A Randomized Controlled Trial of the Effect of Thyroxine Replacement on Cognitive Function in Community-Living Elderly Subjects with Subclinical Hypothyroidism: The Birmingham Elderly Thyroid Study


College of Medical and Dental Sciences, University of Birmingham, Edgbaston, Birmingham B15 2TT, United Kingdom

Context: Subclinical hypothyroidism (SCH) and cognitive dysfunction are both common in the elderly and have been linked. It is important to determine whether T4 replacement therapy in SCH confers cognitive benefit.

Objective: Our objective was to determine whether administration of T4 replacement to achieve biochemical euthyroidism in subjects with SCH improves cognitive function.

Design and Setting: We conducted a double-blind placebo-controlled randomized controlled trial in the context of United Kingdom primary care.

Patients: Ninety-four subjects aged 65 yr and over (57 females, 37 males) with SCH were recruited from a population of 147 identified by screening.

Intervention: T4 or placebo was given at an initial dosage of one tablet of either placebo or 25 μg T4 per day for 12 months. Thyroid function tests were performed at 8-weekly intervals with dosage adjusted in one-tablet increments to achieve TSH within the reference range for subjects in treatment arm. Fifty-two subjects received T4 (31 females, 21 males; mean age 73.5 yr, range 65–94 yr); 42 subjects received placebo (26 females, 16 males; mean age 74.2 yr, 66–84 yr).

Main Outcome Measures: Mini-Mental State Examination, Middlesex Elderly Assessment of Mental State (covering orientation, learning, memory, numeracy, perception, attention, and language skills), and Trail-Making A and B were administered.

Results: Eighty-two percent and 84% in the T4 group achieved euthyroidism at 6- and 12-month intervals, respectively. Cognitive function scores at baseline and 6 and 12 months were as follows: Mini-Mental State Examination T4 group, 28.26, 28.9, and 28.28, and placebo group, 28.17, 27.82, and 28.25 [not significant (NS)]; Middlesex Elderly Assessment of Mental State T4 group, 11.72, 11.67, and 11.78, and placebo group, 11.21, 11.47, and 11.44 (NS); Trail-Making A T4 group, 45.72, 47.65, and 44.52, and placebo group, 50.29, 49.00, and 46.97 (NS); and Trail-Making B T4 group, 110.57, 106.61, and 96.67, and placebo group, 131.46, 119.13, and 108.38 (NS). Linear mixed-model analysis demonstrated no significant changes in any of the measures of cognitive function over time and no between-group difference in cognitive scores at 6 and 12 months.

Conclusions: This RCT provides no evidence for treating elderly subjects with SCH with T4 replacement therapy to improve cognitive function. (J Clin Endocrinol Metab 95: 3623–3632, 2010)
Subclinical hypothyroidism [defined as an elevated serum TSH concentration in the presence of a normal serum free T\textsubscript{4} concentration] is a common condition with prevalence ranging from 2.9% in a recent United Kingdom population-based study of subjects aged 65 yr and over (1) to 9% in a U.S. population of 18 yr and over (2). Whether subclinical hypothyroidism should be treated continues to be controversial (3–5).

Cognitive dysfunction is also common and is recognized as an important public health problem, with up to 16.8% of older adults having cognitive impairment of sufficient severity to affect completion of everyday tasks but that is not associated with a formal diagnosis of dementia (6). Hypothyroidism has traditionally been regarded as a cause of reversible cognitive impairment; indeed, recent functional magnetic resonance imaging evidence [not from a randomized controlled trial (RCT)] suggests reversible domain-specific cognitive defects in hypothyroidism (7). However, the extent of cognitive dysfunction associated with subclinical hypothyroidism is unclear (8). Also unclear is the sensitivity of the adult brain to changes in thyroid function and the potential reversibility of these changes with treatment (9). In addition, cognitive function has been reported as impaired compared with population norms in patients receiving T\textsubscript{4} replacement treatment for hypothyroidism (10). Recent evidence from a large (n = 5865) cross-sectional community-based study of subjects aged 65 yr and over measuring cognitive function has, however, shown no association between subclinical thyroid dysfunction and cognition or depression (11), although a smaller study of 1047 United Kingdom subjects aged 64 yr and over did show correlation between TSH values and cognitive performance (12). Detailed reviews of the association between cognitive dysfunction and thyroid dysfunction have come to no definitive conclusions (4, 13).

There have been few previous RCTs of the effects of T\textsubscript{4} therapy on cognitive function in subjects with subclinical hypothyroidism, and their results are conflicting [see Biondi and Cooper (4) and Table 1]. Nyström et al. (14) reported improvement in cognitive function in four of 17 women who completed a double-blind randomized crossover trial. Jaeschke et al. (15) in an RCT of nine men and 28 women with subclinical hypothyroidism reported an improvement in composite memory score with T\textsubscript{4} treatment. Samuels et al. (16) recently performed an RCT of subclinical hypothyroidism (induced by reducing the dose of T\textsubscript{4} replacement) in 19 females and reported significant reduction in working memory at the end of the subclinical hypothyroidism phase compared with that measured at the end of the euthyroid phase. Lastly, Jorde et al. (17) in the largest RCT reported to date (69 subjects, 37 men and 32 women) performed an extensive battery of cognitive function tests and found no differences with thyroid status, although after secondary analyses, they did report that serum TSH was negatively correlated with the Trail-Making test A, and serum free T\textsubscript{4} concentration was positively associated with the word association test (both at <1% significance level).

Improvement of impaired cognitive function in older adults could have an important impact on continued independence and quality of life as well as produce cost savings in terms of the maintenance of independent living. Thus, it is important to determine whether T\textsubscript{4} therapy in subjects with subclinical hypothyroidism confers any cognitive benefit. We therefore report on a double-blind RCT of the effects of T\textsubscript{4} replacement on cognitive function in elderly subjects (aged 65 yr and over) identified as having subclinical hypothyroidism in a community screening study. To our knowledge, this is the largest study to measure cognitive function in an RCT of T\textsubscript{4} replacement (registered as ISRCTN 23090699) to date.

Subjects and Methods

Design overview

This is a double-blind, placebo-controlled, RCT of 12 months titrated T\textsubscript{4} replacement to achieve euthyroidism in elderly community-living subjects of both sexes recruited from a cross-sectional study of prevalence of thyroid dysfunction.

Setting and participants

Patients with subclinical hypothyroidism were recruited from a community-based cross-sectional study describing the prevalence of thyroid dysfunction in the elderly and defining the association of cognitive function and depression with thyroid function (1, 10). Details of recruitment for the prevalence study are as previously published (1), but in brief, all subjects were aged 65 yr or over and were registered with one of 20 family practices in the greater Birmingham area of the United Kingdom. Subjects on T\textsubscript{4} therapy or antithyroid medications or those who had had recent treatment for hyperthyroidism were excluded from the prevalence study, as were those unable to provide informed consent. The overall prevalence of subclinical hypothyroidism (defined as a TSH concentration >5.5 mU/liter with free T\textsubscript{4} concentration from 9–20 pmol/liter) in this population was 2.9%. Subjects identified as having subclinical hypothyroidism over the 2-yr period of recruitment into the prevalence study were invited to enter the RCT.

Randomization and interventions

Eligible subjects were randomly allocated to receive 12 months of either T\textsubscript{4} treatment or placebo using a blocked randomization list (block size = 10) produced in advance and held by A.R. Eligible subjects were randomized, and prescriptions were prepared. Subjects were then invited to attend the trial clinic at which consent was obtained. The prescribed medication (T\textsubscript{4} or placebo) was then dispensed immediately at the clinic visit to
<table>
<thead>
<tr>
<th>Author</th>
<th>Protocol</th>
<th>Cognitive function tests</th>
<th>Number of subjects</th>
<th>Age (yr)</th>
<th>Sampling frame</th>
<th>Entry criteria and TSH values</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nystrom et al. (14)</td>
<td>Randomized to T₄ or placebo (50 µg for 2 wk, 100 µg for 2 wk, 150 µg for 2 wk), then crossed over</td>
<td>Identical forms test, Bingley's memory test, and reaction time to combined light and sound stimuli</td>
<td>20 females</td>
<td>51–73</td>
<td>Population, Sweden</td>
<td>Subclinical hypothyroidism. Mean TSH 7.7 (SD 3.7), range 2.9–16.3</td>
<td>4 of the 17 women improved on T₄</td>
</tr>
<tr>
<td>Jaeschke et al. (15)</td>
<td>Randomized to T₄ or placebo (25 µg for 1 month, 50 µg for 1 month, then dose titrated at monthly intervals against TSH to achieve biochemical euthyroidism); total length 40 wk</td>
<td>Composite psychomotor speed: subsuming word fluency, digit symbol substitution, and Trail Making; composite memory score: subsuming logical memory and word learning</td>
<td>37 (12 females, 6 males in active arm; 16 females, 3 males in placebo arm)</td>
<td>≥55 (mean 68 in both arms)</td>
<td>Not defined</td>
<td>Subclinical hypothyroidism; active arm (n = 18) mean TSH 12.1 (SD 6.8), range 6–32.4; placebo arm (n = 19) mean TSH 9.4 (SD 3.1), range 6–16.5</td>
<td>Composite psychomotor speed scores, NS; composite memory score improved in treatment group (mean change in z-score 0.58 (P = 0.01; 95% CI = 0.14–1.03)</td>
</tr>
<tr>
<td>Jorde et al. (17)</td>
<td>Randomized to T₄ or placebo in increasing doses (50 µg for 6 wk, 100 µg for 6 wk, then dose titrated at 3-monthly intervals according to TSH to achieve euthyroidism); total length of treatment 52 wk</td>
<td>Attention and working memory: digit span forward, digit span backward; seashore rhythm test; psychomotor/ cognitive speed: Trail Making Test A, Stroop test, parts 1 and 2, digit Symbol test; memory: verbal recall, visual recall, word list test; language: controlled word association test; cognitive flexibility/executive function: Stroop test part 3, Trail Making test B; speed of information processing: California Computerised Assessment Package; intelligence: vocabulary (Wechsler Adult Intelligence Scale); composite cognitive score</td>
<td>69 (17 females, 19 males in active arm; 15 females, 18 males in placebo arm)</td>
<td>29–75 (mean 61.6 in T₄ arm, 63 in placebo arm)</td>
<td>Population study, Tromso, Norway</td>
<td>Subclinical hypothyroidism; active arm (n = 36) mean TSH 5.81 (SD 1.76); placebo arm (n = 33): mean TSH 5.32 (SD 1.25)</td>
<td>No significant differences in any tests</td>
</tr>
</tbody>
</table>

(Continued)
avoid the inconvenience of re-attendance for the medication to be supplied. Subjects had thyroid function tests (TSH and free T₄) performed at 8-weekly intervals. After each blood test, the results were reviewed by a clinician blinded to the patient randomization. The number of tablets (25 µg of T₄ or placebo) per day was adjusted in one-tablet increments (the initial dosage being one tablet of placebo or 25 µg of T₄ per day) with the aim of achieving and then maintaining a serum TSH within the reference range (0.4–5.5 mU/liter) for subjects in the T₄ treatment arm. Subjects receiving placebo had the dose of placebo adjusted in the same way (i.e., the number of tablets was increased or reduced depending on thyroid function tests, but the patient received inactive tablets) to maintain double blinding. In this way, we aimed to achieve a similar experience of being prescribed single or multiple tablets in both the treatment and the placebo groups. All subjects received a new supply of T₄ or placebo along with new dose instructions after every test. The patients and all those who had direct contact with them, including those who performed cognitive function testing, were blinded to treatment allocation.

Outcomes and follow-up

All patients had cognitive function tests performed at entry to the trial and at 6 and 12 months after entry (aiming for a minimum of 6 months of biochemically euthyroid state for those who were receiving T₄ therapy). Any subjects receiving placebo and whose thyroid function tests on two consecutive occasions showed overt hypothyroidism (defined as TSH concentration above the upper limit of the reference range with a free T₄ concentration below the lower limit) were removed from the trial and referred to their family doctor for consideration of treatment with T₄.

Cognitive function tests performed

Trained research nurses administered cognitive function tests, using alternative forms for Middlesex Elderly Assessment of Mental State (MEAMS), Trail Making, and Speed and Capacity of Language Processing test (SCOLP) on retesting to reduce any learning effect. For all subjects, the following tests were completed: 1) Folstein Mini-Mental State Examination (MMSE) (18), which is widely used to determine cognitive status in clinical and research settings; 2) MEAMS (19), which was developed as a screening test to detect gross impairment of specific cognitive skills in elderly persons and systematically surveys the major areas of cognitive performance; 3) SCOLP (22), a measure of speed of processing, accounting for premorbid intelligence; and 4) Trail-Making Test (Parts A and B), a well-established psychomotor test of executive function. Slower times to complete the test indicate greater cognitive dysfunction; the difference score, calculated as the difference in times taken to complete the simpler Part A and more complex Part B (Part B – Part A), can be used to control for the effect of motor speed on performance to give a more accurate measure of executive function than the performance on either part alone (23).

TABLE 1. Continued

<table>
<thead>
<tr>
<th>Author</th>
<th>Protocol</th>
<th>Cognitive function tests</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Samuels et al.</td>
<td>Induced subclinical hypothyroidism by reducing dose of T₄ by average 45% for a total of 12 wk; randomized to intervention arm or placebo then crossed over at 12 wk; total length of induced subclinical hypothyroidism 12 wk</td>
<td>Declarative memory: paragraph recall and complex figure test; working memory: N-back test, subject-ordered pointing, and digit span backwards</td>
<td>Significant reduction in working memory when subclinical hypothyroidism was induced and maintained for 12 wk; working memory was notably worse in the cognitive impairment group.</td>
</tr>
<tr>
<td>Parle et al.</td>
<td><strong>Measurements</strong></td>
<td>Declarative memory: paragraph recall and complex figure test; working memory: N-back test, subject-ordered pointing, and digit span backwards</td>
<td>Significant reduction in working memory when subclinical hypothyroidism was induced and maintained for 12 wk; working memory was notably worse in the cognitive impairment group.</td>
</tr>
</tbody>
</table>
provides a score of 0–30 (20, 21). Subtests in MEAMS can be used alone or a combined score produced (range 0–12). In both tests, higher scores indicate less dysfunction. For SCOLP, a higher discrepancy score indicates greater cognitive dysfunction, with a score of 4 or more indicating a severe problem. The Hospital Anxiety and Depression Scale (HADS) (26) was also completed, because depression is potentially confounded with cognitive dysfunction (27); scores range from 0–21 with a score of 11 or above indicating caseness, that is, a diagnosis of depression.

Statistical analysis and power calculations

Of the 168 subjects identified with subclinical hypothyroidism in the cross-sectional study of thyroid function and cognition (11), 147 were eligible for recruitment into the RCT. Of these, 94 subjects (64% of those eligible) consented to enter the RCT (see Consort flow chart, Fig. 1). Power calculations performed before the prevalence study indicated that 400 patients randomized to T4 or placebo would be sufficient to detect a 0.75-U difference in MMSE (SD = 2.2) (28) and a 1-U difference in HADS (SD = 3.07) (29) with 90% power and 5% significance. We retained sufficient power (90%) to identify a 1.5-U difference in MMSE (SD = 2.0) and a 2-U difference in HADS (SD = 2.5).

Data were analyzed on an intention-to-treat basis, and outcome data were obtained for 82 of 94 subjects at 6 months and 85 of 94 subjects at 12 months. Baseline characteristics were compared across the groups, including age, gender, and deprivation. Deprivation was measured by the Index of Multiple Deprivation 2004 (30), a composite proxy measure of deprivation based on each patient’s place of residence (postcode/zip code). Cognitive function tests (MEAMS, MMSE, SCOLP, and Trail Making A and B) and depression score (HADS) were compared between the intervention and placebo groups at the 6- and 12-month follow-up using linear mixed modeling. Treatment alone and in combination with time were considered as fixed effects, with baseline measurement as a covariate, time (6 and 12 months) as a repeated factor, and participants as the random factor. Statistical analyses were performed using SPSS software, version 14.0 (SPSS Inc., Chicago, IL) and SAS version 9.1 (SAS Institute Inc., Cary, NC).

Biochemical assessment of thyroid function

Serum free T4 and TSH were measured by chemiluminescent immunoassay (Advia Centaur; Siemens Healthcare Diagnostics UK, Camberley, Surrey, UK). The laboratory reference range for free T4 was 9.0–20.0 pmol/liter with an interassay coefficient of variation of 8.2–9.8% over the range 8.2–54.9 pmol/liter. Serum TSH had a reference range of 0.4–5.5 mU/liter with an inter-assay coefficient of variation of 4.4–10.9% over the range 0.41–24.5 mU/liter. The lower limit of reporting for the TSH assay was 0.1 mU/liter with a mean functional sensitivity 0.02 mU/liter as quoted by the manufacturer. Serum free T4 and TSH concentrations were determined in all subjects on each occasion. The reference ranges employed for free T4 and TSH were those recommended by the manufacturer and used in other studies of this cohort and studies from our unit (1, 11, 31).

Results

Table 2 presents baseline characteristics by randomization arm and also for those patients declining to participate in the trial. A total of 94 subjects (57 females and 37 males) consented to recruitment into the trial. Differential consent rates between those randomized to treatment or placebo resulted in 52 subjects (mean age 73.5 yr, range 65–94 yr) receiving T4 replacement and 42 subjects (mean age 74.2 yr, range 66–84 yr) receiving placebo. No substantial baseline differences were observed between the randomized groups. No significant differences were found between those willing and those refusing to take part in the trial: mean difference in age 0.6 yr (95% confidence interval (CI) = −0.9–1.1), baseline TSH 0.28 (95% CI = −0.56–1.11) with a mean change at 6 months of 0.18 (95% CI = −0.64–1.0).

Statistical analysis and power calculations

Of the 168 subjects identified with subclinical hypothyroidism in the cross-sectional study of thyroid function and cognition (11), 147 were eligible for recruitment into the RCT. Of these, 94 subjects (64% of those eligible) consented to enter the RCT (see Consort flow chart, Fig. 1). Power calculations performed before the prevalence study indicated that 400 patients randomized to T4 or placebo would be sufficient to detect a 0.75-U difference in MMSE (SD = 2.2) (28) and a 1-U difference in HADS (SD = 3.07) (29) with 90% power and 5% significance. We retained sufficient power (90%) to identify a 1.5-U difference in MMSE (SD = 2.0) and a 2-U difference in HADS (SD = 2.5).

Data were analyzed on an intention-to-treat basis, and outcome data were obtained for 82 of 94 subjects at 6 months and 85 of 94 subjects at 12 months. Baseline characteristics were compared across the groups, including age, gender, and deprivation. Deprivation was measured by the Index of Multiple Deprivation 2004 (30), a composite proxy measure of deprivation based on each patient’s place of residence (postcode/zip code). Cognitive function tests (MEAMS, MMSE, SCOLP, and Trail Making A and B) and depression score (HADS) were compared between the intervention and placebo groups at the 6- and 12-month follow-up using linear mixed modeling. Treatment alone and in combination with time were considered as fixed effects, with baseline measurement as a covariate, time (6 and 12 months) as a repeated factor, and participants as the random factor. Statistical analyses were performed using SPSS software, version 14.0 (SPSS Inc., Chicago, IL) and SAS version 9.1 (SAS Institute Inc., Cary, NC).

Biochemical assessment of thyroid function

Serum free T4 and TSH were measured by chemiluminescent immunoassay (Advia Centaur; Siemens Healthcare Diagnostics UK, Camberley, Surrey, UK). The laboratory reference range for free T4 was 9.0–20.0 pmol/liter with an interassay coefficient of variation of 8.2–9.8% over the range 8.2–54.9 pmol/liter. Serum TSH had a reference range of 0.4–5.5 mU/liter with an inter-assay coefficient of variation of 4.4–10.9% over the range 0.41–24.5 mU/liter. The lower limit of reporting for the TSH assay was 0.1 mU/liter with a mean functional sensitivity 0.02 mU/liter as quoted by the manufacturer. Serum free T4 and TSH concentrations were determined in all subjects on each occasion. The reference ranges employed for free T4 and TSH were those recommended by the manufacturer and used in other studies of this cohort and studies from our unit (1, 11, 31).

Results

Table 2 presents baseline characteristics by randomization arm and also for those patients declining to participate in the trial. A total of 94 subjects (57 females and 37 males) consented to recruitment into the trial. Differential consent rates between those randomized to treatment or placebo resulted in 52 subjects (mean age 73.5 yr, range 65–94 yr) receiving T4 replacement and 42 subjects (mean age 74.2 yr, range 66–84 yr) receiving placebo. No substantial baseline differences were observed between the randomized groups. No significant differences were found between those willing and those refusing to take part in the trial: mean difference in age 0.6 yr (95% confidence interval (CI) = −0.9–1.1), baseline TSH 0.28 (95% CI = −0.56–1.11) with a mean change at 6 months of 0.18 (95% CI = −0.64–1.0).

Statistical analysis and power calculations

Of the 168 subjects identified with subclinical hypothyroidism in the cross-sectional study of thyroid function and cognition (11), 147 were eligible for recruitment into the RCT. Of these, 94 subjects (64% of those eligible) consented to enter the RCT (see Consort flow chart, Fig. 1). Power calculations performed before the prevalence study indicated that 400 patients randomized to T4 or placebo would be sufficient to detect a 0.75-U difference in MMSE (SD = 2.2) (28) and a 1-U difference in HADS (SD = 3.07) (29) with 90% power and 5% significance. We retained sufficient power (90%) to identify a 1.5-U difference in MMSE (SD = 2.0) and a 2-U difference in HADS (SD = 2.5).

Data were analyzed on an intention-to-treat basis, and outcome data were obtained for 82 of 94 subjects at 6 months and 85 of 94 subjects at 12 months. Baseline characteristics were compared across the groups, including age, gender, and deprivation. Deprivation was measured by the Index of Multiple Deprivation 2004 (30), a composite proxy measure of deprivation based on each patient’s place of residence (postcode/zip code). Cognitive function tests (MEAMS, MMSE, SCOLP, and Trail Making A and B) and depression score (HADS) were compared between the intervention and placebo groups at the 6- and 12-month follow-up using linear mixed modeling. Treatment alone and in combination with time were considered as fixed effects, with baseline measurement as a covariate, time (6 and 12 months) as a repeated factor, and participants as the random factor. Statistical analyses were performed using SPSS software, version 14.0 (SPSS Inc., Chicago, IL) and SAS version 9.1 (SAS Institute Inc., Cary, NC).
There were no significant changes in any of the measures of cognitive function over time. Inclusion of depression as a covariate did not alter these findings. Results remained the same when smoking status was included as a covariate in the analyses.

The differences in mean change between the groups are presented in Table 4. Restricting the analysis to those subjects in the T4 group who had achieved at least 6 months of euthyroidism by the time of the 12-month cognitive function test did not alter these findings [i.e., 37 of 45 patients (82.2%) with available data in the T4 group were euthyroid at 6 months; no significant differences were found between these T4 patients and the placebo group for any of the outcomes at 12 months]. Furthermore, multiple imputation of missing data demonstrated the robustness of these findings.

**Discussion**

In this double-blind RCT of T4 replacement therapy in subjects with subclinical hypothyroidism aged 65 yr and over, we have shown no significant improvement in tests of cognitive function in those receiving T4 replacement compared with those receiving placebo. The majority (82%) of subjects receiving T4 replacement were biochemically euthyroid by the time of the first cognitive function test at 6 months, and 84% were euthyroid by the time of their 12-month test. We have therefore shown no improvement in cognitive function both in the short term, i.e., when euthyroidism is achieved, and after a substantial period of time has elapsed while rendered euthyroid.

There have been few previous RCTs of T4 replacement in subclinical hypothyroidism and its effect on cognitive function. Previous studies have recruited primarily from specialist or unspecified populations (15, 16) and have been based on smaller samples (14–16). Some have been restricted to females only (14, 16), or the length of T4 treatment has been restricted to 12 wk (16). Our present study recruited an unselected sample from the community and included both men and women in a proportion similar to the community at large [39% of our subjects were male, compared with 54% males in Jorde’s study (17) and 24%...
### TABLE 3. Thyroid function over time

<table>
<thead>
<tr>
<th>Group</th>
<th>Baseline TSH [median, (IQR), range]</th>
<th>6 months TSH [median, (IQR), range]</th>
<th>12 months TSH [median, (IQR), range]</th>
<th>Proportion in euthyroid range (%)</th>
<th>Free T4 [median, (IQR), range]</th>
<th>Free T4 [median, (IQR), range]</th>
<th>Proportion in euthyroid range (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>T4</td>
<td>6.6 (6–8.5), 5.6–16.8</td>
<td>4.0 (2.7–4.6), 0.8–20.6</td>
<td>3.7 (2.8–4.9), 0.2–6.9</td>
<td>82.2</td>
<td>15.4 (14.9–17.4), 9.5–19.4</td>
<td>16.2 (14.2–17.3), 12.8–24.8</td>
<td>84.4</td>
</tr>
<tr>
<td>Placebo</td>
<td>6.65 (5.9–8.3), 9.4–16.8</td>
<td>6.4 (5.0–8.5), 1.2–19.0</td>
<td>5.45 (3.9–9.2), 0.9–17.3</td>
<td>34.5</td>
<td>12.5 (11.2–14.2), 9.6–21.1</td>
<td>12.85 (11.4–14.4), 9.7–22.2</td>
<td>50.0</td>
</tr>
</tbody>
</table>

Significant difference in TSH level between the placebo and T4 groups at both 6 and 12 months (Mann-Whitney U test z = 5.1, P < 0.0001; z = 3.8, P = 0.0002).

### TABLE 4. Depression and cognitive function scores over time

<table>
<thead>
<tr>
<th>Score</th>
<th>Group</th>
<th>Baseline, mean (SE)</th>
<th>n</th>
<th>Mean (SE)</th>
<th>Adjusted* mean (SE)</th>
<th>Group difference between adjusted means (95% CI)</th>
<th>n</th>
<th>Mean (SE)</th>
<th>Adjusted* mean (SE)</th>
<th>Group difference between adjusted means (95% CI)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>HADS</td>
<td>T4</td>
<td>3.38 (0.37)</td>
<td>49</td>
<td>3.92 (0.40)</td>
<td>3.78 (0.27)</td>
<td>0.28 (–0.56–1.11)</td>
<td>49</td>
<td>3.61 (0.36)</td>
<td>3.55 (0.27)</td>
<td>0.18 (–0.64–1.00)</td>
<td>0.82</td>
</tr>
<tr>
<td></td>
<td>Placebo</td>
<td>2.88 (0.45)</td>
<td>33</td>
<td>3.18 (0.43)</td>
<td>3.50 (0.32)</td>
<td></td>
<td>36</td>
<td>3.31 (0.47)</td>
<td>3.37 (0.31)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MEAMS</td>
<td>T4</td>
<td>11.72 (0.13)</td>
<td>46</td>
<td>11.67 (0.07)</td>
<td>11.56 (0.09)</td>
<td>–0.02 (–0.32–0.27)</td>
<td>46</td>
<td>11.78 (0.07)</td>
<td>11.67 (0.09)</td>
<td>0.07 (–0.21–0.36)</td>
<td>0.57</td>
</tr>
<tr>
<td></td>
<td>Placebo</td>
<td>11.21 (0.16)</td>
<td>32</td>
<td>11.47 (0.16)</td>
<td>11.58 (0.11)</td>
<td></td>
<td>36</td>
<td>11.44 (0.20)</td>
<td>11.60 (0.11)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MMSE</td>
<td>T4</td>
<td>28.26 (0.30)</td>
<td>48</td>
<td>28.90 (0.19)</td>
<td>29.00 (0.38)</td>
<td>1.19 (0.01–2.36)</td>
<td>46</td>
<td>28.28 (0.29)</td>
<td>28.24 (0.38)</td>
<td>0.03 (–1.12–1.17)</td>
<td>0.18</td>
</tr>
<tr>
<td></td>
<td>Placebo</td>
<td>28.17 (0.36)</td>
<td>33</td>
<td>27.82 (0.91)</td>
<td>27.82 (0.45)</td>
<td></td>
<td>36</td>
<td>28.25 (0.37)</td>
<td>28.22 (0.43)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SCOLP</td>
<td>T4</td>
<td>1.91 (0.46)</td>
<td>49</td>
<td>1.43 (0.38)</td>
<td>1.04 (0.30)</td>
<td>0.21 (–0.71–1.14)</td>
<td>49</td>
<td>1.69 (0.43)</td>
<td>1.29 (0.30)</td>
<td>–0.44 (0.46–1.36)</td>
<td>0.59</td>
</tr>
<tr>
<td></td>
<td>Placebo</td>
<td>0.71 (0.57)</td>
<td>33</td>
<td>0.39 (0.54)</td>
<td>0.82 (0.36)</td>
<td></td>
<td>36</td>
<td>0.22 (0.58)</td>
<td>0.84 (0.35)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trail Making A</td>
<td>T4</td>
<td>45.72 (2.32)</td>
<td>49</td>
<td>47.65 (3.28)</td>
<td>46.54 (2.62)</td>
<td>–3.33 (–11.39–4.73)</td>
<td>48</td>
<td>44.52 (2.62)</td>
<td>45.33 (2.63)</td>
<td>–1.44 (–9.42–6.53)</td>
<td>0.52</td>
</tr>
<tr>
<td></td>
<td>Placebo</td>
<td>50.29 (2.81)</td>
<td>33</td>
<td>49.00 (4.82)</td>
<td>49.87 (3.11)</td>
<td></td>
<td>36</td>
<td>46.97 (3.55)</td>
<td>46.78 (3.05)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trail Making B-Trail Making A</td>
<td>T4</td>
<td>110.57 (15.89)</td>
<td>49</td>
<td>106.61 (8.73)</td>
<td>104.24 (7.71)</td>
<td>–14.00 (–37.83–9.82)</td>
<td>48</td>
<td>96.67 (9.62)</td>
<td>100.65 (7.75)</td>
<td>–13.46 (–37.15–10.22)</td>
<td>0.95</td>
</tr>
<tr>
<td></td>
<td>Placebo</td>
<td>131.46 (19.72)</td>
<td>32</td>
<td>119.13 (18.56)</td>
<td>118.24 (9.21)</td>
<td></td>
<td>34</td>
<td>108.38 (14.12)</td>
<td>114.11 (9.07)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trail Making B</td>
<td>T4</td>
<td>65.76 (14.81)</td>
<td>49</td>
<td>58.96 (6.42)</td>
<td>57.97 (6.75)</td>
<td>–11.01 (–31.98–9.96)</td>
<td>48</td>
<td>52.15 (6.53)</td>
<td>54.55 (6.80)</td>
<td>–12.72 (–33.50–8.06)</td>
<td>0.86</td>
</tr>
<tr>
<td></td>
<td>Placebo</td>
<td>82.63 (18.38)</td>
<td>32</td>
<td>70.31 (14.21)</td>
<td>68.97 (8.14)</td>
<td></td>
<td>34</td>
<td>63.76 (11.72)</td>
<td>67.27 (7.97)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Adjusted by baseline score.

b Group by time interaction.
males in Jaeschke’s study (15)]. We provided T₄ replacement for a substantial period [52 wk, as did Jorde et al. (17); all the other RCTs (14–16) had shorter periods of T₄ therapy] with a large majority of those in our study treated with T₄ achieving euthyroidism. T₄ dosage was titrated (i.e. the dosage was modified according to repeated thyroid function tests) in our study and in those by Jaeschke et al. (15) and Jorde et al. (17) but not in those by Nystrom et al. (14) and Samuels et al. (16). This is also the largest RCT of T₄ replacement in subclinical hypothyroidism in the literature to date, being 36% bigger than the previous largest study (17), and our population was also significantly older than that reported on by Jorde et al. (17) (mean age in our study 73.8 yr with SD 5.8 yr; mean age in Jorde study 62.3 yr with SD 11.9 yr).

Three previous small RCTs (14–16) reported marginal differences in cognitive function in the T₄-treated group compared with controls, but our study of 94 subjects did not, our findings being consistent with Jorde’s study (17) of 69 subjects in which no differences were found in cognitive function after an extensive battery of cognitive function tests.

In a study of induced subclinical hypothyroidism secondary to a T₄ dose reduction, Samuels et al. (16) reported a reduction in working memory function during the time subjects were subclinically hypothyroid compared with when they were euthyroid, in contrast to our findings. This difference may reflect a chance finding or inducement of subclinical hypothyroidism in subjects already on replacement therapy for overt hypothyroidism in this study, as opposed to the correction of subclinical hypothyroidism in the other reported RCTs, including ours.

We defined hypothyroidism according to our laboratory’s reference ranges and aimed for the achievement of biochemical euthyroidism using the same ranges; as Biondi and Cooper (4) recently commented, “there is no consensus on the thyroid hormone and thyrotropin cut-off values at which treatment should be contemplated.”

Limitations of the study

Of the 168 potential participants, 53 declined to take part, but there were no significant differences in general characteristics nor in baseline measures of MMSE, MEAMS, or HADS between those who agreed to take part and those who refused. Another 15 individuals could not be entered to the trial because they were identified too late in the screening study for 12 months trial participation to have been supported. Only one of our subjects demonstrated cognitive dysfunction with a MEAMS score of 6 (mean MEAMS scores being 11.21 and 11.72 at baseline in the T₄ and placebo group, respectively; a score of 10 and upward is in the normal range) before the study, although even within the normal population reference ranges, there could have been improvement if T₄ replacement was effective in improving cognitive function as distinct from cognitive dysfunction. It is possible that those who declined to participate had a greater level of dysfunction than participants that was not assessed by baseline testing and that use of a more healthy group has reduced the potential to demonstrate improvement, although in the light of similar baseline scores in MMSE and MEAMS, any such difference would have to be subtle or highly localized. It is also possible that use of more subtle tests of higher executive function [see Burgess et al. (32)] may have revealed lesser differences in cognitive function than MMSE, MEAMS, SCOLP, or Trail Making, although the clinical significance of very minor deficiencies in higher cognitive functioning is unclear.

Patients eligible for entry into RCTs are typically randomized after consent is obtained. However, in this study, due to logistical (prescribing) issues, eligible patients’ names were randomized. This resulted in an imbalance between the group sizes due to different consent rates between those randomized to placebo and those randomized to treatment. There was also a lower than expected prevalence of subclinical hypothyroidism in the study from which the subjects for the RCT were recruited, so the number of subjects treated was less than anticipated, although we retained sufficient power (90%) to identify a 1.5-U difference in MMSE (SD = 2.0), a 0.6-U difference in MEAMS (SD = 0.8), and a 2-U difference in HADS (SD = 2.5). Stable patients retaking the MMSE on a second occasion improve their score by an average of 0.9 U (18, 33), so clinically important differences must be above this level. The minimally important clinical difference (MICD) in elderly patients with a transitory cognitive deficit (delirium) has been calculated as a negative change of 2 U or a positive change of 3 U (33). For a sample of elderly patients being tested with a longer time interval (1.5 yr), the MICD has been estimated to be between 2 and 4 U (21). One study has shown that the MICD on the HADS for patients with depression associated with chronic obstructive pulmonary disease is 1.68 U (34). No previous research has calculated the MICD for the MEAMS, but a 1-U difference would indicate failing one subtest, which would be clinically significant (19). Among those taking placebo, 50% normalized their TSH values at either 6 or 12 months; however, no significant differences in cognitive function tests within the placebo group between those who normalized their TSH and those who didn’t at either 6 or 12 months follow-up were found. The adopted prescribing strategy for the two groups was devised as logistically feasible and using only one algorithm to reduce human error. Because many patients in the T₄ group stabilized on low
doses of $T_4$, the number of changes of dosage was marginally greater in the placebo group (Mann-Whitney $U$ test; $z = 1.9; P = 0.06$). However, given that patients were aware that different dose tablets were being used, we consider that such differences are unlikely to have influenced subjects’ responses and are of limited relevance.

Recruitment into the study was based on a single thyroid function test; as is well known, thyroid function tests vary, and many return to normal from outside the reference range (35). There is therefore a possibility that ineligible euthyroid patients may have been recruited, increasing the chance of a type 2 error and thus reducing the power of the study to detect an effect on cognitive function.

Conclusion

Our previous cross-sectional survey showed no association between serum TSH concentration and cognitive function (11), an important finding that when combined with the presently reported lack of effect of $T_4$ treatment of subclinical hypothyroidism on cognitive function suggests that the case for treating subjects with subclinical hypothyroidism with replacement therapy to improve their cognitive state should not be accepted. Our results therefore militate against screening for thyroid dysfunction in community-based elderly subjects if the aim of such screening is to identify patients to treat to improve cognitive function. Finally, the potential adverse consequences of $T_4$ replacement need consideration: there is some evidence that untreated elevated TSH is associated with longevity (36, 37) (although the mechanism is unclear) and that subclinical hyperthyroidism (which can be due to $T_4$ therapy) may be associated with increased mortality (38). Thus, careful consideration is indicated before treating subclinical hypothyroidism, an essentially biochemical abnormality, by instituting $T_4$ hormone replacement therapy, when such treatment may also be harmful.

The research costs for the RCT were funded by the Health Foundation (http://www.health.org.uk/) and the service support costs from the Support for Science allocations to the Midlands Research Practices Consortium (MidReC). Both $T_4$ and identical placebo tablets were provided at no cost by Goldshield Pharmaceuticals (Croydon, Surrey, UK). The funding sources had no direct influence on the design and conduct of the study; collection, management, analysis, and interpretation of the data; or preparation, review, or approval of the manuscript. Trial support was provided by the Primary Care Clinical Research and Trials Unit. S.W. was funded by a Department of Health Career Scientist Award during the period of the study.

All subjects gave written informed consent for the study, which was approved by the multicenter research ethics committee of Scotland (MREC/01/0/24). This RCT is registered as ISRCTN 23090699.

Disclosure Summary: The authors declare there are no conflicts of interest in relation to this study.

References

5. Pham CB, Shaughnessy AF 2008 Should we treat subclinical hypothyroidism? BMJ 337:a834

Acknowledgments

We acknowledge E. Kidney of the School of Health and Population Sciences, College of Medical and Dental Sciences, University of Birmingham, Edgbaston, Birmingham B152TT, UK, who assisted with data collection, P. Clark, consultant clinical scientist at the University Hospital Birmingham National Health Service Foundation Trust, Edgbaston, Birmingham, UK, who led the laboratory analyses.

Address all correspondence and requests for reprints to: Professor J. Parle, School of Health and Population Sciences, College of Medical and Dental Sciences, University of Birmingham, Edgbaston, Birmingham B152TT, United Kingdom. E-mail: j.v.parle@bham.ac.uk.