

Treatment of Subclinical Hypothyroidism or Hypothyroxinemia in Pregnancy

B.M. Casey, E.A. Thom, A.M. Peaceman, M.W. Varner, Y. Sorokin, D.G. Hirtz, U.M. Reddy, R.J. Wapner, J.M. Thorp, Jr., G. Saade, A.T.N. Tita, D.J. Rouse, B. Sibai, J.D. Iams, B.M. Mercer, J. Tolosa, S.N. Caritis, and J.P. VanDorsten, for the Eunice Kennedy Shriver National Institute of Child Health and Human Development Maternal–Fetal Medicine Units Network*

ABSTRACT

BACKGROUND

Subclinical thyroid disease during pregnancy may be associated with adverse outcomes, including a lower-than-normal IQ in offspring. It is unknown whether levothyroxine treatment of women who are identified as having subclinical hypothyroidism or hypothyroxinemia during pregnancy improves cognitive function in their children.

METHODS

We screened women with a singleton pregnancy before 20 weeks of gestation for subclinical hypothyroidism, defined as a thyrotropin level of 4.00 mU or more per liter and a normal free thyroxine (T_4) level (0.86 to 1.90 ng per deciliter [11 to 24 pmol per liter]), and for hypothyroxinemia, defined as a normal thyrotropin level (0.08 to 3.99 mU per liter) and a low free T_4 level (<0.86 ng per deciliter). In separate trials for the two conditions, women were randomly assigned to receive levothyroxine or placebo. Thyroid function was assessed monthly, and the levothyroxine dose was adjusted to attain a normal thyrotropin or free T_4 level (depending on the trial), with sham adjustments for placebo. Children underwent annual developmental and behavioral testing for 5 years. The primary outcome was the IQ score at 5 years of age (or at 3 years of age if the 5-year examination was missing) or death at an age of less than 3 years.

RESULTS

A total of 677 women with subclinical hypothyroidism underwent randomization at a mean of 16.7 weeks of gestation, and 526 with hypothyroxinemia at a mean of 17.8 weeks of gestation. In the subclinical hypothyroidism trial, the median IQ score of the children was 97 (95% confidence interval [CI], 94 to 99) in the levothyroxine group and 94 (95% CI, 92 to 96) in the placebo group ($P=0.71$). In the hypothyroxinemia trial, the median IQ score was 94 (95% CI, 91 to 95) in the levothyroxine group and 91 (95% CI, 89 to 93) in the placebo group ($P=0.30$). In each trial, IQ scores were missing for 4% of the children. There were no significant between-group differences in either trial in any other neurocognitive or pregnancy outcomes or in the incidence of adverse events, which was low in both groups.

CONCLUSIONS

Treatment for subclinical hypothyroidism or hypothyroxinemia beginning between 8 and 20 weeks of gestation did not result in significantly better cognitive outcomes in children through 5 years of age than no treatment for those conditions. (Funded by the Eunice Kennedy Shriver National Institute of Child Health and Human Development and the National Institute of Neurological Disorders and Stroke; ClinicalTrials.gov number, NCT00388297.)

The authors' full names, academic degrees, and affiliations are listed in the Appendix. Address reprint requests to Dr. Casey at the Department of Obstetrics and Gynecology, University of Texas Southwestern Medical Center, 5323 Harry Hines Blvd., Dallas, TX 75235, or at brian.casey@utsouthwestern.edu.

*A complete list of the investigators in the Eunice Kennedy Shriver National Institute of Child Health and Human Development Maternal–Fetal Medicine Units Network is provided in the Supplementary Appendix, available at NEJM.org.

N Engl J Med 2017;376:815–25.

DOI: 10.1056/NEJMoa1606205

Copyright © 2017 Massachusetts Medical Society.

OBSERVATIONAL STUDIES SPANNING almost three decades suggest that subclinical thyroid disease during pregnancy is associated with adverse outcomes.¹⁻⁵ In 1999, interest in undiagnosed maternal thyroid dysfunction was heightened by studies suggesting an association between subclinical thyroid hypofunction and impaired fetal neuropsychological development.^{6,7} In one report, children of women whose serum thyrotropin levels during pregnancy were greater than the 98th percentile had a lower IQ than children of matched controls who had a normal thyrotropin level.⁶ In another study, children whose mothers had a serum free thyroxine (T_4) level of less than the 10th percentile in early pregnancy had impaired psychomotor development at 10 months of age, as compared with children whose mothers had a higher free T_4 level.⁷ Subclinical hypothyroidism has also been associated with increased risks of preterm birth, placental abruption, admission to the intensive care nursery (or neonatal intensive care unit), and other adverse pregnancy outcomes that could explain neurodevelopmental delay.⁸⁻¹¹ However, the risks of these adverse outcomes are not increased among women with hypothyroxinemia in pregnancy.^{12,13}

These findings led several professional organizations to recommend routine prenatal screening for and treatment of subclinical thyroid disease during pregnancy.¹⁴ This recommendation could affect more than 15% of pregnant women, depending on the thyrotropin and free T_4 thresholds used.¹⁵ However, the American College of Obstetricians and Gynecologists has maintained that recommendations for routine screening are premature in the absence of trials showing an improvement in these outcomes with levothyroxine treatment.¹⁶ The Controlled Antenatal Thyroid Screening (CATS) study showed that cognitive function in 3-year-old children was not better than that in controls when mothers who had been identified with subclinical hypothyroidism or hypothyroxinemia were treated with levothyroxine.¹⁷ Despite this evidence, the treatment of subclinical thyroid dysfunction is still recommended by several organizations in their clinical practice guidelines.^{18,19} The primary objective of our trials involving women with either subclinical hypothyroidism or hypothyroxinemia was to assess the effect of screening and thyroxine replacement during pregnancy on the IQ of children at 5 years of age.

METHODS

STUDY POPULATION

We conducted two multicenter, randomized, placebo-controlled trials in parallel at 15 centers within the Eunice Kennedy Shriver National Institute of Child Health and Human Development Maternal-Fetal Medicine Units Network. The protocol, available with the full text of this article at NEJM.org, was approved by the institutional review board at each center. All the women with a singleton pregnancy who presented for prenatal care before 20 weeks of gestation were invited to undergo thyroid screening for serum thyrotropin and free T_4 values. The serum samples obtained from women who provided written informed consent were analyzed at a centralized laboratory (see the Supplementary Appendix, available at NEJM.org).

The criteria that were used to diagnose subclinical hypothyroidism at the origin of the trial were a thyrotropin level of 3.00 mU or more per liter (which was presumed to correspond to the 97.5th percentile) and a free T_4 level in the normal range (0.86 to 1.90 ng per deciliter [11 to 24 pmol per liter]). After 10 months of screening, during a planned evaluation, the prevalence of subclinical hypothyroidism was found to exceed 6%. On the basis of an analysis of the 97.5th percentile in the first 15,000 women who underwent screening, the cutoff for the thyrotropin level was subsequently increased to 4.00 mU per liter. Hypothyroxinemia was defined as a thyrotropin level in the normal range (0.08 to 3.99 mU per liter) and a free T_4 level of less than 0.86 ng per deciliter.

Women with either subclinical hypothyroidism or hypothyroxinemia and an ultrasonographically verified singleton pregnancy between 8 weeks 0 days of gestation and 20 weeks 6 days of gestation met the inclusion criteria. The complete eligibility criteria are listed in the Supplementary Appendix. Women who were found to have overt hypothyroidism or hyperthyroidism at the time of screening were excluded, and obstetrical providers were notified of abnormal test results for subsequent follow-up.

TRIAL REGIMENS

Eligible women who provided written informed consent were given a 7-day supply of placebo capsules (adherence run-in phase). Women who took 50% or more of the capsules and returned within 2 weeks were randomly assigned, in a 1:1



A Quick Take
is available at
NEJM.org

ratio, to receive either levothyroxine or placebo in one of two trials, the subclinical hypothyroidism trial or the hypothyroxinemia trial, according to their results on the thyrotropin and free T_4 tests. Separate randomization sequences were prepared at the independent data coordinating center with the use of the simple urn method,²⁰ with stratification according to clinical site. Numbered trial-regimen kits were prepared, and at randomization, each patient was assigned to the next sequentially numbered kit. Blood and urine samples were obtained from all the participants for analysis, at the central laboratory, of thyroid peroxidase antibodies and urinary iodine concentration.

Participants with subclinical hypothyroidism began taking 100 μ g of levothyroxine or matching placebo daily. Participants with hypothyroxinemia began taking 50 μ g of levothyroxine or matching placebo daily; the lower dose in the hypothyroxinemia trial was intended to avoid overtreatment in women with mild suppression of free T_4 at trial entry. Women in the two trials were seen monthly, and blood samples for thyrotropin and free T_4 testing were sent to the same laboratory. Results were reported to the coordinating center, which notified the clinical center whether a dose adjustment was required according to the algorithm shown in Table S2 in the Supplementary Appendix. Sham adjustments were communicated for the placebo group. Adjustments were made within 7 days after the blood test. The goal for women with subclinical hypothyroidism was a thyrotropin level between 0.1 and 2.5 mU per liter, with a maximum daily dose of 200 μ g of levothyroxine. The goal for participants with hypothyroxinemia was a free T_4 level between 0.86 and 1.90 ng per deciliter, with the same maximum dose of levothyroxine.

Pregnancy and neonatal outcomes were abstracted from the medical records by certified research staff. The children underwent developmental testing annually for 5 years. Examiners were trained to administer each test, and they submitted a videotaped encounter to two expert psychologists for initial certification. Annual recertification was required. Research staff, examiners, and participants were unaware of the trial-group assignments.

TRIAL OUTCOMES

The primary outcome was the full-scale IQ as assessed with the use of the Wechsler Preschool

and Primary Scale of Intelligence III (WPPSI-III) at 5 years of age, or with the overall (general conceptual ability) score from the Differential Ability Scales-II (DAS) at 3 years of age if the WPPSI-III score was not available, or death before 3 years of age (because it was a competing event for IQ score). Results are expressed as age-standardized scores, with an expected population mean of 100 and a standard deviation of 15. The DAS and WPPSI-III scores correlate well ($r=0.89$). Prespecified subgroup analyses for the primary outcome were performed according to gestational age at randomization, race or ethnic group, and baseline thyroid peroxidase antibody, thyrotropin, free T_4 , and iodine levels.

Secondary outcomes in infants and children included the cognitive, motor, and language scores on the Bayley Scales of Infant Development, Third Edition (Bayley-III), at 12 months and 24 months of corrected age; DAS overall scores at 36 months of age; specific scores on the DAS (subtests regarding recall of digits forward and recognition of pictures) plus the Conners' Rating Scales-Revised at 48 months of age for assessment of attention; and scores on the Child Behavior Checklist at 36 months and 60 months of age for assessment of behavioral and social competency. Maternal and neonatal secondary outcomes included preterm delivery, pregnancy complications, fetal death, and neonatal morbidity and mortality. A complete list of secondary outcomes is provided in the Supplementary Appendix.

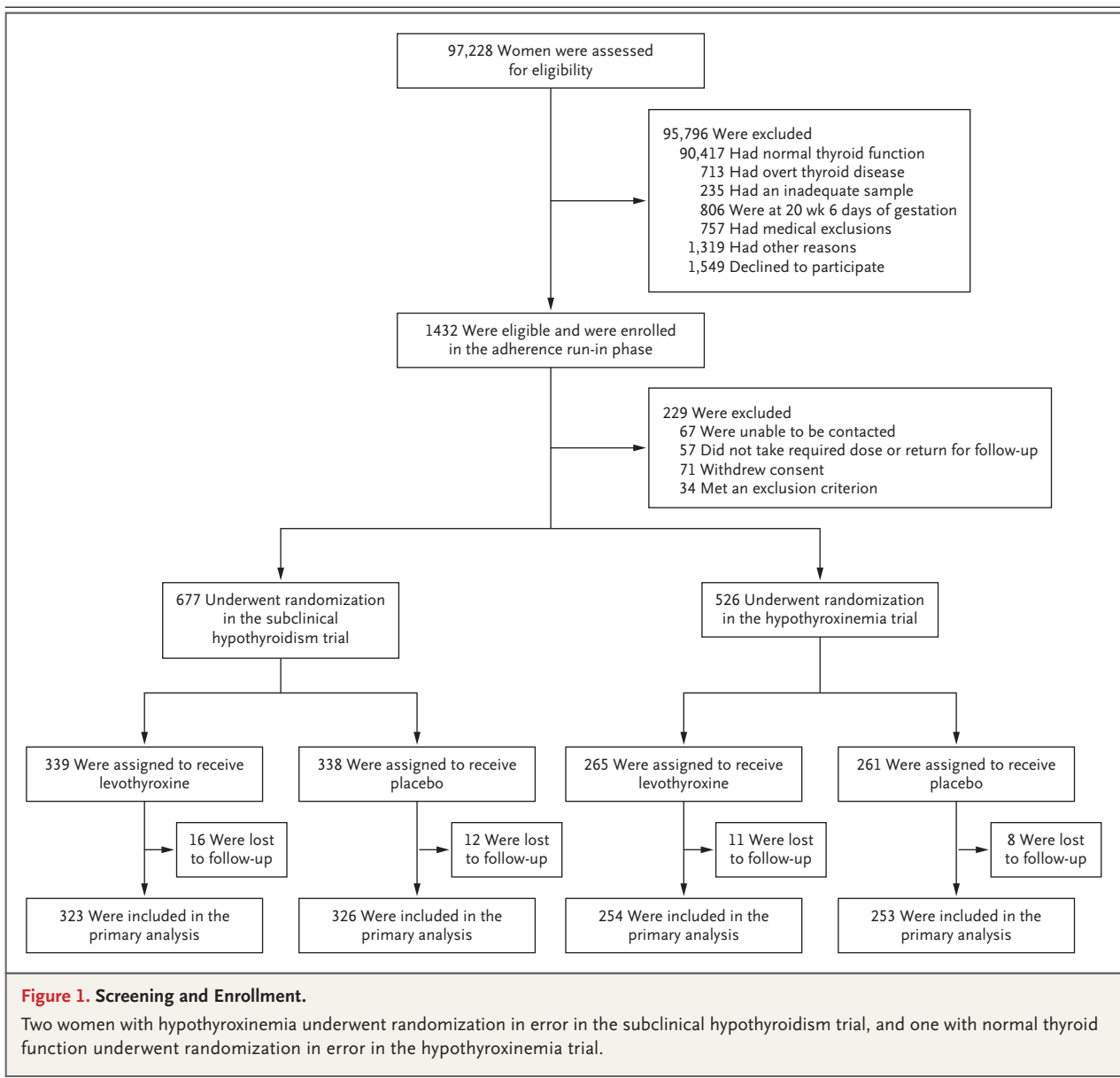
STATISTICAL ANALYSIS

Assuming an analysis that would be based on a Wilcoxon test with a 5-point difference between the group median IQ scores, a death rate before the age of 3 years (including spontaneous abortions, stillbirths, and neonatal and infant deaths) of 2 to 5%, and 15% loss to follow-up, we initially calculated that a sample of 500 patients in each trial (250 women per group) would provide the trial with a power of at least 80%, at a two-sided type I error rate of 5%. When the eligibility criteria for the subclinical hypothyroidism trial changed, the sample was adjusted to 670 participants under the assumption that there would be no between-group difference in the children's IQ scores if the mothers' thyrotropin levels were between 3.00 and 4.00 mU per liter, but there would be a between-group difference of 5 points if the mothers' thyrotropin levels were 4.00 mU or more per liter.

The analysis was performed according to the intention-to-treat principle. The primary outcome and other continuous variables were compared with the use of the Wilcoxon test or van Elteren's test for stratified analysis. For the primary outcome, death before 3 years of age was assigned a score of 0 (lowest possible rank) and was included in the estimation of the median. Differences between groups were estimated with the use of the Hodges–Lehmann estimator, and 95% confidence intervals were reported. Categorical variables were analyzed with the use of the chi-

square test or Fisher's exact test, as appropriate. To test for interaction in the prespecified subgroup analyses, we used regression with normal-order scores.

An independent data and safety monitoring committee monitored the trials. Since recruitment was completed before any 5-year outcomes were available, there was no interim analysis of the primary outcome. For secondary outcomes, nominal P values of less than 0.05 were considered to indicate statistical significance. No adjustments were made for multiple comparisons.



RESULTS

TRIAL POPULATIONS

From October 2006 through October 2009, a total of 97,228 pregnant women underwent thyroid screening. A total of 90,417 women (93%) had results that were considered to be normal, 463 (<1%; prevalence, 5 cases per 1000 pregnant women) had overt hypothyroidism, and 250 (<1%; prevalence, 3 cases per 1000 pregnant women) had overt hyperthyroidism. Of the 3057 women (3%) with subclinical hypothyroidism, 800 were eligible and consented to participate in the adherence run-in phase, and 677 underwent randomization. Of the 2805 women (3%) with hypothyroxinemia, 632 were eligible and consented to participate in the adherence run-in phase, and

526 underwent randomization (Fig. 1). IQ scores were not available for 4% of the offspring in each trial (28 offspring in the subclinical hypothyroidism trial and 19 in the hypothyroxinemia trial).

There were no significant differences at baseline between the levothyroxine group and the placebo group in either trial (Table 1). The population in each trial was iodine-sufficient (median urinary iodine concentration, $\geq 150 \mu\text{g}$ per liter).²¹ On average, women in the subclinical hypothyroidism trial underwent randomization before 17 weeks of gestation, and 93% of the women in the levothyroxine group had a thyrotropin level between 0.1 and 2.5 mU per liter by a median gestational age of 21 weeks. In the hypothyroxinemia trial, women underwent randomization

Table 1. Maternal Characteristics at Baseline.*

Characteristic	Subclinical Hypothyroidism		Hypothyroxinemia	
	Levothyroxine (N=339)	Placebo (N=338)	Levothyroxine (N=265)	Placebo (N=261)
Age — yr	27.7 \pm 5.7	27.3 \pm 5.7	27.8 \pm 5.7	28.0 \pm 5.8
Race or ethnic group — no. (%) [†]				
Black	27 (8)	25 (7)	61 (23)	65 (25)
Hispanic	195 (58)	185 (55)	131 (49)	125 (48)
White	109 (32)	117 (35)	69 (26)	69 (26)
Other	8 (2)	11 (3)	4 (2)	2 (1)
Body-mass index [‡]	28.1 \pm 6.4	28.2 \pm 6.4	30.3 \pm 6.4	30.2 \pm 7.1
Nulliparous — no. (%)	124 (37)	134 (40)	69 (26)	64 (25)
Baseline thyrotropin — mU/liter				
Median	4.5	4.3	1.5	1.4
95% CI	4.4–4.7	4.2–4.5	1.4–1.6	1.3–1.5
Baseline free thyroxine — ng/dl [§]				
Median	1.01	1.02	0.83	0.83
95% CI	1.00–1.02	1.01–1.04	0.82–0.83	0.82–0.83
Urinary iodine — μg /liter [¶]				
Median	199	196	185	191
95% CI	184–238	172–229	167–219	164–208
No. of weeks of gestation at randomization	16.6 \pm 3.0	16.7 \pm 3.0	18.0 \pm 2.8	17.7 \pm 2.9

* Plus-minus values are means \pm SD. There were no significant differences at baseline between the levothyroxine group and the placebo group in either trial ($P>0.05$). CI denotes confidence interval.

[†] Race and ethnic group were determined by the research nurses.

[‡] The body-mass index is the weight in kilograms divided by the square of the height in meters.

[§] To convert thyroxine values to picomoles per liter, multiply by 12.87.

[¶] One patient in the levothyroxine group and one in the placebo group in the subclinical hypothyroidism trial were missing the urinary iodine measurement.

at a mean gestational age of 18 weeks, and 83% of the women in the levothyroxine group met the treatment goal (free T_4 level between 0.86 and 1.90 ng per deciliter) by a median gestational age of 23 weeks.

PREGNANCY AND NEONATAL OUTCOMES

The frequencies of adverse pregnancy and neonatal outcomes did not differ significantly be-

tween the groups in either trial (Table 2). After neonatal discharge, there was one death in the subclinical hypothyroidism trial and none in the hypothyroxinemia trial. Two women were lost to follow-up before delivery. There was no significant difference between the levothyroxine group and the placebo group in either trial with regard to the mean gestational age at delivery (subclinical hypothyroidism trial: 39.1 ± 2.5 weeks and

Table 2. Pregnancy and Neonatal Outcomes.*

Outcome	Subclinical Hypothyroidism			Hypothyroxinemia		
	Levothyroxine (N=339)	Placebo (N=338)	P Value	Levothyroxine (N=263)	Placebo (N=261)	P Value
Maternal						
Week of gestation at delivery	39.1 ± 2.5	38.9 ± 3.1	0.57	39.0 ± 2.4	38.8 ± 3.1	0.46
Preterm birth — no. (%)						
At <34 wk	9 (3)	10 (3)	0.81	10 (4)	7 (3)	0.47
At <37 wk	31 (9)	37 (11)	0.44	31 (12)	20 (8)	0.11
Placental abruption — no. (%)	1 (<1)	5 (1)	0.12	3 (1)	2 (1)	1.00
Gestational hypertension — no. (%)	33 (10)	36 (11)	0.69	20 (8)	24 (9)	0.51
Preeclampsia — no. (%)	22 (6)	20 (6)	0.76	9 (3)	11 (4)	0.64
Gestational diabetes — no. (%)	25 (7)	22 (7)	0.66	21 (8)	24 (9)	0.62
Fetal or neonatal†						
Stillbirth or miscarriage — no. (%)	4 (1)	7 (2)	0.36	2 (1)	5 (2)	0.28
Neonatal death — no. (%)	0	1 (<1)	0.50	1 (<1)	1 (<1)	1.00
Apgar score at 1 min <4 — no. (%)	6 (2)	7 (2)	0.76	6 (2)	7 (3)	0.76
Apgar score at 5 min <7 — no. (%)	2 (1)	3 (1)	0.69	2 (1)	4 (2)	0.45
Admission to NICU — no. (%)	29 (9)	21 (6)	0.24	31 (12)	31 (12)	0.97
Birth weight <10th percentile — no. (%)	33 (10)	27 (8)	0.45	23 (9)	20 (8)	0.68
Head circumference — cm	33.9 ± 1.8	33.9 ± 1.7	0.46	33.9 ± 1.8	34.2 ± 1.6	0.19
Respiratory distress syndrome — no. (%)	9 (3)	6 (2)	0.45	4 (2)	5 (2)	0.75
Retinopathy of prematurity — no. (%)	1 (<1)	0	1.00	0	0	—
Necrotizing enterocolitis — no. (%)	1 (<1)	1 (<1)	1.00	2 (1)	0	0.50
Bronchopulmonary dysplasia — no. (%)	0	1 (<1)	0.50	0	1 (<1)	0.49
Composite neonatal outcome — no. (%)‡	7 (2)	12 (4)	0.24	5 (2)	7 (3)	0.55
Respiratory therapy ≥ 1 day — no. (%)	11 (3)	11 (3)	0.99	13 (5)	12 (5)	0.85
No. of days in hospital nursery			0.43			0.39
Median	2	2		2	2	
95% CI	2–2	2–2		2–2	2–2	

* Plus-minus values are means \pm SD. NICU denotes neonatal intensive care unit.

† Analyses of neonatal outcomes of the respiratory distress syndrome, retinopathy of prematurity, necrotizing enterocolitis, and bronchopulmonary dysplasia did not include stillbirths or miscarriages in the levothyroxine group (four offspring) or the placebo group (seven) in the subclinical hypothyroidism trial or in the levothyroxine group (two) or the placebo group (four) in the hypothyroxinemia trial. One infant in the placebo group in the hypothyroxemia trial was born out of network and was also not included in the analyses.

‡ The composite neonatal outcome was defined as periventricular leukomalacia, intraventricular hemorrhage of grade III or IV, necrotizing enterocolitis (stage \geq II), severe retinopathy of prematurity (stage \geq III), the severe respiratory distress syndrome, bronchopulmonary dysplasia, neonatal death, stillbirth, or serious infectious complication.

38.9±3.1 weeks, respectively; $P=0.57$; and hypothyroxinemia trial: 39.0±2.4 weeks and 38.8±3.1 weeks, respectively; $P=0.46$). There were no significant differences in overall or serious adverse events between the two groups in either trial; serious adverse events were rare in the two trials (Tables S4 and S5 in the Supplementary Appendix).

NEURODEVELOPMENTAL AND BEHAVIORAL OUTCOMES

In the subclinical hypothyroidism trial, data on the primary outcome were available for 649 offspring (96%) (Table 3). A total of 11 children had DAS scores at 3 years that were substituted for the WPPSI-III IQ score, and 13 offspring died

Table 3. Developmental and Behavioral Outcomes in Offspring of Mothers with Subclinical Hypothyroidism.*

Outcome	Levothyroxine		Placebo		Difference (95% CI)†	P Value
	No. of Children	Median Value (95% CI)	No. of Children	Median Value (95% CI)		
Primary outcome‡	323	97 (94 to 99)	326	94 (92 to 96)	0 (–3 to 2)	0.71
Bayley-III score§						
At 12 mo						
Cognitive	311	100 (95 to 100)	315	100 (95 to 100)	0 (0 to 0)	0.63
Motor	312	97 (97 to 97)	314	97 (97 to 97)	0 (0 to 3)	0.83
Language	309	94 (94 to 97)	312	94 (94 to 97)	0 (0 to 3)	0.48
At 24 mo						
Cognitive	308	90 (90 to 90)	302	90 (90 to 90)	0 (0 to 0)	0.59
Motor	304	97 (97 to 97)	300	97 (97 to 100)	0 (0 to 3)	0.31
Language	300	89 (89 to 91)	296	91 (89 to 94)	0 (0 to 3)	0.30
Differential Ability Scales–II scores						
Overall at 36 mo	304	90 (88 to 93)	308	90 (87 to 93)	0 (–2 to 3)	0.90
Recall of digits forward at 48 mo	298	84 (76 to 91)	299	84 (76 to 91)	0 (–5 to 7)	0.60
Recognition of pictures at 48 mo	298	74 (74 to 80)	302	74 (74 to 80)	0 (–6 to 0)	0.52
Child Behavior Checklist T score¶						
At 36 mo	306	46 (45 to 48)	309	46 (45 to 48)	0 (–2 to 2)	0.99
At 60 mo	314	44 (43 to 46)	313	44 (42 to 46)	0 (–2 to 2)	0.96
Conners' Rating Scales–Revised ADHD score at 48 mo	302	48 (47 to 49)	303	49 (47 to 51)	0 (–1 to 2)	0.37
WPPSI-III at 60 mo	311	97 (95 to 99)	314	95 (93 to 97)	0 (–3 to 2)	0.89

* For all outcomes except the primary outcome, the potential follow-up cohort consisted of 335 children in the levothyroxine group and 329 in the placebo group (offspring who were not lost to follow-up at maternal delivery, who were discharged alive after birth, and who did not die before 1 year of age).

† Shown is the Hodges–Lehmann estimate of the absolute difference between the placebo group and the levothyroxine group. The Hodges–Lehmann estimate is the median of all paired differences between the observations in the two samples, and negative numbers reflect lower scores in the placebo group.

‡ The primary outcome was death or IQ score at 5 years of age (or at 3 years of age if the 5-year examination was missing). The full-scale IQ was assessed with the use of the Wechsler Preschool and Primary Scale of Intelligence III (WPPSI-III) at 5 years of age or the overall (general conceptual ability) score from the Differential Ability Scales–II at 3 years of age if the WPPSI-III score was not available. Results are expressed as an age-standardized score, with an expected population mean of 100 and a standard deviation of 15. Death before 3 years of age was assigned a score of 0 (lowest possible rank) and was included in the estimation of the median.

§ Results on the Bayley Scales of Infant Development, Third Edition (Bayley-III) are expressed as an age-standardized score, with an expected population mean of 100 and a standard deviation of 15.

¶ A Child Behavior Checklist T score of less than 60 is considered to be in the normal range, a T score of 60 to 63 is a borderline score, and a T score of more than 63 is in the clinical range.¹⁷

|| The Conners' Rating Scales–Revised were used to assess attention deficit–hyperactivity disorder (ADHD). A T score of 45 to 55 is considered to be typical or average; a T score of 44 or less is not a concern, a T score of 56 to 60 is considered to be a borderline score, and a T score of 61 or higher indicates a possible or clinically significant problem.

before 3 years of age (fetal death, neonatal death, or infant death), including 4 in the levothyroxine group and 9 in the placebo group ($P=0.16$). The median IQ score was 97 (95% confidence interval [CI], 94 to 99) in the levothyroxine group and 94 (95% CI, 92 to 96) in the placebo group ($P=0.71$). Annual developmental-testing scores and the results of the behavioral and attention assessments (the Child Behavior Checklist and Connors' Rating Scales–Revised, respectively) did not differ significantly between the groups for any test.

In the hypothyroxinemia trial, data on the primary outcome were available for 507 offspring (96%) (Table 4). A total of 12 children had a DAS score substituted for the WPPSI-III IQ

score, and 9 offspring died before 3 years of age, including 3 in the levothyroxine group and 6 in the placebo group ($P=0.34$). The median IQ score was 94 (95% CI, 91 to 95) in the levothyroxine group and 91 (95% CI, 89 to 93) in the placebo group ($P=0.30$). There were no significant between-group differences in any of the annual measures.

The median T scores for the Child Behavior Checklist and the Connors' Rating Scales–Revised in all comparison groups (in the two trials) were within the normal range.^{22,23} None of the subgroup interaction tests were significant (Table S6 in the Supplementary Appendix). Stratification according to clinical center did not materially alter the results in either trial.

Table 4. Developmental and Behavioral Outcomes in Offspring of Mothers with Hypothyroxinemia.*

Outcome	Levothyroxine		Placebo		Difference (95% CI)†	P Value
	No. of Children	Median Value (95% CI)	No. of Children	Median Value (95% CI)		
Primary outcome	254	94 (91 to 95)	253	91 (89 to 93)	–1 (–4 to 1)	0.30
Bayley-III score						
At 12 mo						
Cognitive	247	100 (100 to 100)	238	100 (100 to 100)	0 (0 to 0)	0.89
Motor	246	97 (94 to 97)	236	97 (94 to 97)	0 (0 to 3)	0.54
Language	246	94 (91 to 94)	237	94 (91 to 97)	0 (–3 to 3)	0.92
At 24 mo						
Cognitive	235	90 (85 to 90)	235	90 (85 to 90)	0 (0 to 0)	0.70
Motor	233	97 (94 to 100)	232	97 (94 to 97)	0 (–3 to 0)	0.20
Language	232	89 (89 to 94)	229	89 (89 to 91)	0 (–3 to 2)	0.71
Differential Ability Scales–II scores						
Overall at 36 mo	244	90 (87 to 92)	235	89 (87 to 91)	–1 (–3 to 2)	0.64
Recall of digits forward at 48 mo	236	91 (84 to 99)	224	84 (84 to 91)	0 (–8 to 0)	0.22
Recognition of pictures at 48 mo	234	74 (74 to 80)	226	74 (74 to 80)	0 (–4 to 0)	0.91
Child Behavior Checklist T score						
At 36 mo	244	48 (46 to 50)	237	48 (45 to 49)	0 (–2 to 2)	0.65
At 60 mo	244	45 (43 to 46)	243	43 (42 to 45)	–1 (–3 to 1)	0.44
Conner's Rating Scales–Revised ADHD score at 48 mo	238	50 (49 to 51)	228	49 (48 to 51)	0 (–2 to 2)	0.98
WPPSI-III score at 60 mo	243	94 (91 to 95)	243	92 (90 to 95)	–1 (–3 to 2)	0.48

* For all outcomes except the primary outcome, the potential follow-up cohort consisted of 260 children in the levothyroxine group and 255 children in the placebo group (those who were not lost to follow-up at maternal delivery, who were discharged alive after birth, and who did not die before 1 year of age).

† Shown is the Hodges–Lehmann estimate of the absolute difference between the placebo group and the levothyroxine group. The Hodges–Lehmann estimate is the median of all paired differences between the observations in the two samples, and negative numbers reflect lower scores in the placebo group.

DISCUSSION

These two parallel, randomized, placebo-controlled trials involving women with subclinical hypothyroidism or hypothyroxinemia in the first half of pregnancy showed no significant effect of thyroid hormone-replacement therapy on the cognitive function of the children or on other indexes of neurodevelopment through 5 years of age. There were no significant differences in measures of behavior, attention deficits, or hyperactivity in either trial. Moreover, treatment of women who had either an elevated thyrotropin level or a low free T_4 level had no significant effect on pregnancy or neonatal outcomes.

A previous study indicated that children who were born to untreated women who had a thyrotropin value above the 98th percentile had diminished school performance and an average IQ that was 7 points lower than the average IQ of control children.⁶ These results are often offered as evidence that the offspring of women with subclinical hypothyroidism during pregnancy are at risk for subnormal brain function. However, most women who were included in this previous study had a thyrotropin level greater than 10 mU per liter and a free T_4 level that was more consistent with the diagnosis of overt hypothyroidism.⁶ Studies involving children of women with hypothyroxinemia that was identified before 12 weeks of gestation also showed significantly lower scores on the Bayley mental and psychomotor subscales at 2 years of age than infants of euthyroid women.^{7,24} These studies suggested that offspring would benefit from maternal treatment for subclinical thyroid disease and resulted in recommendations by some medical organizations for routine maternal screening and treatment to prevent subnormal cognitive development in offspring.¹⁴

The results of our trials are consistent with those of the CATS study, which was a thyroid-screening trial involving 21,846 pregnant women, primarily from the United Kingdom. Women in that trial were either screened immediately and treated with levothyroxine if they were identified as having subclinical hypothyroidism or hypothyroxinemia or had their serum frozen to be analyzed on completion of the pregnancy. The results of IQ testing of the children at 3 years of age did not differ significantly between children whose mothers had been immediately treated during the pregnancy and those whose mothers

had treatment that was deferred. Thyroid hormone-replacement therapy in the CATS study was initiated at a median gestational age of 13 weeks 3 days, as compared with 16 weeks 4 days and 18 weeks in our two trials.¹⁷ However, 24% of the children in the CATS study were lost to follow-up. In the current trials, the follow-up rate at 5 years of age was more than 92%.

Several studies have also suggested that either a high maternal thyrotropin level or hypothyroxinemia is associated with externalizing behavioral problems such as attention deficit-hyperactivity disorder (ADHD).^{25,26} In the Generation R Study, 8-year-old children of women from an iodine-deficient geographic area who had been identified with hypothyroxinemia before 20 weeks of gestation were found to have higher ADHD index scores than children of women without hypothyroxinemia.²⁶ This association persisted after adjustment for IQ. We did not identify any significant between-group differences in ADHD index scores in either of our trials, and the scores were well within the normal range. Moreover, the Child Behavior Checklist scores did not reveal any evidence of behavioral problems in children whose mothers with subclinical hypothyroidism or hypothyroxinemia were in the placebo group. The CATS study also did not detect any behavioral improvements in children who had been exposed to thyroid hormone replacement.¹⁷

Subclinical hypothyroidism has been associated with several obstetrical complications,⁸⁻¹¹ but there has been no direct evidence that levothyroxine therapy reduces these risks.¹⁹ One study involving women with thyroid peroxidase antibodies showed a lower rate of preterm delivery among women treated with levothyroxine during pregnancy than among those who were untreated²⁷; however, these women had normal thyroid-function tests. We did not detect any significant improvement in pregnancy or neonatal outcomes that was associated with levothyroxine therapy in women with subclinical thyroid hypofunction.

A limitation of the two trials is the relatively late time during gestation at which women were randomly assigned to the trial groups. The fetal thyroid gland begins producing thyroid hormone between 10 weeks and 12 weeks of gestation, and on average, women underwent randomization in our trials several weeks after this time (at a mean of 16.7 weeks of gestation in mothers with subclinical hypothyroidism and 17.8 weeks

of gestation in those with hypothyroxinemia). However, in a previous study, a small cohort of children whose mothers had a low free T_4 level in the first trimester but whose level increased before 24 weeks of gestation had scores that were similar to those of children of euthyroid mothers, which suggests a possible benefit with the initiation of supplementation after the first trimester.²⁴

In our trials, the mothers who were treated with levothyroxine met therapy goals by a median gestational age of less than 24 weeks. We also found no significant interaction according to the time of initiation of therapy (<17 weeks of gestation vs. ≥17 weeks of gestation). Furthermore, post hoc analyses of the CATS study did not suggest any significant differences in IQ scores between the treated group and the control group among women who were screened before 14 weeks of gestation or who met the target thyrotropin levels within 6 weeks after screening.¹⁷ Although we did not enroll women earlier than at 8 weeks of gestation in order to avoid enrolling women who might have an early miscarriage, we believe that these trials, which were performed at centers that were actively screening women on presentation for prenatal care, probably reflect what could be accomplished by means of routine thyroid screening during pregnancy in the United States.

In conclusion, on the basis of a comprehensive battery of tests through 5 years of age, we did not find significantly better neurodevelopmental outcomes in children whose mothers had received thyroxine treatment for subclinical hypothyroidism or hypothyroxinemia during pregnancy than in children whose mothers did not receive such treatment. Our trials also showed no significant effect of thyroxine treatment on pregnancy and neonatal outcomes.

The views expressed in this article are those of the authors and do not necessarily represent the views of the National Institutes of Health.

Supported by grants (HD34116, HD40512, HD27917, HD34208, HD40485, HD40560, HD53097, HD27869, HD40500, HD40545, HD27915, HD40544, HD53118, HD21410, and HD36801) from the Eunice Kennedy Shriver National Institute of Child Health and Human Development and the National Institute of Neurological Disorders and Stroke.

No potential conflict of interest relevant to this article was reported.

Disclosure forms provided by the authors are available with the full text of this article at NEJM.org.

We thank Lisa Moseley, R.N., B.S.N., and Gail Mallet, R.N., B.S.N., C.C.R.C., for protocol development and coordination between clinical research centers; Barbara Jones-Binns, J.D., M.P.H., for protocol and data management, overall coordination, and quality control; Lisa Mele, Sc.M., for statistical analysis; Victoria Watson, M.S., and Terri Leach, M.S., for training and certification of examiners; M. María Elena López Ramírez, L. Natalia Aguilar, Paul J. Lamothe, M.D., and Maria Jose Rangel, M.D., of the American British Cowdray Medical Center, Mexico City, for assistance in examining children and obtaining outcomes in mothers who moved to Mexico; and Catherine Y. Spong, M.D., for protocol development and oversight.

APPENDIX

The authors' full names and academic degrees are as follows: Brian M. Casey, M.D., Elizabeth A. Thom, Ph.D., Alan M. Peaceman, M.D., Michael W. Varner, M.D., Yoram Sorokin, M.D., Deborah G. Hirtz, M.D., Uma M. Reddy, M.D., M.P.H., Ronald J. Wapner, M.D., John M. Thorp, Jr., M.D., George Saade, M.D., Alan T.N. Tita, M.D., Ph.D., Dwight J. Rouse, M.D., Baha Sibai, M.D., Jay D. Iams, M.D., Brian M. Mercer, M.D., Jorge Tolosa, M.D., Steve N. Caritis, M.D., and J. Peter VanDorsten, M.D.

The authors' affiliations are as follows: the University of Texas Southwestern Medical Center, Dallas (B.M.C.), the University of Texas Medical Branch, Galveston (G.S.), and the University of Texas Health Science Center at Houston, McGovern Medical School—Children's Memorial Hermann Hospital, Houston (B.S.) — all in Texas; George Washington University Biostatistics Center, Washington, DC (E.A.T.); Northwestern University, Chicago (A.M.P.); the University of Utah Health Sciences Center, Salt Lake City (M.W.V.); Wayne State University, Detroit (Y.S.); the National Institute of Neurological Disorders and Stroke (D.G.H.) and the Eunice Kennedy Shriver National Institute of Child Health and Human Development (U.M.R.), Bethesda, MD; Columbia University, New York (R.J.W.); the University of North Carolina at Chapel Hill, Chapel Hill (J.M.T.); the University of Alabama at Birmingham, Birmingham (A.T.N.T.); Brown University, Providence, RI (D.J.R.); Ohio State University, Columbus (J.D.I.), and MetroHealth Medical Center—Case Western Reserve University, Cleveland (B.M.M.) — both in Ohio; Oregon Health and Science University, Portland (J.T.); University of Pittsburgh, Pittsburgh (S.N.C.); and the Medical University of South Carolina, Charleston (J.P.V.).

REFERENCES

1. Davis LE, Leveno KJ, Cunningham FG. Hypothyroidism complicating pregnancy. *Obstet Gynecol* 1988;72:108-12.
2. Leung AS, Millar LK, Koonings PP, Montoro M, Mestman JH. Perinatal outcome in hypothyroid pregnancies. *Obstet Gynecol* 1993;81:349-53.
3. Abalovich M, Gutierrez S, Alcaraz G, Maccallini G, Garcia A, Levalle O. Overt and subclinical hypothyroidism complicating pregnancy. *Thyroid* 2002;12:63-8.
4. Männistö T, Väärasmäki M, Pouta A, et al. Perinatal outcome of children born to mothers with thyroid dysfunction or antibodies: a prospective population-based cohort study. *J Clin Endocrinol Metab* 2009;94:772-9.
5. Su PY, Huang K, Hao JH, 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. *J Clin Endocrinol Metab* 2011;96:3234-41.
6. Haddow JE, Palomaki GE, Allan WC, et al. Maternal thyroid deficiency during pregnancy and subsequent neuropsychological

- logical development of the child. *N Engl J Med* 1999;341:549-55.
7. Pop VJ, Kuipers 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;50:149-55.
 8. Tudela CM, Casey BM, McIntire DD, Cunningham FG. Relationship of subclinical thyroid disease to the incidence of gestational diabetes. *Obstet Gynecol* 2012;119:983-8.
 9. Wilson KL, Casey BM, McIntire DD, Halvorson LM, Cunningham FG. Subclinical thyroid disease and the incidence of hypertension in pregnancy. *Obstet Gynecol* 2012;119:315-20.
 10. Casey BM, Dashe JS, Wells CE, et al. Subclinical hypothyroidism and pregnancy outcomes. *Obstet Gynecol* 2005;105:239-45.
 11. Allan WC, Haddow JE, Palomaki GE, et al. Maternal thyroid deficiency and pregnancy complications: implications for population screening. *J Med Screen* 2000;7:127-30.
 12. Casey BM, Dashe JS, Spong CY, McIntire DD, Leveno KJ, Cunningham GF. Perinatal significance of isolated maternal hypothyroxinemia identified in the first half of pregnancy. *Obstet Gynecol* 2007;109:1129-35.
 13. Cleary-Goldman J, Malone FD, Lambert-Messerlian G, et al. Maternal thyroid hypofunction and pregnancy outcome. *Obstet Gynecol* 2008;112:85-92.
 14. Gharib H, Tuttle RM, Baskin HJ, Fish LH, Singer PA, McDermott MT. Subclinical thyroid dysfunction: a joint statement on management from the American Association of Clinical Endocrinologists, the American Thyroid Association, and the Endocrine Society. *J Clin Endocrinol Metab* 2005;90:581-5.
 15. Blatt AJ, Nakamoto JM, Kaufman HW. National status of testing for hypothyroidism during pregnancy and postpartum. *J Clin Endocrinol Metab* 2012;97:777-84.
 16. American College of Obstetrics and Gynecology. ACOG practice bulletin: thyroid disease in pregnancy. Number 37, August 2002. *Int J Gynaecol Obstet* 2002;79:171-80.
 17. Lazarus JH, Bestwick JP, Channon S, et al. Antenatal thyroid screening and childhood cognitive function. *N Engl J Med* 2012;366:493-501.
 18. Abalovich M, Amino N, Barbour LA, et al. Management of thyroid dysfunction during pregnancy and postpartum: an Endocrine Society Clinical Practice Guideline. *J Clin Endocrinol Metab* 2007;92:Suppl:S1-S47.
 19. Stagnaro-Green A, Abalovich M, Alexander E, et al. Guidelines of the American Thyroid Association for the diagnosis and management of thyroid disease during pregnancy and postpartum. *Thyroid* 2011;21:1081-125.
 20. Rosenberger W, Lachin JM. Randomization in clinical trials. New York: Wiley, 2002.
 21. Assessment of iodine deficiency disorders and monitoring their elimination: a guide for programme managers, 3rd ed. Geneva: World Health Organization, 2007 (http://apps.who.int/iris/bitstream/10665/43781/1/9789241595827_eng.pdf).
 22. Conners CK, Sitarenios G, Parker JD, Epstein JN. The revised Conners' Parent Rating Scale (CPRS-R): factor structure, reliability, and criterion validity. *J Abnorm Child Psychol* 1998;26:257-68.
 23. Achenbach TM, Howell CT, Quay HC, Conners CK. National survey of problems and competencies among four- to sixteen-year-olds: parents' reports for normative and clinical samples. *Monogr Soc Res Child Dev* 1991;56:1-131.
 24. 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;59:282-8.
 25. Vermiglio F, Lo Presti VP, Moleti M, 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. *J Clin Endocrinol Metab* 2004;89:6054-60.
 26. Modesto T, Tiemeier H, Peeters RP, et al. Maternal mild thyroid insufficiency in early pregnancy and attention-deficit/hyperactivity disorder symptoms in children. *JAMA Pediatr* 2015;169:838-45.
 27. 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;91:2587-91.

Copyright © 2017 Massachusetts Medical Society.

TRACK THIS ARTICLE'S IMPACT AND REACH

Visit the article page at NEJM.org and click on the Metrics tab for a dashboard that logs views, citations, media references, and commentary, with easy linking. Learn more at www.nejm.org/page/article-metrics-faq.