Starting Dose of Levothyroxine for the Treatment of Congenital Hypothyroidism

A Systematic Review

Ihor Hrytsiuk, MB, ChB; Ruth Gilbert, MD, MSc(Epid), FRCPC; Stuart Logan, MB, ChB, MSc(Epid), FRCPC; Sima Pindoria, MSc(Med Stats); Charles G. D. Brook, MD, FRCP, FRCPC

Objective: To determine the effect of levothyroxine sodium starting dose on cognitive development, growth, or behavior in children with congenital hypothyroidism identified by neonatal screening.

Design: Systematic review of cohort studies. Two analyses were performed: a between-study comparison of mean starting dose with mean developmental score and an analysis of the within-study effects of starting dose on cognitive development, growth, or behavior.

Results: The between-study comparison (14 cohort studies based on 1321 patients) found that the standardized mean IQ or developmental quotient scores ranged from 90 to 115 but were not associated with the mean starting dose of levothyroxine ($P=0.48$). The within-study comparison of 4 cohort studies (based on 558 patients) that reported the effect of the starting dose of levothyroxine on cognitive development found no consistent effects. There was weak evidence for an effect of starting dose on growth (1 study) and on behavior problems (1 study).

Conclusions: The evidence for an effect of starting dose of levothyroxine on cognitive development, growth, or behavior is too weak to justify recommendations in favor of high- or standard-dose regimens. More reliable information, based on a randomized controlled trial of starting dose or a meta-analysis of the individual patient data currently available, is required to inform treatment policies.

Arch Ophthalmol. 2002;120:485-491

Neonatal screening for congenital hypothyroidism was introduced in the early 1970s in response to the finding that children who were diagnosed based on symptoms of congenital hypothyroidism had better developmental outcomes the earlier treatment was started. The screening test involves the detection of raised levels of thyroid-stimulating hormone (TSH) and/or low levels of thyroxine (T4) in filter paper blood samples collected shortly after birth. The sensitivity of these screening tests is 90% to 100% depending on the method used. Confirmation of the diagnosis is based on sequential analysis of serum levels of free thyroxine and TSH to rule out transient hypothyroidism.

Although early studies clearly support the treatment of children with hypothyroidism, evidence to support specific dose regimens is lacking. Replacement therapy for congenital hypothyroidism was initially determined empirically and was based on the use of desiccated thyroid. In the early 1970s, children were treated with a combination of triiodothyronine (50 µg/m² per day) and standardized preparations from desiccated thyroid or levothyroxine sodium (100 µg/kg per day), and later with levothyroxine sodium alone at doses of approximately 100 µg/m² per day (4-8 µg/kg per day). A starting dose of levothyroxine sodium has gradually increased since the introduction of neonatal screening, and current regimens vary from 5 to 10 µg/kg per day in some centers to 10 to 15 µg/kg per day (or 50 µg/d as a uniform dose) in others. Starting doses of 10 to 15 µg/kg per day have been reported to result in normalization of total thyroxine levels within a few days to 3 weeks, whereas doses of less than 8 µg/kg per day result in normalization within 6 to 8 weeks. However, there is a lack of evidence that earlier normalization of biochemical markers improves developmental outcome or growth after taking into account the severity of hypothyroidism at diagnosis. In addition, there are possible adverse effects of high-dose regimens and elevated T4 levels on temperament and behavior in later childhood.
METHODS
LITERATURE SEARCH

We conducted searches of the Cochrane Trials Register, Medline, and EMBASE (to June 1999), using the terms “congenital,” “hypo/levothyrox,” “myxedem,” “cretin,” “levothyrox,” “l-thyrox,” “drug,” “therap*,” “treatm*,” “manag*”). (An asterisk symbol is used to include all terms that contain the preceding word or word part.) There were no language restrictions. In addition, we hand-searched reference lists in review articles, relevant textbooks, and expert committee reports. Forward citation searches were conducted for all studies that met the inclusion criteria up to March 2000.

STUDY SELECTION

One reviewer (I.H.) scanned the titles and abstracts of all studies generated by the search. Hard copies of all potentially relevant articles were retrieved. These were assessed by 2 independent reviewers (R.G. and Tessa Parsons, PhD) to determine whether they met the inclusion criteria.

The inclusion criteria for the between-study comparison required that participants were children with congenital hypothyroidism identified by neonatal screening whose global development or cognitive development had been assessed using a population-standardized measure. For the within-study comparison, inclusion criteria required that the study must (1) be a cohort study of children with congenital hypothyroidism identified by neonatal screening and (2) either include an analysis of the relationship between starting dose and any measure of global or cognitive development or growth, adjusted for severity of hypothyroidism at diagnosis based on serum T4 levels or bone age or include an analysis of the relationship between individual starting dose and any measure of behavior.

DATA EXTRACTION

Two reviewers (I.H. and R.G.) independently extracted the data. For the between-study comparison, we recorded the mean starting dose of levothyroxine, the number of patients identified with congenital hypothyroidism, the number of patients excluded from the developmental assessment and reasons for this, the number undergoing testing for IQ developmental quotient (DQ), and the mean IQ or DQ score at each age of assessment. We preferentially recorded the Full-Scale IQ or Global Developmental Assessment scores rather than subscales.

For the within-study comparison, we preferentially reported results for associations between starting dose and Full-Scale IQ or Global DQ unless results were only reported for subscales. We preferentially used adjusted measures of association between the starting dose of levothyroxine and IQ or DQ obtained using continuous measures for starting dose, severity of hypothyroidism, and any other confounding variables. In the event of uncertainty about whether analyses represented crude or adjusted measures of association, we contacted the authors. We also recorded results from studies that examined the relationship between starting dose and any measure of growth, after adjusting for severity of hypothyroidism at diagnosis. We sought any reports that examined the relationship between starting dose and behavior because behavioral problems have been reported to be an adverse effect of high-dose levothyroxine. Because the evidence for an association between severity of hypothyroidism at diagnosis and later behavior is inconsistent,18,24 we included reports regardless of whether they took account of severity of hypothyroidism.

ANALYSIS

Between-Study Comparison

Our aim was to determine whether the mean IQ or DQ for the cohort was correlated with the mean starting dose of levothyroxine. We ranked studies according to the mean or median (when the mean was not available) starting dose. For the UK [United Kingdom] II study, only a range of total daily doses during the first 6 months was given but was reported to be the same as the starting dose. We used the mean starting dose of 7 µg/kg per day quoted in the UK I study. We plotted the mean IQ or DQ with 95% confidence intervals for each study at each age in order of starting dose to visualize any relationship between starting dose and outcome (Figure 1). Because of potential biases due to loss to follow-up at older ages, we also plotted the mean scores for the youngest age of assessment (Figure 2). Confidence intervals were derived from the SDs for the tests in the normative population and the sample size. The association between starting dose and mean IQ or DQ was investigated in a weighted linear regression analysis after allowing for between-cohort variation, using a random effects model. Studies were weighted by the inverse of the variance, and the model included terms for the proportion of children followed-up and age at assessment. Because the mean starting dose in the UK II study was not known for certain, we carried out sensitivity analyses, imputing doses between 0 and 20 µg/d. These were applied to the measurements at the youngest and oldest age of assessment separately. All analyses were performed using STATA software (Stata Corp, College Station, Tex).25

Within-Study Comparison

We did not attempt a quantitative analysis of the effect of starting dose on development due to the lack of consistency in the measures used for starting dose and other covariates and in the measures of association.

We aimed to systematically evaluate the evidence of an effect of levothyroxine starting dose on development, growth, and behavior. Because to our knowledge there have been no controlled trials comparing starting doses of levothyroxine treatment, we first compared mean population standardized developmental scores in cohorts with different mean starting doses of levothyroxine (between-study comparison). An underlying assumption in this ecologic analysis was that the spectrum of patients identified by screening would be similar, and population differences in mean developmental scores might therefore be attributable to differences in the start-
ing dose. In the second analysis, we sought studies that compared development, growth, or behavior in children treated with different starting doses of levothyroxine within a cohort (within-study comparison). We stipulated that studies included in the within-study comparison had to address the potential confounding effects of severity of hypothyroidism at diagnosis.23

RESULTS

One thousand one hundred sixty-six articles (excluding duplicates) were retrieved from the electronic databases and 106 from additional search methods, including hand-searches of textbooks, committee reports, and the forward citation search. We did not find any randomized controlled trials. Thirty-three potentially relevant studies were evaluated against the inclusion criteria (references available).

BETWEEN-STUDY COMPARISON

Characteristics of Studies Included

We identified 14 studies, involving 1321 patients, described in 31 articles11,12,18,20,21,26-51 for the between-study comparison. All of these studies enrolled children whose hypothyroidism was identified by neonatal screening (Table 1). Six more cohort studies identified by the searches failed to meet all of the inclusion criteria because mean development scores were not given13,52 or there was no information on the starting dose.23-37 All 14 studies excluded children with transient hypothyroidism. Most studies excluded children because of language difficulties and the largest studies (UK II, New England, Quebec I, and Toronto) excluded children with comorbid conditions likely to affect development (eg, Down syndrome). The age at final assessment ranged from 1 to 14 years. The number of children lost to follow-up or refusing to participate increased with age at assessment. Between 26% (Quebec I) and 88% (Norway) of the original cohort underwent the final developmental assessment.

Relationship Between Mean Starting Dose of Levothyroxine and IQ or DQ

We found no evidence for an effect of mean starting dose on mean IQ or DQ score at the youngest age of assessment (P=.48) after adjusting for age at assessment or percentage of children followed-up. In a sensitivity analysis, variation of the mean starting dose in the UK II study did not produce an effect that was significant at the 5% level. Similar results were observed at the oldest age assessed.

WITHIN-STUDY COMPARISON

Characteristics of Studies

Five studies20,38,42,51,58 met the inclusion criteria. Twelve studies* were excluded because the authors had not determined the effect of the starting dose of levothyroxine on development, growth, or behavior, or had not allowed for severity of hypothyroidism at diagnosis.

Development

Four studies examined the effect of the starting dose of levothyroxine on development, growth, or behavior, or had not allowed for severity of hypothyroidism at diagnosis.
results were inconsistent. The largest study42 (n=361) was based on a national UK register of children with congenital hypothyroidism. There was no evidence for an association (P=.6) between average levothyroxine dose during the first 6 months of life and Full-Scale IQ at age 5 years after adjusting for serum T4 level at diagnosis (measured as a continuous variable), socioeconomic status (categorized as manual or nonmanual), and age at the start of treatment (measured as a continuous variable).

In the Norwegian study38 (n=45), the initial dose of levothyroxine was associated with verbal IQ at age 6 years (accounting for 12% of the variance; P<.05). No association was detected between starting dose and the Mental Development Index (MDI) at age 2 years. The analysis adjusted for serum T4 level at diagnosis (measured as a continuous variable) and socioeconomic status (measurement not specified).

The Toronto study19 was based on 91 children and reported an association between starting dose of levothyroxine and Full-Scale IQ. Starting dose was divided into 2 groups using a cutoff of 7.7 µg/kg per day. The authors did not assess T4 levels at diagnosis but reported results stratified by bone age at diagnosis (above and below 36 weeks of gestational age). No justifications were given for the cutoffs used and no other confounding factors were included in the analysis.

### Table 1. Studies Reporting Cognitive or Global Development in Children With Congenital Hypothyroidism Used in the Between-Study Comparison

<table>
<thead>
<tr>
<th>Study Center, Years Patients Recruited</th>
<th>Levothyroxine Sodium Starting Dose, Mean (SD) or Range, µg/kg Per Day</th>
<th>No. of Patients</th>
<th>Age at Assessment, y Test</th>
<th>Mean (SD) Developmental Score‡</th>
</tr>
</thead>
<tbody>
<tr>
<td>New York,36 1987-1989</td>
<td>6.6 (1.0)</td>
<td>23</td>
<td>5 ± 1</td>
<td>WPPSI 101.4 (12.2)</td>
</tr>
<tr>
<td>Finland,37-38 1979-1988</td>
<td>6.8</td>
<td>50</td>
<td>2</td>
<td>Bayley 113 (22)</td>
</tr>
<tr>
<td>Italy (Naples),30 1982-1985</td>
<td>6.8 (1.8)</td>
<td>64</td>
<td>12</td>
<td>WISC-R 88.4 (14)</td>
</tr>
<tr>
<td>UK II,41 1982-1984</td>
<td>12-75§</td>
<td>472</td>
<td>5</td>
<td>WPPSI 109.3 (15.1)</td>
</tr>
<tr>
<td>UK I,39 1982-1984 before 1988</td>
<td>7.0 (1.6)</td>
<td>Unknown</td>
<td>1</td>
<td>Brunet-Lezine 97.6 (15.4)</td>
</tr>
<tr>
<td>Kansas,51 1980-1990</td>
<td>7.4 (2.6)</td>
<td>23</td>
<td>9.6 ± 2.1</td>
<td>WISC-R 103 (13.5)</td>
</tr>
<tr>
<td>Toronto,11,13,14,15 1987-1985</td>
<td>8.9 (1.3)§</td>
<td>106</td>
<td>2</td>
<td>Griffiths 107.5 (7.6)</td>
</tr>
<tr>
<td>Italy (Tuscany),36 before 1988</td>
<td>8.7 (2.8)</td>
<td>68</td>
<td>3-3.5</td>
<td>Griffiths 100.2 (4.8)</td>
</tr>
<tr>
<td>Norway,36 1979-1981</td>
<td>8.5 (3.3)</td>
<td>49</td>
<td>2</td>
<td>WPPSI 108.5 (2.3)</td>
</tr>
<tr>
<td>Sweden,36,37 1980-1981</td>
<td>8.7 (2.8)</td>
<td>68</td>
<td>7</td>
<td>Griffiths 100.2 (4.8)</td>
</tr>
<tr>
<td>Quebec I,42,43 1975-1985</td>
<td>8-10¶</td>
<td>105</td>
<td>1</td>
<td>WISC-R 103 (SEM, 2.3)</td>
</tr>
<tr>
<td>The Netherlands,30 1993-1996</td>
<td>12.1¶</td>
<td>59</td>
<td>1.5</td>
<td>Griffiths 108.5 (6)</td>
</tr>
</tbody>
</table>

*WISC-R indicates Wechsler Intelligence Scale for Children–Revised (Full-Scale IQ); WPPSI, Wechsler Preschool and Primary Scales of Intelligence (Full-Scale IQ); and UK, United Kingdom.
†Total includes patients with transient hypothyroidism, comorbidity or language difficulties, or those who moved away, refused consent, or were lost to follow-up. Number assessed excludes these patients.
‡Intelligence Quotient for WPPSI, WISC-R, Stanford-Binet; Developmental Quotient for Griffiths, General Cognitive Scale for McCarthy, Mental Developmental Index for Bayley.
§Analyses of mean starting dose of levothyroxine sodium used, 7 µg/kg per day.
¶Analyses of mean starting dose of levothyroxine sodium used, 9 µg/kg per day.
††Analyses of mean starting dose of levothyroxine sodium used, value shown.
Finally, the Dutch study^51^ involved 61 children and found no direct association between starting dose and MDI or Psychomotor Developmental Index (PDI) measured between age 10 months and 2.5 years after adjusting for serum thyroxine levels at diagnosis. However, the authors did report a significant effect of a combined variable (defined as the initial levothyroxine dose / log of the age at onset of treatment [measured as continuous variables]), MDI (P = .03), and PDI (P = .006). Dose and age explained only 7.5% and 3.4% of the variance, respectively.

**Growth**

Four cohort studies^10,32,58,60,61^ reported growth in children with congenital hypothyroidism but only 2^58,61^ examined the effect of starting dose on growth and only one^28^ took account of the severity of hypothyroidism at diagnosis. Heyerdahl et al^58^ compared serial growth measurements in children whose hypothyroidism was categorized as mild (T4 ≥ 3.1 ng/dL [40 nmol/L]) or severe (T4 < 3.1 ng/dL [40 nmol/L]). There was an inverse correlation (r = −0.40; P = .02) between the starting dose of levothyroxine and the age at onset of the childhood component of growth in children with severe hypothyroidism (n = 34) but not in those with mild hypothyroidism (n = 42). The age at onset of the childhood component of growth is derived from serial growth measurements up to age 3 years and represents the onset of the slow deceleration in linear growth that occurs between the ages of 6 and 12 months in normal children. This change probably reflects the start of the influence of growth hormone on linear growth.^62^

**Behavior**

Although behavior was assessed in several cohort studies,^11,18,19,21,22,39,63^ the effect of the starting dose of levothyroxine on behavior was only reported in the Toronto study. Clinician-assessed behavior indices (social withdrawal, internalizing problems, somatic complaints, social problems, thought problems, delinquency, and aggression on the Child Behavior Checklist and the Conner's Parent hyperactivity scale) measured in 88 children at age 8 years were increased in children who began receiving high-dose (>7.7 µg/kg per day) compared with low-dose (≤7.7 µg/kg per day) levothyroxine therapy (P < .05). A subsequent report of 83 children from the same cohort^29^ assessed by parents and teachers at mean

---

**Table 2. Studies Reporting the Association Between Starting Dose of Levothyroxine and Cognitive Development or Growth Analyzed in the Within-Study Comparison**

<table>
<thead>
<tr>
<th>Study Center, Years</th>
<th>Levothyroxine Sodium Starting Dose, µg/kg per Day</th>
<th>No. of Patients Studied</th>
<th>Age at Assessment, y</th>
<th>Confounding Variables Included in Analysis</th>
<th>Outcome Measure</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cognitive Development</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Norway,^37^ 1979-1981</td>
<td>Mean (SD), 8.5 (3.3)</td>
<td>45</td>
<td>2</td>
<td>Mean levothyroxine dose during age 1 and 2 years, T4, SES</td>
<td>Verbal IQ (WPPSI)</td>
<td>Starting dose explained 12% of the variance (P &lt; .05)</td>
</tr>
<tr>
<td>UK II,^41^ 1982-1984</td>
<td>Range, 12-75 µg/d</td>
<td>361</td>
<td>5</td>
<td>T4, average T4 level during 1st year, SES</td>
<td>Full-Scale IQ (WPPSI)</td>
<td>No association (P = .6)</td>
</tr>
<tr>
<td>Toronto,^19^ 1976-1985</td>
<td>≤7.7</td>
<td>45</td>
<td>7</td>
<td>Bone age at diagnosis (&gt; or ≤36 weeks)</td>
<td>Full-Scale IQ (WISC-R)</td>
<td>Mean IQ lower in group treated with ≤7.7 µg/kg per day than in group treated with &gt;7.7 µg/kg per day (P &lt; .001)</td>
</tr>
<tr>
<td>Netherlands,^50^ 1993-1996</td>
<td>Median, 9.2</td>
<td>61</td>
<td>1-2.5</td>
<td>T4 level at diagnosis, during first 3 and subsequent 9 months, SD of dose during latter period, SES</td>
<td>MDI and PDI (Bayley)</td>
<td>No association between starting dose and MDI or PDI†</td>
</tr>
</tbody>
</table>

| **Growth** |
| Norway and Sweden^57^ | 8.6 (3.4) | 76 | 0-6 | T4 level | No association if initial T4 =3.1 ng/dL (40 nmol/L)† If T4 <3.1 ng/dL (40 nmol/L), low dose correlated with delayed onset of childhood component of growth (r = −0.40, P = .02) |

^T4 indicates serum T4 level at diagnosis; SES, socioeconomic status; MDI, Mental Developmental Index; WPPSI, Wechsler Preschool and Primary Scales of Intelligence; UK, United Kingdom; and PDI, Psychomotor Developmental Index. Both indices were obtained from the Bayley Mental Scale.

†No further details were given.

©2002 American Medical Association. All rights reserved.
ages of 8.8 and 11.8 years, respectively, reported more internalizing (P < .05) and externalizing (P < .05) behavior problems in the high-dose group. The analyses did not adjust for the severity of hypothyroidism at diagnosis or other potential confounding factors.

Adverse Effects Other Than Behavior

Two studies each reported a single episode of thyrotoxicosis. Both involved children with mild forms of hypothyroidism receiving high doses of levothyroxine.

Although neonatal screening for hypothyroidism has been in operation for more than 20 years, we found no randomized controlled trials comparing the effects of different starting doses on clinical outcomes. We analyzed the available cohort data using 2 approaches and found no clear evidence for an effect of starting dose of levothyroxine on cognitive or global development, growth, or behavior.

The results of the between-study comparison should be interpreted with caution since an association between mean development scores and the starting dose of levothyroxine may have been masked by other center-specific characteristics. These include differential loss to follow-up, the spectrum of severity of hypothyroidism, the types of children refusing or excluded from participation, the performance of the IQ or DQ assessment, and secular trends in IQ score.

Within-study examination of the relationship between starting dose and outcome is less susceptible to bias but nonetheless may be affected by confounding. Analysis of starting dose as a categorical variable runs the risk of obscuring a real effect or, if not decided a priori, may produce spurious associations. Similarly, analysis of serum T4 level as a categorical variable reduces the power of the study to control adequately for confounding. Only 2 studies, UK II and the Netherlands, analyzed the starting dose of levothyroxine and serum T4 (a measure of the severity of hypothyroidism) as continuous variables. Neither study reported an association between starting dose and developmental outcome. The report by Bongers-Schokking et al of a weak association between a composite measure (dose divided by log age at the start of treatment) and developmental outcome up to 2.5 years, should be interpreted as a chance finding unless confirmed in independent data sets.

The relevance of the weak correlation, found in the Norwegian-Swedish study, between starting dose and the onset of childhood component of growth for growth in later childhood and final adult height is uncertain. We found weak evidence for increased behavior problems in children given a high starting dose of levothyroxine. This important, potentially adverse effect of a high starting dose needs to be confirmed in other studies that adjust for possible confounding by severity of hypothyroidism at diagnosis and social factors.

The limitations of the evidence found do not support clinical recommendations of high or standard starting doses of levothyroxine. Although higher-dose regimens lead to more rapid normalization of biochemical indices, we found no clear evidence for an effect on cognitive or global development. Our findings may be partly due to poor study methods and do not exclude a moderate but clinically important effect. More reliable information on the effect of high- vs standard-dose regimens on development, particularly in severely affected infants, might be provided by combining several large cohort studies in a meta-analysis of individual patient data. Alternatively or in addition, unbiased estimates of the effect of starting dose of levothyroxine on development, growth, and behavior in the long term could be determined by undertaking a multicenter randomized controlled trial.

Accepted for publication January 17, 2002.

This study was supported in part by the Department of Health Directorate of Research and Development, London, England.

We thank Leanne Jones, MSc, for help with searching, Tim Cole for help with statistical analysis, Tessa Parsons, PhD, for independent assessment of eligible studies, Peter Hindmarsh, and an anonymous reviewer for helpful comments on the manuscript.

Corresponding author and reprints: Ruth Gilbert, MD, MSc, FRCPCH, Department of Paediatric Epidemiology and Biostatistics, Institute of Child Health, 30, Guilford St, London, WC1N 1EH, England (e-mail: r.gilbert@ich.ucl.ac.uk).

REFERENCES
