Congenital hypothyroidism: no adverse effects of high dose thyroxine treatment on adult memory, attention, and behaviour

B Oerbeck, K Sundet, B F Kase, S Heyerdahl

Background: In congenital hypothyroidism (CH) it has been questioned whether high dose thyroxine replacement therapy has detrimental effects on memory, attention, and behaviour.

Aims: To describe memory, attention, and behaviour problems in young adults with CH, and to study possible negative effects of high dose thyroxine replacement therapy.

Methods: A cohort based follow up study of 49 young adults (mean age 20 years) with early treated CH, and sibling controls (n = 41).

Results: Controlled for age and sex, the CH group attained significantly lower scores than sibling controls on some tests of memory (Wechsler Logical Memory part II: 12.9 versus 17.8; difference 5.2, 95% CI 3.6 to 6.8) and attention (Wechsler Freedom From Distractibility factor: 95.6 versus 104.8; difference 9.9, 95% CI 6.4 to 13.4). They rated themselves with more behaviour problems than did sibling controls (52.7 versus 44.7; difference −7.6, 95% CI −11.2 to −4.0) on the Achenbach Self Report. A high thyroxine starting dose, high serum thyroxine treatment levels during the first six childhood years, and high levels at assessment had no adverse effects on outcome measures at age 20. On the contrary, the results suggest better outcome with higher childhood treatment levels.

Conclusions: Long term outcome revealed deficits in some aspects of memory, attention, and behaviour in young adults with CH relative to sibling controls. No adverse effects of high dose thyroxine therapy were found on measures of memory, attention, and behaviour problems.

A large number of studies on children and adolescents with early treated congenital hypothyroidism (CH) have reported results on intelligence (IQ), while there are fewer studies describing memory, attention, and behaviour. In the limited literature discussing these higher order cognitive skills, children and adolescents with CH are described as deficient in memory and attention compared to controls. More behavioural problems were found in terms of activity regulation and internalising symptoms, while one study reported no significant group differences. To our knowledge, adult outcome studies have not been reported.

Over the years, the level of thyroxine treatment has been debated, and higher thyroxine starting doses have been recommended (10–15 μg/kg/day). The lack of studies with a randomised controlled treatment design is a major obstacle in getting definite answers to the question of optimal thyroxine dosages. Hrytsiuk et al. concluded, in a review paper of clinical studies, that the evidence for an effect of thyroxine starting dose on IQ is weak. However, some studies have specifically studied children treated early with a high initial thyroxine dosage and found that even children with a severe CH had normal IQ. (For review see Heyerdahl and Oerbeck.)

On the other hand, studies from Toronto suggested that a high starting dose of thyroxine was associated with more behavioural problems, and that high serum levels of thyroxine at assessment were associated with more deficits in memory and attention.

In a previous article we reported enduring deficits in IQ, motor function, and school associated skills in young adults with CH compared to sibling controls. CH severity was associated with motor outcome, suggesting a prenatal effect, while early thyroxine treatment factors accounted for a significant portion of the variance in verbal IQ and school associated outcome. We found no negative effects of high thyroxine treatment on IQ, motor, and school associated outcome at age 20.

The objectives of the present paper are: (1) to investigate whether memory, attention, and behavioural function are affected in young adults with CH compared to sibling controls; and (2) to study whether high levels of thyroxine treatment have detrimental effects on memory, attention, and behaviour.

METHODS

Subjects

Forty nine children (29 girls) with CH were identified during the first three years of the Norwegian national screening programme (November 1978 to 1981). They were included in the present follow up study, which has been described previously, and have participated in a previous study at the age of 2 and 6 years. All subjects agreed to participate in the present follow up study at mean age 20.2 years (SD 0.9, range 18.3–21.7 years). Forty one siblings (16 girls) functioned as controls (mean age 21.4 years, SD 4.0, range

Abbreviations: CAVLT, Children’s Auditory Verbal Learning Test; CH, congenital hypothyroidism; CVMT, Continuous Visual Memory Test; DVT, Digit Vigilance Test; FFD, Wechsler Revised Freedom from Distractibility factor; RCFT, Rey Osterrieth Complex Figure; SES, socioeconomic status; T4, thyroxine; TMT, Trail Making Test; TSH, thyroid stimulating hormone; WCST, Wisconsin Card Sorting Test; WMS-R-LM, Wechsler Logical Memory Test, Story A
Biomedical diagnostic and early treatment data were CH variables scale based on the profession and education of head of Parental socioeconomic status (SES) was rated on a five point variables. Research Ethics Committee approved the study.20 Informed consent was obtained, and the Norwegian Medical Procedure were not treated early and continuously.20 Mean neurop- Norwegian well enough to be assessed. Four CH subjects disorder. One CH/sibling pair was of foreign origin, but spoke sibling was excluded due to a pervasive developmental of the CH subjects had central nervous system disorders. One known to influence cerebral development or function. None 12.3–30.0 years). Exclusion criteria were other disorders 100 mU/l. Disease and treatment characteristics

<table>
<thead>
<tr>
<th>Table 1 Disease and treatment characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td>CH variables</td>
</tr>
<tr>
<td>n</td>
</tr>
<tr>
<td>CH severity</td>
</tr>
<tr>
<td>Skeletal maturity at diagnosis</td>
</tr>
<tr>
<td>Type of CH</td>
</tr>
<tr>
<td>CH treatment, early years</td>
</tr>
<tr>
<td>Thyroxine starting dose; μg/kg/24 hours‡</td>
</tr>
<tr>
<td>Thyroxine dose age 1 year; μg/kg/24 hours</td>
</tr>
<tr>
<td>Thyroxine dose age 2 years; μg/kg/24 hours</td>
</tr>
<tr>
<td>Thyroxine dose age 6 years; μg/kg/24 hours</td>
</tr>
<tr>
<td>Mean serum T4 0–2.0 years; mmol/l</td>
</tr>
<tr>
<td>Mean serum T4 2.1–6 years; mmol/l</td>
</tr>
<tr>
<td>CH treatment of age 20</td>
</tr>
<tr>
<td>Serum TSH at testing; mU/l</td>
</tr>
</tbody>
</table>

*Knee epiphyses scores (0–4), score 0–1: absent or incipient epiphyseal development.
†Treatment was postponed for three children because of suspected transitory hypothyroidism, and one child had a drug withdrawal period. Mean age at start of treatment for the group with early and continuous treatment (n = 45): 18.5 ± 9.3 days.
‡ Mean thyroxine starting dose for the group with early and continuous treatment: 8.5 ± 3.3 μg/kg/24 hours.
†Mean serum T4 values were computed for each child for defined age periods, for the first year from samples drawn after 14 days of treatment.
*Drawn at mean 3.9 ± 7.7 days from day of neuropsychological testing.

12.3–30.0 years). Exclusion criteria were other disorders known to influence cerebral development or function. None of the CH subjects had central nervous system disorders. One sibling was excluded due to a pervasive developmental disorder. One CH/sibling pair was of foreign origin, but spoke Norwegian well enough to be assessed. Four CH subjects were not treated early and continuously.20, 21 Mean neuropsychological test results are presented for the total CH group (n = 49) and sibling controls. However, in the CH-sibling comparisons, results were not substantially different if these four children were excluded (data not shown). In analyses regarding the impact of CH variables on outcome, only the Norwegian CH subjects with early and continuous thyroxine treatment were included (n = 44).

Procedure

Informed consent was obtained, and the Norwegian Medical Research Ethics Committee approved the study.20

Variables

Background variables

Parental socioeconomic status (SES) was rated on a five point scale based on the profession and education of head of household.20

CH variables

Biomedical diagnostic and early treatment data were obtained from medical records (table 1). CH severity measures include serum thyroxine (T4) concentration at diagnosis, skeletal maturity at diagnosis (knee epiphyses score (range 0–4; score 0–1, absent or incipient knee epiphyses to 4, both femoral and tibial epiphyseal diameters >3 mm))20, 21 and scintigraphic classification of CH. All CH severity measures were included in the comparison between high and low starting dose effects on outcome, while serum T4 at diagnosis was used as the measure of CH severity in the multivariate analyses.24

Thyroxine treatment variables include thyroxine starting dose, mean serum T4 values calculated for each child from all serum T4 values during defined periods (first two years of life and from 2.1 to 6 years of age), and serum levels of thyroxine and TSH at age 20 years. The CH group was divided into two groups based on their thyroxine starting dose as done by Rovet and Ehrlich; low dose (<7.8 μg/kg/day) versus high dose (>7.8 μg/kg/day), to replicate their findings.27

The CH group was divided into four groups by their TSH levels at age 20 based on the TSH reference range of the laboratory: low TSH, <0.5 mU/l; normal TSH, 0.5–4.3 mU/l; increased TSH, 4.4–15 mU/l; and highly increased TSH, 16–100 mU/l.

Outcome variables

Memory was assessed in the visual domain with Delayed Recall from the Rey Osterrieth Complex Figure (RCFT)25 and the Immediate Total Score and Delayed Recognition Score from the Continuous Visual Memory Test (CVMT),25 and in the verbal domain with the Immediate and Delayed Recall Scores of story A from the WMS-R Logical Memory (WMS-R-LM)25 and Total Learning Score from the Children’s Auditory Verbal Learning Test (CAVLT).26

Attention was assessed within four sub-domains:

(1) Reaction time was measured as the difference between Choice Reaction Time minus Simple Reaction Time on the California Computerized Assessment Package (CalCAP)27
(2) Distractibility was measured with the Wechsler Revised Freedom From Distractibility factor (FFD)25 and the difference in time to complete Part A from Part B on the Trail Making Test (TMT)25
(3) Vigilance was measured with the Digit Vigilance Test (DVT)25
(4) Executive function was measured with the Stroop Test Interference Score29 and Perseverative Responses from the Wisconsin Card Sorting Test (WCST).25

Raw scores are presented for all tests except the Wechsler Revised Freedom From Distractibility factor, where standard scores were adapted from the age relevant American Wechsler norms.25

Behavioural problems were assessed with the age relevant ASEBA Self Report,30, 31 and T-scores of internalising, externalising, and total behaviour problems are reported. In addition, the number and percentage of CH subjects with mean results more than one standard deviation below the sibling group mean are presented.
Table 2  Memory, attention, and behaviour problems in young adults with CH and sibling controls*

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Tests and questionnaires</th>
<th>CH, n = 49 (Mean (SD))</th>
<th>Siblings, n = 41 (Mean (SD))</th>
<th>Estimated difference (95% CI)</th>
<th>p†</th>
<th>Number/% of CH subjects with results &lt; 1 SD below sibling mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Memory</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Visual</td>
<td>RCFT, delayed memory</td>
<td>22.6 (6)</td>
<td>24.2 (5)</td>
<td>1.9 (0.6 to 4.4)</td>
<td>0.14</td>
<td>n = 14/27%</td>
</tr>
<tr>
<td></td>
<td>CVMT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Immediate</td>
<td></td>
<td>75.2 (7)</td>
<td>82.9 (6)</td>
<td>7.7 (5.0 to 10.3)</td>
<td>0.003b</td>
<td>n = 27/55%</td>
</tr>
<tr>
<td>Delayed</td>
<td></td>
<td>3.9 (2)</td>
<td>6.0 (7)</td>
<td>2.6 (0.5 to 4.6)</td>
<td>0.014</td>
<td>n = 2/4%</td>
</tr>
<tr>
<td>Verbal</td>
<td></td>
<td>35.8 (6)</td>
<td>39.7 (6)</td>
<td>5.2 (2.3 to 8.1)</td>
<td>0.001b</td>
<td>n = 17/35%</td>
</tr>
<tr>
<td></td>
<td>WMS-R LM</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Immediate</td>
<td></td>
<td>14.6 (4)</td>
<td>18.5 (3)</td>
<td>4.2 (2.8 to 5.5)</td>
<td>0.003b</td>
<td>n = 28/57%</td>
</tr>
<tr>
<td>Delayed</td>
<td></td>
<td>12.9 (4)</td>
<td>17.8 (3)</td>
<td>5.2 (3.6 to 6.8)</td>
<td>0.003b</td>
<td>n = 34/69%</td>
</tr>
<tr>
<td>Attention</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reaction time</td>
<td>CalCAP, reaction time†</td>
<td>0.14 (0.06)</td>
<td>0.11 (0.05)</td>
<td>−0.022 (−0.044 to −0.036)</td>
<td>0.047</td>
<td>n = 16/33%</td>
</tr>
<tr>
<td>Distraction</td>
<td>Trail Making Test (B minus A)†</td>
<td>56.90 (54)</td>
<td>39.15 (21)</td>
<td>−25.1 (−40.9 to −9.4)</td>
<td>0.003b</td>
<td>n = 25/51%</td>
</tr>
<tr>
<td></td>
<td>Freedom From Distractibility</td>
<td>95.6 (8)</td>
<td>104.8 (9)</td>
<td>9.9 (6.4 to 13.4)</td>
<td>0.003b</td>
<td>n = 27/55%</td>
</tr>
<tr>
<td>Vigilance</td>
<td>Time to finish†</td>
<td>446.6 (87)</td>
<td>408.2 (75)</td>
<td>−36.4 (−72.0 to −8.2)</td>
<td>0.045</td>
<td>n = 24/49%</td>
</tr>
<tr>
<td></td>
<td>Errors†</td>
<td>4.1 (5)</td>
<td>2.0 (2)</td>
<td>−2.1 (−3.9 to −6.3)</td>
<td>0.007b</td>
<td>n = 15/22%</td>
</tr>
<tr>
<td>Executive</td>
<td>Stroop Interference score</td>
<td>3.12 (6)</td>
<td>2.74 (7)</td>
<td>−0.3 (−2.7 to 2.6)</td>
<td>0.73</td>
<td>n = 6/12%</td>
</tr>
<tr>
<td>function</td>
<td>WCST Perseverative responses†</td>
<td>12.5 (9)</td>
<td>8.8 (5)</td>
<td>−4.2 (−7.6 to −0.8)</td>
<td>0.016</td>
<td>n = 11/22%</td>
</tr>
<tr>
<td>Behavioural</td>
<td>ASEA internalising problems‡</td>
<td>50.3 (8)</td>
<td>43.4 (8)</td>
<td>−6.7 (−9.6 to −3.7)</td>
<td>0.000b</td>
<td>n = 25/51%</td>
</tr>
<tr>
<td>problems</td>
<td>ASEA externalising problems‡</td>
<td>50.9 (9)</td>
<td>45.5 (9)</td>
<td>−4.4 (−8.5 to −0.2)</td>
<td>0.041</td>
<td>n = 20/41%</td>
</tr>
<tr>
<td></td>
<td>ASEA total problems‡</td>
<td>52.7 (9)</td>
<td>44.7 (9)</td>
<td>−7.6 (−11.2 to −4.0)</td>
<td>0.000b</td>
<td>n = 22/45%</td>
</tr>
</tbody>
</table>

SD, standard deviation; CI, confidence interval.

*Group differences are analysed with a linear mixed model adjusted for age and sex.
†High scores on this test or questionnaire indicate more problems than low scores.
‡p values are the original ones from the mixed model analysis. B marks those that are significant after Bonferroni correction.

Statistical analyses
Development in CH versus controls
A linear mixed model was used to analyse group differences controlling for age and gender. To adjust for dependence between siblings, variation between sibling pairs was introduced as a random effect in the mixed model. Significance level was set to 0.05. Bonferroni correction was applied to adjust for multiple comparisons, thus setting the critical values to p = 0.008 for the memory measures, p = 0.007 for the attention measures, and p = 0.017 for the behavioural measures.

Adverse effects of thyroxine treatment levels
To replicate the Rovet and Ehrlich study, independent t tests were used to evaluate outcome differences between the groups of high (>7.8 μg/kg/day) and low (<7.8 μg/kg/day) thyroxine starting dose within the CH group. Multivariate analyses of variance (MANOVAs) were used to analyse the effect of thyroxine starting dose and TSH level at age 20 on outcome (within each psychological domain: memory, attention, and behaviour), with background (SES, gender), CH severity (serum T4 at diagnosis), and mean serum T4 levels during childhood years as covariates. Hierarchical linear multiple regression analyses were used to analyse the effect of thyroxine treatment variables during the first six childhood years on outcome, controlling for SES, gender, and CH severity, with forced entry of SES, gender, and CH severity and stepwise introduction of the thyroxine treatment variables. Missing mean substitution was used. R adjusted was used as an expression of explained variance.

RESULTS
Development in CH versus controls
Table 2 presents neuropsychological test results and behaviour problems.

Memory
The CH group showed significant memory deficits compared to controls. The memory deficits were evident on all verbal measures, while group differences were not consistent in the visual domain.

Attention
The CH group showed a significant deficit in distractibility, while group differences were not consistent on the vigilance measures. The groups did not differ significantly on measures of reaction time and executive function.

Behavioural problems
The CH group reported significantly more total and internalising behaviour problems, predominantly due to increased scores on the subscales Anxiety and Somatic Complaints (subscale data not shown).

Adverse effects of high dose thyroxine treatment levels
Early treatment level
There were no significant group differences in background factors, CH variables, or outcome at age 20 between the two groups with high (>7.8 μg/kg/day) and low (<7.8 μg/kg/day) thyroxine starting dose (table 3). MANOVA revealed no effects of starting dose on memory (F = 0.81, df = 6,37, p = 0.37), attention (F = 0.48, df = 7,29, p = 0.84), or behaviour (F = 1.14, df = 3,33, p = 0.35) at age 20 (Wilks Lambda).

Multiple regression analyses were performed on all dependent variables; table 4 presents analyses with significant results. There were no detectable adverse effects of higher levels of thyroxine treatment during early childhood years on memory, attention, and behaviour. On the contrary, the results suggest better results with higher treatment levels. The thyroxine starting dose and mean serum T4 level during
the first two years predicted the Freedom From Distractibility factor. Mean serum T4 level during the first two years predicted performance on the CalCAP reaction time task.

**Thyroxine treatment level (TSH) at age 20**

A large number of CH subjects had increased TSH in young adulthood (table 1). MANOVA revealed no effects of concurrent thyroxine treatment level (TSH) on memory (F = 1.56, df = 7,29, p = 0.19), attention (F = 1.28, df = 8,27, p = 0.30), and behaviour (F = 0.68, df = 3,33, p = 0.57) at age 20 (Wilks Lambda), with no adverse effects of either low or increased TSH.

**DISCUSSION**

**Development in CH versus controls**

**Memory**

The CH subjects performed significantly worse than controls on measures of memory, predominantly in the verbal domain. Story recall seem to be an especially sensitive measure, both in terms of mean results and the number/percentage of CH subjects scoring more than one standard deviation below sibling mean (table 2). Poor story recall has previously been reported by the Toronto group, but their statement that visuospatial deficits seemed to be the most striking feature in CH adolescents is inconsistent with the present findings within the visual domain and our former study on intellectual functioning in CH relative to sibling controls with no specific deficits in performance IQ relative to verbal IQ in the CH group.

**Attention**

A number of significant group differences were found on the attention measures (table 2). After Bonferroni correction, the CH group performed consistently and significantly weaker than sibling controls only on measures of distractibility, in line with Rovet and Hepworth. The present study found no consistent group difference in vigilance, in contrast to what was reported in a Dutch study of CH children. Use of different tests could be an explanation, as well as differences
in age (children versus adults). Executive functioning tests showed no consistent group differences in line with Rovet, who found that CH adolescents actually outperformed controls on the Wisconsin Card Sorting Test.7

**Behaviour**

Compared to controls CH subjects reported significantly more behavioural problems, primarily internalising symptoms as also found by Kooistra and colleagues. The present study thus reveals enduring behavioural problems in young adults with CH, inconsistent with results suggesting less behavioural problems with age.17

**Adverse effects of high dose thyroxine treatment levels?**

**Early treatment level**

The negative effects of a high thyroxine starting dose, reported in the Toronto studies by Rovet and Ehrlich16 were important, considering the recent recommendations of the use of a high thyroxine starting dose.4 10 11 In contrast to Rovet and Ehrlich,17 the present study found no adverse effects of a comparable high thyroxine starting dose on higher order cognitive skills. The mean results presented by Rovet and Ehrlich17 disclose, in general, better results in the high dose treatment group, and one could question whether the single negative finding (more behaviour problems in the high dose group) has been given more weight than warranted.

One could question our definition of high and low thyroxine starting dose (<7.8 μg/kg/day and >7.8 μg/kg/ day). This cut-off value was chosen simply to replicate the above mentioned study. Originally it was used by the Toronto study as a median split, and made two equally balanced groups in their material. This is also the case in our study, suggesting that the chosen cut-off value is a representative median thyroxine treatment level in studies from the 1980s.

In agreement with the findings comparing outcome in the two groups receiving either a high or low starting dose (table 3), regression analyses revealed no adverse effects of higher levels of thyroxine treatment during childhood years on higher order cognitive skills. On the contrary, on two tests (freedom from distractibility and cognitive reaction time), higher thyroxine treatment levels during the first two years significantly predicted better outcome.

**Thyroxine treatment level at age 20**

The present study found no adverse effects of high thyroxine treatment level at assessment. This is in contrast to earlier findings from the Toronto group,4 6 and also different from their latest findings indicating that both high and low levels of thyroid hormones at time of testing contribute to attention problems.33 34

**Strengths and limitations**

The strength of this study is the inclusion of a total three year cohort of CH subjects followed from infancy to young adulthood, and a sibling control group. We had systematic data on thyroxine treatment in childhood years and at assessment in young adulthood. We report findings from an observational study with great variations in treatment practice; however, treatment variables during different time periods were included in multivariate analyses. This study does not include a great number of CH subjects treated with very high thyroxine dosages (only 12 of the 49 CH subjects in the present study received a thyroxine starting dose >10 μg); thus definite answers to the outcome in high dose treatment groups await further studies.

**Conclusions**

The present study found significant group differences between young adults with CH and sibling controls on some measures of memory, attention, and behaviour problems. There were no adverse effects of higher thyroxine treatment levels during infancy, early childhood, or at assessment in young adulthood on higher order cognitive skills. The results support the guidelines advocating higher thyroxine treatment levels in CH.

**Table 4** Multiple regression analyses; effects of background variables, CH severity, and CH treatment variables on attention outcome

<table>
<thead>
<tr>
<th>Cognitive reaction time (CalCAP)</th>
<th>Freedom From Distractability</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>SES</strong></td>
<td><strong>β</strong></td>
</tr>
<tr>
<td>Gender (male = 1, female = 2)</td>
<td>0.01</td>
</tr>
<tr>
<td>Serum T4 at diagnosis</td>
<td>0.02</td>
</tr>
<tr>
<td>Thyroxine starting dose</td>
<td>0.00</td>
</tr>
<tr>
<td>Mean serum T4 0–2.0 years</td>
<td>0.03</td>
</tr>
<tr>
<td>Mean serum T4 2.1–6 years</td>
<td>0.00</td>
</tr>
<tr>
<td>R² adjusted</td>
<td>0.00</td>
</tr>
<tr>
<td>Significance of model</td>
<td>0.00</td>
</tr>
</tbody>
</table>

*Background (SES, gender) and CH severity (serum T4 at diagnosis) included by forced entry (block 1) and CH treatment variables (starting dose, mean serum T4 levels) included by stepwise entry (block 2) in the regression analyses. We report unstandardised coefficients (β) and standard error (SE); standardised coefficients (β), and significance (p). High scores on this test indicate more problems than low scores.

**Authors’ affiliations**

B Oerbeck, S Heyerdahl, Centre for Child and Adolescent Mental Health, Eastern and Southern Norway

K Sundet, Department of Psychology, University of Oslo, Norway

B F Kase, Department of Pediatric Research, Rikshospitalet University Hospital, Oslo, Norway

Grants: The Norwegian Research Council

Competing interests: none declared

**REFERENCES**

High dose thyroxine treatment