia or subclinical hypothyroidism started in mid-pregnancy (week 16 to 18), a statistically non-significant 3 points greater median child IQ was found after levothyroxine treatment compared with placebo. The associations of hypothyroxinemia with a 3.8- and 2.8-point lower non-verbal and verbal IQ, respectively, found in our study compared with euthyroid women might seem small on an individual level. However, on a population level, this might have effects on educational achievements and capita per income, among others [41].

The consistent association of low maternal FT4 with adverse child neurocognitive outcomes, specifically lower non-verbal IQ in three independent cohorts, is particularly relevant given that all three cohorts used a different immunoassay to measure FT4. The value of an FT4 measurement during pregnancy has been under debate, because the absolute values of FT4 might have been under- or overestimated when measured using immunoassays in pregnancy, especially in the third trimester [42–44]. However, these results suggest that FT4 is a reliable clinical marker of the fetal thyroid state in early pregnancy, a period when maternal FT4 is the sole source of thyroid hormones for the fetus and influences the developmental processes, including proliferation, migration, and differentiation of neuronal cells in various parts of the brain [45]. No conclusions about the use of FT4 assays during the later stages of pregnancy, when the fetal thyroid is fully functional, should be drawn from these data.

In our study, the effect estimates for non-verbal IQ were larger than for those for verbal IQ. Non-verbal IQ is a language- and culture-free measure of cognitive ability that is less dependent on the learning stimulus received by the child during the first years of life. Therefore, it might be a better neurodevelopmental outcome for detecting the effects of maternal exposures in early pregnancy, such as thyroid hormone levels.

Our results did not show an association between high maternal FT4 and non-verbal or verbal IQ across the three cohorts, although we confirmed the previously reported association with the Generation R data [4]. The discrepancies in the association of high FT4 and non-verbal IQ among the cohorts might have resulted from population differences such as maternal iodine

status, which differed considerably among the cohorts. Pregnant women in Generation R had an adequate iodine status according to the World Health Organization (median urinary iodine concentration, 229.6 µg/L [19]). In contrast, mild to moderate iodine insufficiency was present in the INMA and ALSPAC cohorts (median, 94 to 168 µg/L depending on the region and 91.1 µg/L, respectively [46,47]). Although mild-to-moderate iodine deficiency has been associated with adverse neurodevelopmental outcomes, such as lower verbal IQ, worse language skills, reduced educational outcomes, impaired executive function, more behavior problems, and worse fine motor skills, this was not found in iodine-deficient women in a iodine sufficient population [19,48–50]. It is unclear how much of the association of iodine deficiency with child neurocognitive outcomes can be attributed to impaired thyroid function in the mother or to impaired thyroid function in the fetus. Further studies should elucidate the mediating role of maternal and fetal thyroid function in the association between maternal iodine status and child neurodevelopment.

To date, only two studies have explored the association between maternal thyroid function and ASD diagnosis or autistic traits. The Danish study was based on registry linkage information and showed that maternal diagnosed or treated hypothyroidism was associated with a greater risk of a diagnosed ASD (hazard ratio, 1.30, 95% CI, 1.11 to 1.53) [51]. The Dutch study from the Generation R cohort found that severe hypothyroxinemia, defined as maternal FT4 $\leq 5^{th}$ percentile with normal TSH, was associated with a greater risk of autistic traits [9]. In the present meta-analysis, including data from Generation R, we also found an association between hypothyroxinemia using the FT4 fifth percentile or less cutoff and a greater risk of autistic traits. However, when using the FT4 $\leq 2.5^{th}$ percentile cutoff, no greater risk of autistic was found, suggesting the possibility of a chance finding. Likewise, high FT4 was associated with a greater risk of autistic traits, although only when the more stringent cutoff was used (*i.e.*, FT4 $\geq 97.5^{th}$ percentile). Considering the crucial role of thyroid hormones in key processes in the pathophysiology of ASD, including neuronal cell migration, synaptogenesis, synapse maintenance, neuronal activity, and fetal growth [52,53], it is biologically plausible that nonoptimal levels of maternal FT4 during early pregnancy are related to a greater risk of ASD. However, the inconsistent results across cohorts or cutoffs limited us from drawing firm conclusions regarding this potential association. Further studies focusing on autistic traits or ASD diagnosis are therefore needed to replicate and better understand the full extent of these results.

TSH is frequently used as a marker of thyroid status during pregnancy. Subclinical hypothyroidism has been associated with a greater risk of miscarriage and preterm delivery, and the beneficial effects of levothyroxine treatment for hypothyroid women have been shown in some trials, especially in TPOAb-positive women [54–57]. Therefore, the current international guidelines recommend screening for TSH first, either directly in combination with TPOAb status [27] or determining TPOAb status and FT4 only when TSH is elevated [58]. The results from the present study call into question the use of TSH as the only first-line parameter to screen

maternal thyroid status in early pregnancy. First, elevated human chorionic gonadotropin concentrations stimulate the thyroid directly to produce thyroid hormone, which induces a decrease in TSH in early pregnancy [59]. Therefore, TSH might not be the best marker for maternal thyroid status in this time window. Second, in our study, maternal TSH was not independently associated with non-verbal IQ, verbal IQ, or autistic traits, in contrast to FT4. However, owing to the absence of available randomized trials demonstrating the benefit of levothyroxine treatment for maternal hypothyroxinemia, screening for FT4 cannot be advocated.

One strength of the present study was that we investigated the association of maternal thyroid function with child neurodevelopmental outcomes in a prospective manner using a large dataset with detailed data on non-verbal IQ, verbal IQ, and autistic traits, assessed using validated tools. Furthermore, by combining data from three different countries, we were able to perform an external replication of previous studies and assess potential differences related to iodine status, after adjusting for many potential confounding variables. We also used advanced statistical methods, including multiple imputation combined with inverse probability weighting, to reduce possible selection bias.

One limitation of the present study was that the child neurodevelopmental outcomes were assessed with different tools at different ages. This might be, for example, reflected in the different prevalence of children with autistic traits within the clinical range across cohorts. The varying occurrence might have resulted from the different ages at the assessment and/or the different types of evaluator but most likely resulted from the different set of questions for assessing autistic traits. For instance, the CAST [33] contains 31 items and is therefore a more extensive questionnaire compared with the CBCL1½–5, with 13 items [34], and the SCDC, with 12 items [35]. The CAST and CBCL1½–5 cover questions on all three domains of ASD. In contrast to the CAST and CBCL1½–5, the SCDC was designed to assess deficits in social and communications skills but does not assess the ASD domain of restricted and repetitive behaviors and interests. To account for the differences as best as possible, we standardized all outcome scores and adjusted all analyses for child age at the IQ or autistic traits ascertainment. We observed little heterogeneity among the cohorts. Another limitation was the low prevalence children with autistic traits within the clinical range, which caused, especially in INMA, issues with statistical power. Furthermore, we only had a single thyroid function measurement available from early pregnancy. Hence, the results should not be generalized to thyroid function in late pregnancy, and the potential effects of individual variations in maternal thyroid hormone availability could not be studied.

In conclusion, the results from the present study have confirmed that a low FT4 is consistently associated with a lower child IQ. We also found a suggestive association of maternal hypothyroxinemia and high FT4 with a greater risk of autistic traits within the clinical range. FT4 seemed a reliably marker of the fetal thyroid state in early pregnancy, regardless of

the type of immunoassay. Further studies should replicate the findings of autistic traits and investigate the potential modifying role of maternal iodine status.

# Acknowledgements

EUthyroid Project: European Union´s Horizon 2020 research and innovation programme under grant agreement No 634453.

INMA, Spain: This study was funded by grants from UE (FP7-ENV-2011 cod 282957 and HEALTH.2010.2.4.5-1) and Spain: Instituto de Salud Carlos III (Red INMA G03/176; CB06/02/0041; FIS-FEDER: PI041436, PI05/1079, PI06/0867, PI081151 , FIS-and PS09/00090, PI11/01007, PI11/02591, PI11/02038, PI13/1944, PI13/2032, PI14/00891, PI14/01687, and PI16/1288; Miguel Servet-FEDER CP11/00178, CP15/00025, and CPII16/00051, MS13/00054), Generalitat Valenciana: FISABIO (UGP 15-230, UGP-15-244, and UGP-15-249), Generalitat de Catalunya-CIRIT 1999SGR 00241, Fundació La marató de TV3 (090430), Department of Health of the Basque Government (2005111093 and 2009111069), and the Provincial Government of Gipuzkoa (DFG06/004 and DFG08/001).

Generation R, the Netherlands: The Generation R Study is conducted by the Erasmus Medical Center in close collaboration with the Faculty of Social Sciences of the Erasmus University Rotterdam, the Municipal Health Service Rotterdam area, Rotterdam, and the Stichting Trombosedienst & Artsenlaboratorium Rijnmond (STAR-MDC), Rotterdam. The Generation R Study is supported by the Erasmus Medical Center, Rotterdam, the Erasmus University Rotterdam, the Netherlands Organization for Health Research and Development (ZonMw), the Netherlands Organization for Scientific Research (NWO), the Ministry of Health, Welfare and Sport. A grant from the Sophia Children's Hospital Research Funds supports the neurodevelopmental work on thyroid; Robin P. Peeters is supported by a clinical fellowship from ZonMw, project number 90700412.

ALSPAC, United Kingdom: We are extremely grateful to all the families who took part in this study, the midwives for their help in recruiting them, and the whole ALSPAC team, which includes interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists, and nurses. The UK Medical Research Council and Wellcome (Grant ref: 102215/2/13/2) and the University of Bristol currently provide core support for ALSPAC. The publication is the work of the authors and Mònica Guxens will serve as guarantor for the contents of this paper. Data collection is funded from a wide range of sources, which are detailed on the ALSPAC website: http://www.bristol.ac.uk/alspac/about/.

We would like to thank the participants of all cohorts for their generous collaboration; Professor Scott Nelson, University of Glasgow, UK, for the ALSPAC thyroid measurements which were funded by the Chief Scientist Office, Scotland (ETM 97/0357/130024782); Ane M. Castilla of the Public Health Laboratory of Bilbao, Spain, for the thyroid measurements in

INMA, and Professor dr. Yolanda de Rijke, of the department of Clinical Chemistry, Erasmus MC, University Medical Center, the Netherlands for design and measurements of the thyroid measurements in Generation R.

# Supplemental material

**Supplemental Table 1** Type of immunoassay used by the cohorts.

| Cohort | Assay | TPOAb positivity[a] |
|---|---|---|
| INMA | AutoDEL-FIA (PerkinElmer Life and Analytical Sciences, Wallac Oy, Turku, Finland) and a lanthanide metal europium (Eu) label. | NA |
| Generation R | Vitros ECI immunodiagnostic (Ortho Clinical Diagnostics, Rochester, NY, USA) | ≥60 IU/ml |
| ALSPAC | Abbott Architect i2000 | ≥6 IU/ml |

[a] Assay cut-off. NA: not available

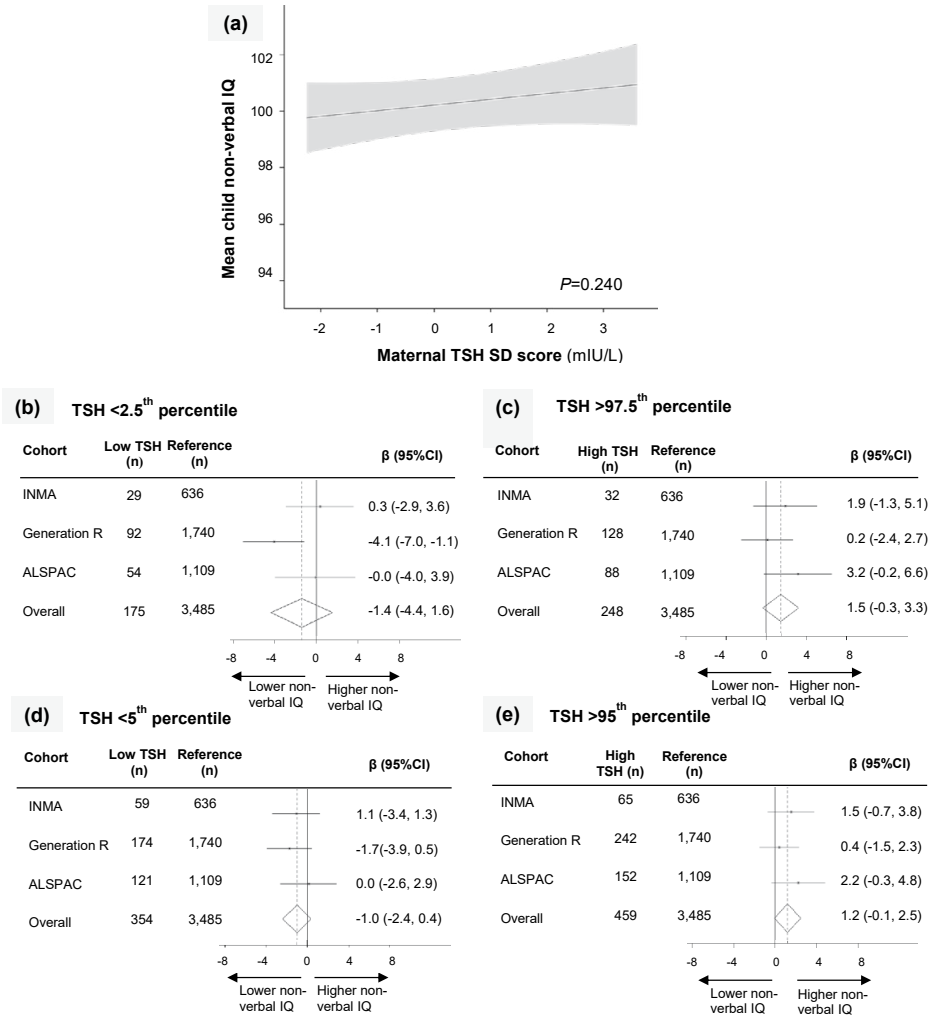
**Supplemental Table 2** Distribution and comparison of maternal and child characteristics in the included and excluded population.

| | INMA | | | Generation R | | | ALSPAC | | |
|---|---|---|---|---|---|---|---|---|---|
| | **Included (n=1,289)** | **Excluded (n=679)** | ***P*-value** | **Included (n=4,660)** | **Excluded (n=2,358)** | ***P*-value** | **Included (n=3,087)** | **Excluded (n=1,790)** | ***P*-value** |
| Maternal TSH, mIU/L | 1.24 (0.84-1.81) | 1.25 (0.77-1.91) | .660 | 1.36 (0.85-2.03) | 1.28 (0.78-2.03) | .066 | 1.00 (0.64-1.46) | 0.96 (0.62-1.43) | .088 |
| Maternal FT4, pmol/L | 10.6 (9.7-11.6) | 10.4 (9.5-11.5) | .018 | 14.8 (13.2-16.7) | 14.6 (12.9-16.6) | .071 | 16.2 (14.8-17.7) | 16.3 (14.9-17.9) | .034 |
| TPOAb positivity, n (%) | NA | NA | | 254 (5.8) | 96 (6.7) | .253 | 392 (12.8) | 212 (11.9) | .377 |
| Gestational age at blood sampling, weeks | 13.1 (1.3) | 13.2 (1.3) | .074 | 13.4 (2.0) | 13.8 (2.2) | <.001 | 11.0 (3.2) | 11.0 (3.2) | .942 |
| Maternal educational level, n (%) | | | | | | | | | |
| Low | 281 (21.9) | 216 (31.1) | | 353 (8.0) | 295 (14.8) | | 736 (24.7) | 562 (39.8) | |
| Medium | 537 (41.8) | 279 (40.1) | <.001 | 1,904 (42.9) | 1,022 (51.3) | <.001 | 1,828 (61.3) | 735 (52.0) | <.001 |
| High | 468 (36.4) | 200 (28.8) | | 2,179 (49.1) | 676 (33.9) | | 416 (14.0) | 116 (8.2) | |
| Maternal ethnicity, n (%) | | | | | | | | | |
| Spanish | 1,202 (93.4) | 592 (85.2) | | NA | NA | | NA | NA | |
| Latin-american | 60 (4.7) | 68 (9.8) | | NA | NA | | NA | NA | |
| European/other | 25 (1.9) | 35 (5.0) | | NA | NA | | NA | NA | |
| Dutch | NA | NA | | 2,606 (56.7) | 844 (41.0) | | NA | NA | |
| Indonesian | NA | NA | | 150 (3.3) | 50 (2.4) | | NA | NA | |
| Cape verdian | NA | NA | | 170 (3.7) | 108 (5.3) | | NA | NA | |
| Moroccan | NA | NA | | 225 (4.9) | 179 (8.7) | | NA | NA | |
| Dutch Antilles | NA | NA | <.001 | 104 (2.3) | 89 (4.3) | <.001 | NA | NA | <.001 |
| Surinamese | NA | NA | | 351 (7.6) | 229 (11.1) | | NA | NA | |
| Turkish | NA | NA | | 356 (7.8) | 205 (10.0) | | NA | NA | |
| Asian | NA | NA | | 51 (1.1) | 40 (1.9) | | NA | NA | |
| Other, non-western | NA | NA | | 162 (3.5) | 135 (6.6) | | NA | NA | |
| Other, western | NA | NA | | 418 (9.1) | 178 (8.7) | | NA | NA | |
| White | NA | NA | | NA | NA | | 2,924 (98.6) | 1,349 (96.5) | |
| Non-white | NA | NA | | NA | NA | | 42 (1.4) | 49 (3.5) | |
| Maternal age, years | 31.5 (4.0) | 30.8 (4.8) | <.001 | 30.3 (4.8) | 28.6 (5.5) | <.001 | 28.0 (4.6) | 26.3 (5.0) | <.001 |

**Supplemental Table 2** Distribution and comparison of maternal and child characteristics in the included and excluded population (continued).

| | INMA | | | Generation R | | | ALSPAC | | |
|---|---|---|---|---|---|---|---|---|---|
| | **Included (n=1,289)** | **Excluded (n=679)** | ***P*-value** | **Included (n=4,660)** | **Excluded (n=2,358)** | ***P*-value** | **Included (n=3,087)** | **Excluded (n=1,790)** | ***P*-value** |
| Parity, n (%) | | | | | | | | | |
| 0 | 731 (56.8) | 361 (51.9) | | 2,721 (58.4) | 1,203 (52.7) | | 1,410 (47.2) | 692 (43.9) | |
| 1 | 472 (36.7) | 275 (39.6) | .069 | 1,386 (29.7) | 679 (29.7) | <.001 | 1,033 (34.6) | 530 (33.6) | .002 |
| ≥2 | 84 (6.5) | 59 (8.5) | | 553 (11.9) | 402 (17.6) | | 543 (18.2) | 354 (22.5) | |
| Maternal smoking, n (%) | | | | | | | | | |
| Never smoked | 883 (69.4) | 419 (66.4) | | 3,085 (73.5) | 1,339 (68.6) | | 2,391 (79.2) | 1,077 (67.0) | |
| Smoked at the beginning of pregnancy | 174 (13.7) | 90 (14.3) | .372 | 396 (9.4) | 176 (9.0) | <.001 | 142 (4.7) | 107 (6.7) | <.001 |
| Continued smoking | 216 (17.0) | 122 (19.3) | | 719 (17.1) | 437 (22.4) | | 486 (16.1) | 424 (26.4) | |
| Pre-pregnancy BMI, kg/m$^2$ | 22.5 (20.8-25.1) | 22.5 (20.7-25.2) | .845 | 22.6 (20.7-25.2) | 22.7 (20.7-25.6) | .215 | 22.1 (20.5-24.2) | 22.2 (20.5-24.5) | .380 |
| Child female sex, n (%) | 635 (49.3) | 301 (47.6) | .490 | 2,313 (49.6) | 1,155 (49.0) | .629 | 1,500 (48.6) | 845 (47.2) | .360 |

Values represent median (interquartile range) or mean (SD), unless stated otherwise. *P*-value for differences calculated using Chi-square test for categorical variables, Student´s t-test for continuous normal-distributed variables, and Kruskal-Wallis test for continuous non-normal distributed variables. NA: not available. Numbers may not add up due to rounding.



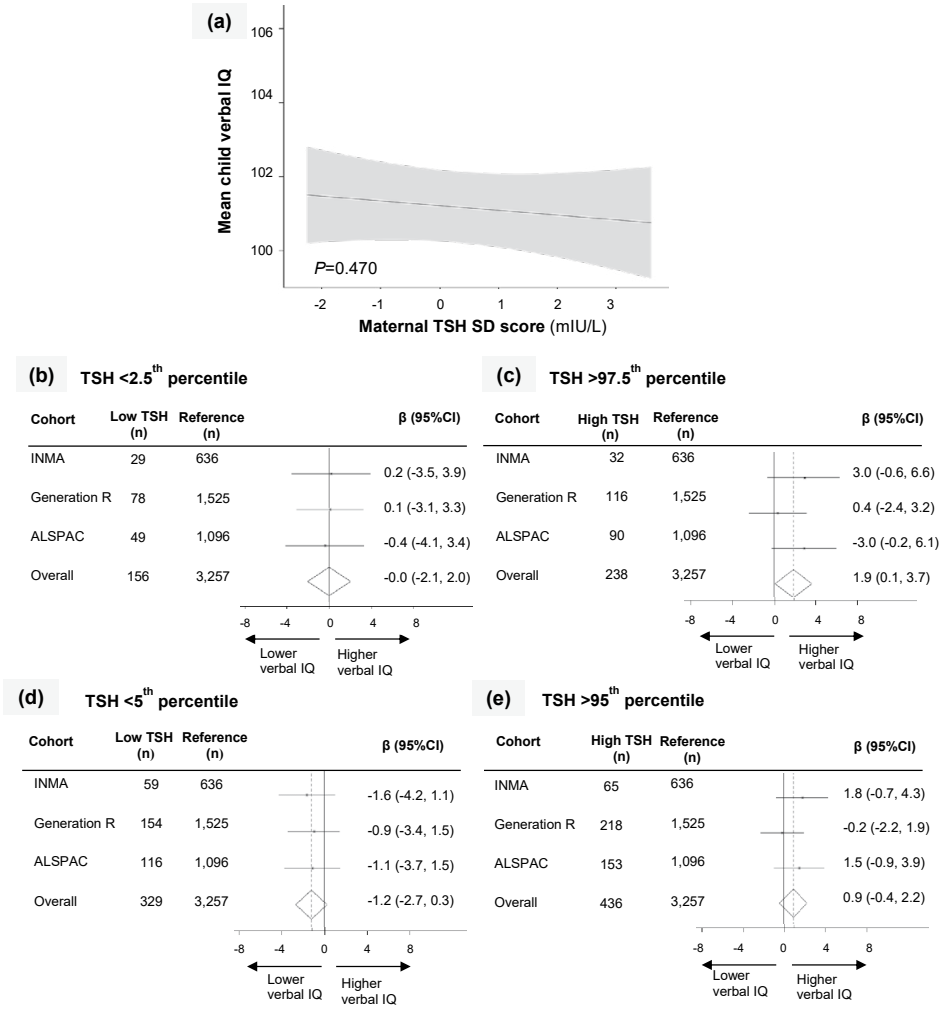

**Supplemental Fig. 1** Association of maternal TSH during early pregnancy with child non-verbal IQ**.** Association shows as (a) a continuous association depicted as the child non-verbal IQ (black line) with 95% CI (gray area) and by cohort-specific maternal TSH concentrations in the (b) <2.5th percentile, (c) >97.5th percentile, (d) <5th percentile, and (e) >95th percentile compared with interquartile range (between 25th and 75th percentiles), depicted as effect estimate (dot) with the 95% CI per cohort and overall as estimated by random-effects meta-analysis (diamond). The $I^2$ for each model is as follows: for TSH<2.5th percentile, $I^2$=57.5%; for TSH>97.5th percentile, $I^2$=4.8%; for TSH<5th percentile, $I^2$=0.0%; for TSH>95th percentile, $I^2$=0.0%.

**Supplemental Table 3** Association of maternal TSH with child non-verbal IQ stratified by FT4.

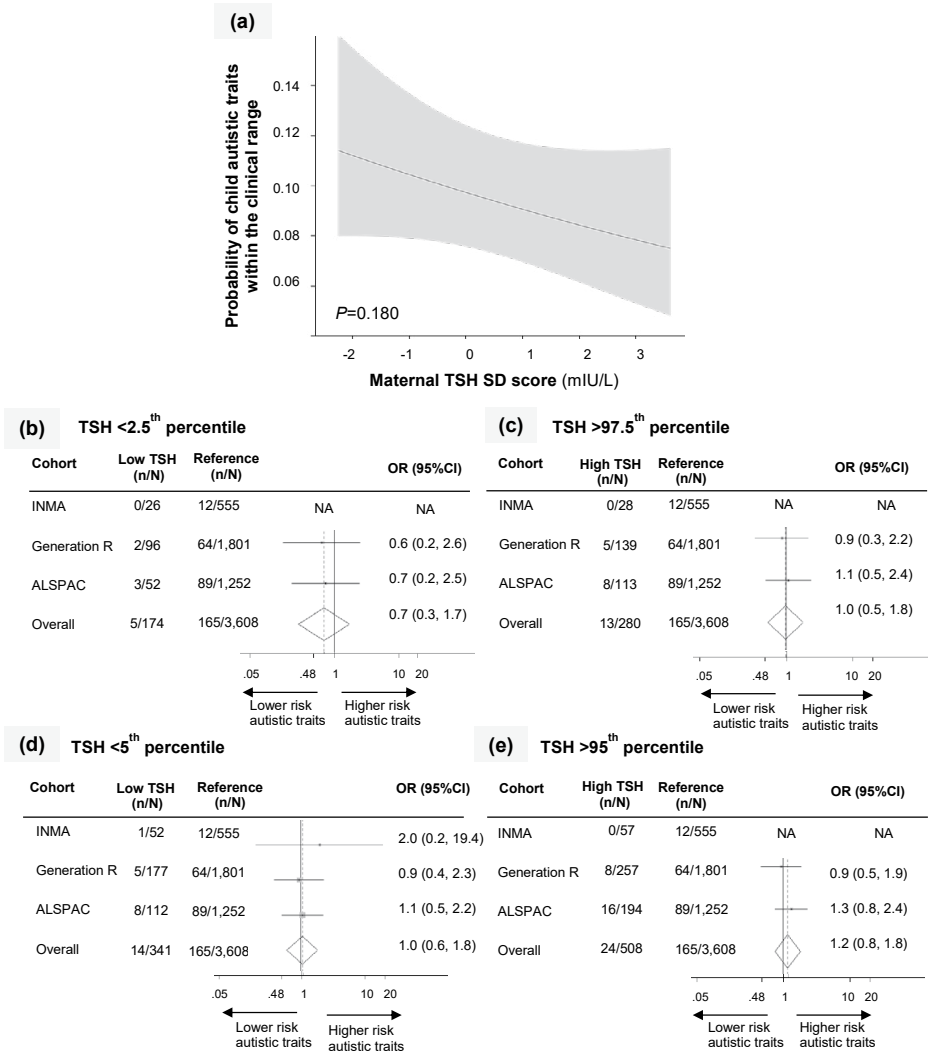
| | β (95% CI) | *P*-value | *P* heterogeneity | $I^2$ |
|---|---|---|---|---|
| **High TSH (>P95) vs. normal TSH (P25-P75)** | 1.2 (-0.1 to 2.5) | 0.063 | 0.503 | 0.0 |
| in low FT4 group (<P25) | 0.9 (-1.3 to 3.0) | 0.421 | 0.522 | 0.0 |
| in medium FT4 group (P25-P75) | 2.1 (0.1 to 4.0) | 0.038 | 0.343 | 6.5 |
| in high FT4 group (>P75) | 1.0 (-2.6 to 4.6) | 0.578 | 0.926 | 0.0 |

**Supplemental Table 4** Association of maternal TSH with child verbal IQ stratified by FT4.

| | β (95% CI) | *P*-value | *P* heterogeneity | $I^2$ |
|---|---|---|---|---|
| **High TSH (>P97.5) vs. normal TSH (P25-P75)** | 1.9 (0.1 to 3.7) | 0.039 | 0.388 | 0.0 |
| in low FT4 group (<P25) | 0.8 (-2.0 to 3.6) | 0.565 | 0.780 | 0.0 |
| in medium FT4 group (P25-P75) | 1.7 (-1.0 to 4.5) | 0.219 | 0.432 | 0.0 |
| in high FT4 group (>P75) | 3.3 (-5.0 to 11.6) | 0.438 | 0.103 | 55.9 |

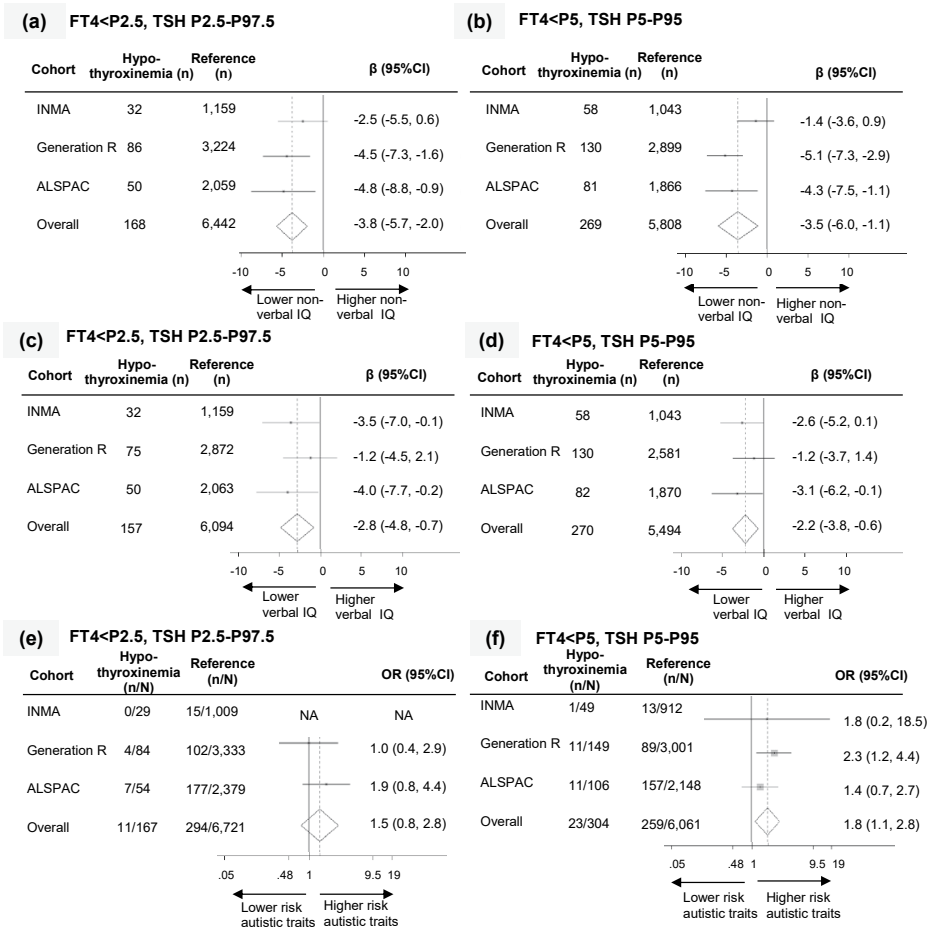



**Supplemental Fig. 2** Association of maternal TSH during early pregnancy with child verbal IQ.
Association shown as (a) a continuous association depicted as the child verbal IQ (black line) with 95% CI (gray area); and by cohort-specific maternal TSH concentrations in the (b) <2.5th percentile, (c) >97.5th percentile, (d) <5th percentile, and (e) >95th percentile compared with interquartile range (between 25th and 75th percentiles), depicted as effect estimate (dot) with the 95% CI per cohort and overall as estimated by random-effects meta-analysis (diamond). The $I^2$ for each model is as follows: for TSH<2.5th percentile, $I^2$=0.0%; for TSH>97.5th percentile, $I^2$=0.0%; for TSH<5th percentile, $I^2$=0.0%; for TSH>95th percentile, $I^2$=0.0%.



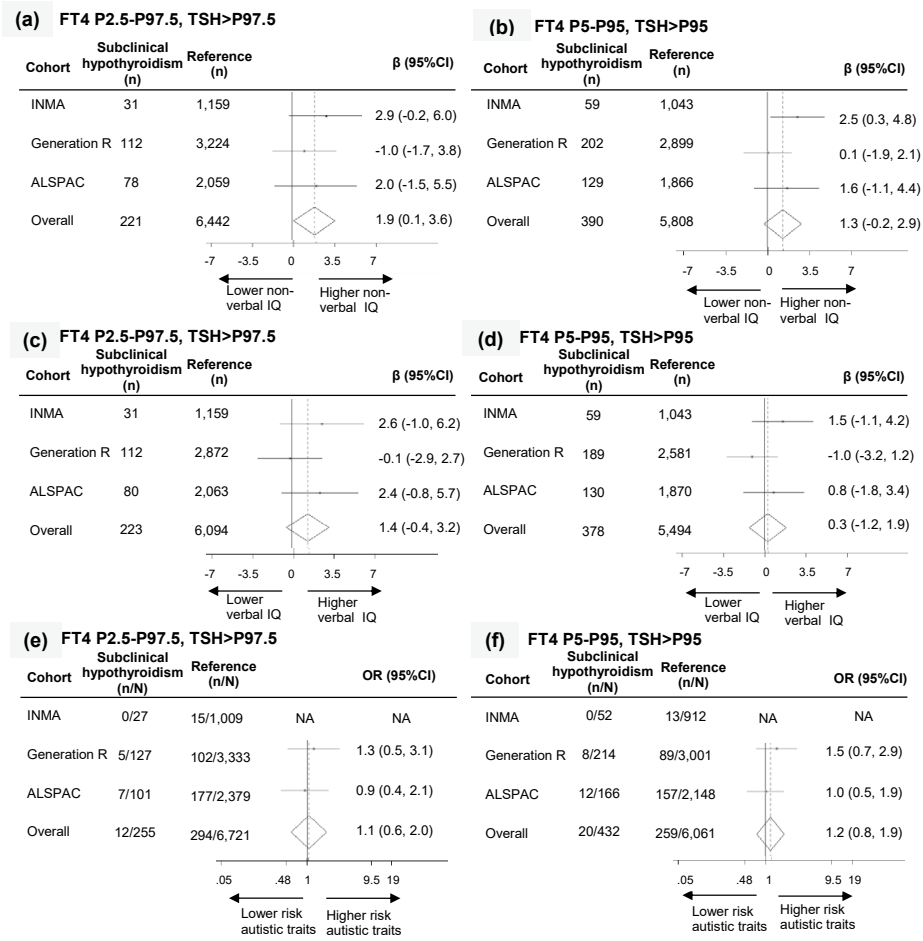

**Supplemental Fig. 3** Association of maternal TSH during early pregnancy with child autistic traits within the clinical range.

Association shown a (a) a continuous association depicted as the child autistic traits within the clinical range (black line) with 95% CI (gray area) and by cohort-specific maternal TSH concentrations in the (b) <2.5th percentile, (c) >97.5th percentile, (d) <5th percentile, and (e) >95th percentile compared with interquartile range (between 25th and 75th percentiles), depicted as effect estimate (dot) with the 95% CI per cohort and overall as estimated by random-effects meta-analysis (diamond). The $I^2$ for each model is as follows: for TSH<2.5th percentile, $I^2$=0.0%; for TSH>97.5th percentile, $I^2$=0.0%; for TSH<5th percentile, $I^2$=0.0%; for TSH>95th percentile, $I^2$=0.0%.



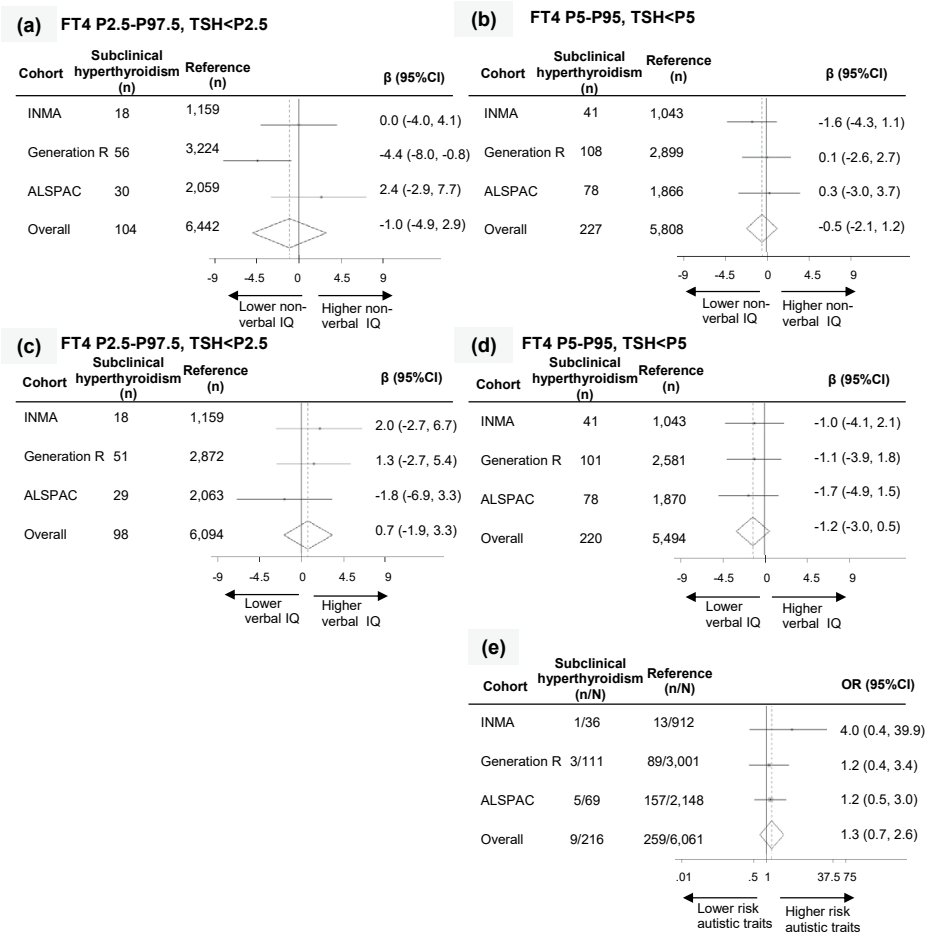

**Supplemental Fig. 4** Association of maternal hypothyroxinemia during early pregnancy with child non-verbal IQ, verbal IQ, and autistic traits within the clinical range.

Figure shows the association of maternal hypothyroxinemia during early pregnancy with child non-verbal IQ (a,b), verbal IQ (c,d), and autistic traits within the clinical range (e,f), defined using the cohort-specific 2.5th and 97.5th population-based percentiles (a,c,e) or the 5th and 95th population-based percentiles (b,d,f), depicted as effect estimate (dot) with the 95% CI per cohort and overall as estimated by random-effects meta-analysis (diamond). The reference group consisted of euthyroid women. The $I^2$ for each model is as follows: for a, $I^2$=0.0%; for b, $I^2$=64.6%; for c, $I^2$=0.0%; for d, $I^2$=0.0%; for e, $I^2$=0.0%; for f, $I^2$=0.0%.




**Supplemental Fig. 5** Association of maternal subclinical hypothyroidism during early pregnancy with child non-verbal IQ, verbal IQ, and autistic traits within the clinical range.

Figure shows the association of maternal subclinical hypothyroidism during early pregnancy with child non-verbal IQ (a,b), verbal IQ (c,d), and autistic traits within the clinical range (e,f), defined using the cohort-specific 2.5th and 97.5th population-based percentiles (a,c,e) or the 5th and 95th population-based percentiles (b,d,f), depicted as effect estimate (dot) with the 95% CI per cohort and overall as estimated by random-effects meta-analysis (diamond). The reference group consisted of euthyroid women. The $I^2$ for each model is as follows: for a, $I^2$=0.0%; for b, $I^2$=26.8%; for c, $I^2$=0.0%; for d, $I^2$=17.1%; for e, $I^2$=0.0%; for f, $I^2$=0.0%.




**Supplemental Fig. 6** Association of maternal subclinical hyperthyroidism during early pregnancy with child non-verbal IQ, verbal IQ, and autistic traits within the clinical range.

Figure shows the association of maternal subclinical hyperthyroidism during early pregnancy with child non-verbal IQ (a,b), verbal IQ (c,d), and autistic traits within the clinical range (e), defined using the cohort-specific 2.5th and 97.5th population-based percentiles (a,c) or the 5th and 95th population-based percentiles (b,d,e), depicted as effect estimate (dot) with the 95% CI per cohort and overall as estimated by random-effects meta-analysis (diamond). The reference group consisted of euthyroid women. The $I^2$ for each model is as follows: for a $I^2$=60.5%; for b, $I^2$=0.0%; for c, $I^2$=0.0%; for d, $I^2$=0.0%; for e, $I^2$=0.0%.

# References


1. Lavado-Autric R, Ausó E, García-Velasco JV, et al. Early maternal hypothyroxinemia alters histogenesis and cerebral cortex cytoarchitecture of the progeny. *J Clin Invest*. 2003;111(7):1073-1082.
2. Bernal J. Thyroid Hormones and Brain Development. *Vitam Horm*. 2005;71:95-122.
3. Thorpe-Beeston JG, Nicolaides KH, Felton CV, Butler J, McGregor AM. Maturation of the secretion of thyroid hormone and thyroid-stimulating hormone in the fetus. *N Engl J Med*. 1991;324(8):532-536.
4. Korevaar TIM, Muetzel R, Medici M, et al. Association of maternal thyroid function during early pregnancy with offspring IQ and brain morphology in childhood: a population-based prospective cohort study. *Lancet Diabetes Endocrinol*. 2016;4(1):35-43.
5. Haddow JE, Palomaki GE, Allan WC, et al. Maternal thyroid deficiency during pregnancy and subsequent neuropsychological development of the child. *N Engl J Med*. 1999;341(8):549–555.
6. Julvez J, Alvarez-Pedrerol M, Rebagliato M, et al. Thyroxine Levels During Pregnancy in Healthy Women and Early Child Neurodevelopment: *Epidemiology*. 2013;24(1):150-157.
7. Henrichs J, Bongers-Schokking JJ, Schenk JJ, et al. Maternal Thyroid Function during Early Pregnancy and Cognitive Functioning in Early Childhood: The Generation R Study. *J Clin Endocrinol Metab*. 2010;95(9):4227-4234.
8. 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. *Clin Endocrinol (Oxf)*. 2003;59(3):282-288.
9. Román GC, Ghassabian A, Bongers-Schokking JJ, et al. Association of gestational maternal hypothyroxinemia and increased autism risk. *Ann Neurol*. 2013;74(5):733-742.
10. Pop VJ, Kuijpens JL, van Baar AL, et al. Low maternal free thyroxine concentrations during early pregnancy are associated with impaired psychomotor development in infancy. *Clin Endocrinol (Oxf)*. 1999;50(2):149-155.
11. Gyllenberg D, Sourander A, Surcel H-M, Hinkka-Yli-Salomäki S, McKeague IW, Brown AS. Hypothyroxinemia During Gestation and Offspring Schizophrenia in a National Birth Cohort. *Biol Psychiatry*. 2016;79(12):962-970.
12. Marta CB, Adamo AM, Soto EF, Pasquini JM. Sustained neonatal hyperthyroidism in the rat affects myelination in the central nervous system. *J Neurosci Res*. 1998;53(2):251-259.
13. Pasquini JM, Adamo AM. Thyroid hormones and the central nervous system. *Dev Neurosci*. 1994;16(1-2):1-8.
14. Nicholson JL, Altman J. Synaptogenesis in the rat cerebellum: effects of early hypo- and hyperthyroidism. *Science*. 1972;176(4034):530-532.
15. Nicholson JL, Altman J. The effects of early hypo- and hyperthyroidism on the development of rat cerebellar cortex. I. Cell proliferation and differentiation. *Brain Res*. 1972;44(1):13-23.
16. Nicholson JL, Altman J. The effects of early hypo- and hyperthyroidism on the development of the rat cerebellar cortex. II. Synaptogenesis in the molecular layer. *Brain Res*. 1972;44(1):25-36.
17. Lauder JM. The effects of early hypo- and hyperthyroidism on the development of rat cerebellar cortex. III. Kinetics of cell proliferation in the external granular layer. *Brain Res*. 1977;126(1):31-51.
18. Lauder JM, Altman J, Krebs H. Some mechanisms of cerebellar foliation: effects of early hypo- and hyperthyroidism. *Brain Res*. 1974;76(1):33-40.
19. Ghassabian A, Steenweg-de Graaff J, Peeters RP, et al. Maternal urinary iodine concentration in pregnancy and children's cognition: results from a population-based birth cohort in an iodine-sufficient area. *BMJ Open*. 2014;4:e005520.

20. Lazarus JH, Bestwick JP, Channon S, et al. Antenatal Thyroid Screening and Childhood Cognitive Function. *N Engl J Med*. 2012;366(6):493-501.
21. Casey BM, Thom EA, Peaceman AM, et al. Treatment of Subclinical Hypothyroidism or Hypothyroxinemia in Pregnancy. *N Engl J Med*. 2017;376(9):815-825.
22. Korevaar TIM, Chaker L, Peeters RP. Improving the clinical impact of randomised trials in thyroidology. *Lancet Diabetes Endocrinol*. 2018;6(7):523-525.
23. Guxens M, Ballester F, Espada M, et al. Cohort Profile: the INMA--INfancia y Medio Ambiente--(Environment and Childhood) Project. *Int J Epidemiol*. 2012;41(4):930-940.
24. Kooijman MN, Kruithof CJ, van Duijn CM, et al. The Generation R Study: design and cohort update 2017. *Eur J Epidemiol*. 2016;31(12):1243-1264.
25. Boyd A, Golding J, Macleod J, et al. Cohort Profile: The 'Children of the 90s'—the index offspring of the Avon Longitudinal Study of Parents and Children. *Int J Epidemiol*. 2013;42(1):111-127.
26. ALSPAC Executive. Access data and samples. http://www.bris.ac.uk/alspac/researchers/data-access/data-dictionary/. Accessed March 12, 2018.
27. Alexander EK, Pearce EN, Brent GA, et al. 2017 Guidelines of the American Thyroid Association for the Diagnosis and Management of Thyroid Disease during Pregnancy and the Postpartum. *Thyroid Off J Am Thyroid Assoc*. 2017;27(3):315-389.
28. McCarthy D. *McCarthy Scales of Children's Abilities*. Psychological Corporation; 1972.
29. Tellegen PJ, Winkel M, Wijnberg-Williams BJ, Laros JA. *Snijders-Oomen Nonverbal Intelligence Test. SON-R 21/2-7 Manual and Research Report*. Lisse: Swets & Zeitlinger B.V.; 1998.
30. Fenson L, Pethick S, Renda C, Cox JL, Dale PS, Reznick JS. Short-form versions of the MacArthur Communicative Development Inventories. *Appl Psycholinguist*. 2000;21(1):95-116.
31. Wechsler D, Golombok S, Rust J. *WISC-III UK Wechsler Intelligence Scale for Children.* Sidcup, UK: Psychological Corporation; 1992.
32. Constantino JN, Todd RD. Autistic traits in the general population: a twin study. *Arch Gen Psychiatry*. 2003;60(5):524-530.
33. Williams J, Scott F, Stott C, et al. The CAST (Childhood Asperger Syndrome Test): Test accuracy. *Autism*. 2005;9(1):45-68.
34. Achenbach TM, Rescorla LA. *ASEBA Preschool Forms & Profiles*. Burlington, VT: University of Vermont, Research Center for Children, Youth and Families; 2000.
35. Skuse DH, Mandy WPL, Scourfield J. Measuring autistic traits: heritability, reliability and validity of the Social and Communication Disorders Checklist. *Br J Psychiatry J Ment Sci*. 2005;187:568-572.
36. Shrier I, Platt RW. Reducing bias through directed acyclic graphs. *BMC Med Res Methodol*. 2008;8:70.
37. Bravata DM, Olkin I. Simple pooling versus combining in meta-analysis. *Eval Health Prof*. 2001;24(2):218–230.
38. Higgins, JP, Thompson, SG. Quantifying heterogeneity in a meta-analysis. *Stat Med*. 2002;21(11):1539–1558.
39. Weisskopf MG, Sparrow D, Hu H, Power MC. Biased Exposure-Health Effect Estimates from Selection in Cohort Studies: Are Environmental Studies at Particular Risk? *Environ Health Perspect*. 2015;123(11):1113-1122.
40. Sterne JAC, White IR, Carlin JB, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *BMJ*. 2009;338:b2393.
41. Lynn R, Vanhanen T. National IQs: A review of their educational, cognitive, economic, political, demographic, sociological, epidemiological, geographic and climatic correlates. *Intelligence*. 2012;40(2):226-234.

42. Jonklaas J, Kahric-Janicic N, Soldin OP, Soldin SJ. Correlations of free thyroid hormones measured by tandem mass spectrometry and immunoassay with thyroid-stimulating hormone across 4 patient populations. *Clin Chem*. 2009;55(7):1380-1388.
43. Anckaert E, Poppe K, Van Uytfanghe K, Schiettecatte J, Foulon W, Thienpont LM. FT4 immunoassays may display a pattern during pregnancy similar to the equilibrium dialysis ID-LC/tandem MS candidate reference measurement procedure in spite of susceptibility towards binding protein alterations. *Clin Chim Acta Int J Clin Chem*. 2010;411(17-18):1348-1353.
44. Lee RH, Spencer CA, Mestman JH, et al. Free T4 immunoassays are flawed during pregnancy. *Am J Obstet Gynecol*. 2009;200(3):260.e1-6.
45. Howdeshell KL. A model of the development of the brain as a construct of the thyroid system. *Environ Health Perspect*. 2002;110(Suppl 3):337-348.
46. Murcia M, Rebagliato M, Espada M, et al. Iodine intake in a population of pregnant women: INMA mother and child cohort study, Spain. *J Epidemiol Community Health*. 2010;64(12):1094-1099.
47. 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):331-337.
48. 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. *J Clin Endocrinol Metab*. 2013;98(5):1954-1962.
49. van Mil NH, Tiemeier H, Bongers-Schokking JJ, et al. Low urinary iodine excretion during early pregnancy is associated with alterations in executive functioning in children. *J Nutr*. 2012;142(12):2167-2174.
50. Abel MH, Caspersen IH, Meltzer HM, et al. Suboptimal Maternal Iodine Intake Is Associated with Impaired Child Neurodevelopment at 3 Years of Age in the Norwegian Mother and Child Cohort Study. *J Nutr*. 2017;147(7):1314-1324.
51. Andersen S, Laurberg P, Wu C, Olsen J. Attention deficit hyperactivity disorder and autism spectrum disorder in children born to mothers with thyroid dysfunction: a Danish nationwide cohort study. *BJOG Int J Obstet Gynaecol*. 2014;121(11):1365-1374.
52. Reiner O, Karzbrun E, Kshirsagar A, Kaibuchi K. Regulation of neuronal migration, an emerging topic in autism spectrum disorders. *J Neurochem*. 2016;136(3):440-456.
53. León G, Murcia M, Rebagliato M, et al. Maternal Thyroid Dysfunction during Gestation, Preterm Delivery, and Birthweight. The Infancia y Medio Ambiente Cohort, Spain. *Paediatr Perinat Epidemiol*. 2015;29(2):113-122.
54. Negro R, Schwartz A, Gismondi R, Tinelli A, Mangieri T, Stagnaro-Green A. Increased pregnancy loss rate in thyroid antibody negative women with TSH levels between 2.5 and 5.0 in the first trimester of pregnancy. *J Clin Endocrinol Metab*. 2010;95(9):E44-48.
55. Negro R, Formoso G, Mangieri T, Pezzarossa A, Dazzi D, Hassan H. Levothyroxine treatment in euthyroid pregnant women with autoimmune thyroid disease: effects on obstetrical complications. *J Clin Endocrinol Metab*. 2006;91(7):2587-2591.
56. Liu H, Shan Z, Li C, et al. Maternal subclinical hypothyroidism, thyroid autoimmunity, and the risk of miscarriage: a prospective cohort study. *Thyroid Off J Am Thyroid Assoc*. 2014;24(11):1642-1649.
57. Lepoutre T, Debiève F, Gruson D, Daumerie C. Reduction of miscarriages through universal screening and treatment of thyroid autoimmune diseases. *Gynecol Obstet Invest*. 2012;74(4):265-273.
58. 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. *Eur Thyroid J*. 2014;3(2):76-94.

59. Glinoer D. What happens to the normal thyroid during pregnancy? *Thyroid*. 1999;9(7):631–635.