



Published in final edited form as:  
*Endocr Pract.* 2005 ; 11(4): 223–233.

## Substitution of Liothyronine at a 1:5 Ratio for a Portion of Levothyroxine: Effect on Fatigue, Symptoms of Depression, and Working Memory Versus Treatment with Levothyroxine Alone

**Tom Rodriguez, DSN, RN, NP-C, APRN, BC [Assistant Professor of Nursing],**  
*College of Nursing, Texas Woman's University, Institute of Health Sciences at Houston and Adult Primary-Care Nurse Practitioner, Population Program, Baylor College of Medicine, Houston, Texas.*

**Victor R. Lavis, MD, FACP [Professor of Internal Medicine (Endocrinology), School of Medicine],**  
*University of Texas Health Science Center at Houston*

**Janet Ck. Meininger, PhD, RN, FAAN [Professor of Nursing, School of Nursing and Professor of Epidemiology],**  
*School of Public Health, University of Texas Health Science Center at Houston*

**Asha S. Kapadia, PhD [Professor of Biometry, School of Public Health and Professor of Preventive Dentistry], and**  
*Dental Branch, University of Texas Health Science Center at Houston*

**Linda F. Stafford, PhD, RN, CS [Assistant Professor of Nursing]**  
*School of Nursing, University of Texas Health Science Center at Houston*

### Abstract

**Background**—This study was planned at a time when important questions were being raised about the adequacy of using one hormone to treat hypothyroidism instead of two.

**Methods**—This trial attempted to replicate prior findings which suggested that substituting 12.5 µg of liothyronine for 50 µg of levothyroxine might improve mood, cognition, and physical symptoms in patients with primary hypothyroidism. Additionally, this trial aimed to extend the above findings to fatigue and to assess for differential effects in subjects with low and high fatigue at baseline. A randomized, double-blind, two-period, crossover design was used. Thirty subjects stabilized on levothyroxine (L-T4) at an endocrinology and diabetes clinic were recruited. Sequence one received their standard L-T4 dose in one capsule and placebo in another. Sequence two received their usual L-T4 dose minus 50 µg in one capsule and 10 µg of liothyronine (L-T3) in the other. At the end of the first six weeks, subjects were crossed over. T tests were used to assess carry-over and treatment effects.

**Results**—Twenty-seven subjects completed the trial. Mean L-T4 dose was 121 µg/d ( $\pm$  26.0) at baseline. There were no significant differences in fatigue and symptoms of depression between treatments. Measures of working memory were unchanged. While on substitution treatment, free thyroxine index was reduced by 0.7 ( $p < 0.001$ ), total serum thyroxine was reduced by 3.0 µg/dl ( $p < 0.001$ ), and total serum triiodothyronine was increased by 20.5 ng/dl ( $p = 0.004$ ).

**Conclusions**— With regard to the outcomes measured, substitution of L-T3 at a 1:5 ratio for a portion of baseline L-T4 was no better than treatment with the original dose of L-T4 alone.

## Keywords

randomized clinical trial; treatment; hypothyroidism; liothyronine; T3; levothyroxine; T4; fatigue; depression; working memory

In the United States, hypothyroidism affects as many as 9.5 percent of the population.<sup>1</sup> Modern medical theory directs that these patients be treated with levothyroxine (L-T4) alone – a synthetic hormone that replaces the inactive thyroid hormone thyroxine (T4). Yet there have been questions about the adequacy of treating hypothyroidism with only one hormone since at least the early 1970s.<sup>2</sup> Reasons for this are that the thyroid gland secretes triiodothyronine (T3) in addition to T4 and that some patients treated with L-T4 alone report they continue to experience symptoms of hypothyroidism. A large, controlled, community-based study recently supported the notion that there is an impairment in psychological well-being and increased symptoms of hypothyroidism in hypothyroid patients treated with L-T4 alone as compared to controls of similar age and gender (i.e., despite normal thyrotropin [TSH] levels).<sup>3</sup>

Most trials to date which compare the use of one hormone versus two in hypothyroid treatment are hindered by limitations which oblige further study. The earliest trial appeared in 1970.<sup>4</sup> At that time, sensitive measures of TSH were not yet widely available, and this resulted in significant hormone over-replacement. In addition to utilizing high doses of L-T4, the 1970 trial replaced liothyronine (L-T3) for L-T4 on a microgram per microgram basis without regard to the fact that L-T3 is three to eight times as potent.<sup>5</sup> The inevitable result was that subjects experienced significant side effects on the combined treatment. In 1999, a trial substituting 12.5 µg of L-T3 for 50 µg of daily L-T4 was reported.<sup>6</sup> This substitution ratio was more physiologic compared to that used in the earlier trial and resulted in improved mood, cognition, and physical symptoms. However, a commentary published simultaneously called for replication of findings before any change in practice could be entertained.<sup>7</sup>

It was in this context that the present study was planned in 2000. Since then, additional trials have been reported including one in infants born with congenital hypothyroidism (CH).<sup>8</sup> None of these trials was able to replicate the original findings however. In the study of infants with CH, subjects were replaced with a ratio of L-T3 for L-T4 of 1:5. Although not statistically significant, at three months the infants on substitution treatment had TSH values twice as high as those on L-T4 alone. Psychometric quotients (psychomotor evaluations) in the CH infants were significantly lower than in healthy matched controls, but there were no significant differences between CH infants treated with L-T4 alone versus substitution treatment.

In a second trial by Bunevicius, Jakubonien, Jurkevicius, Cernicat, Lasas, and Prange,<sup>9</sup> no significant differences were found on measures of mood, cognition, or physiologic variables between treatments. In yet another study – the largest to date,<sup>10</sup> outcomes were tissue hypothyroidism, symptoms of hypothyroidism, quality of life, depression and anxiety, cognitive functioning, and treatment satisfaction. In this study, subjects had significantly higher depression and anxiety scores on substitution treatment. This study also reported a mean TSH twice as high ( $p < 0.001$ ) on substitution treatment versus L-T4 alone (substitution ratio 1:5). Additionally, tissue hypothyroidism (i.e., Zulewski scores<sup>11</sup>) was significantly higher on the substitution intervention. Another study<sup>12</sup> focused on measures of depression and found no significant improvement in self-reported mood when subjects took substitution treatment. In this study however, the actual ratio of substitution varied because the protocol required that the pre-study dose of L-T4 be cut in half and replaced by 25 µg of L-T3. The L-T3 and L-T4 doses were then titrated upward to maintain TSH levels within normal limits. Additionally, the report of this small trial did not include a power analysis. In another trial,<sup>13</sup> neurocognitive functioning and hypothyroid health-related quality of life were the primary outcomes. This

trial reported no significant differences in neurocognition or quality of life between treatments. In this report however, validity and reliability of the hypothyroid health-related quality of life instrument was not reported, and the instrument had many depression items. Finally, the most recently reported trial<sup>14</sup> also failed to demonstrate any advantage to substitution treatment (ratio 1:5). This last trial also tested a ratio of 1:1.7. Limitations to the second regimen included no randomization to treatment and a supraphysiologic dose that resulted in over-replacement.

## METHODS

### Objectives and Outcome Measures

This study attempted to replicate the findings of the 1999 trial<sup>6</sup> and to extend those findings to fatigue as the primary outcome variable. Fatigue is a relevant endpoint because it is a common and early manifestation of hypothyroidism.<sup>15</sup> Secondary outcomes were symptoms of depression, certain symptoms of hypothyroidism, working memory, serum thyroid hormone profile including total T3 level, heart rate, blood pressure, and weight. Additional effort was put forth to identify a subgroup of patients in which substitution treatment might be effective by assessing whether subjects with high fatigue would be affected differentially than subjects with low fatigue at baseline (predefined subgroups).

Fatigue was measured using the Piper Fatigue Scale (PFS).<sup>16</sup> This is a 22 item, paper and pencil, self report that measures four dimensions of fatigue. The PFS was validated in an endocrine population during a small pilot study conducted by the primary investigator (TR). The same instrument was used in 1998 in a study of fatigue in women with breast cancer<sup>16</sup>; in that study, Cronbach alpha for the entire scale measured 0.97. Symptoms of depression were measured by two well-known, valid and reliable, paper and pencil, self-reports: the Beck Depression Inventory-II (BDI-II) and the General Health Questionnaire-30 (GHQ-30).<sup>17,18</sup> Working memory was measured by a well known, valid and reliable, researcher-administered instrument – the Working Memory subtest of the Wechsler Memory Scale-Third Edition (WMS-III).<sup>19</sup> Symptoms of hypothyroidism were measured using a well-known, valid and reliable technique for scaling subjective states known as the visual analogue scaling (VAS) model.<sup>20</sup>

Blood pressure (BP) and pulse were measured using a Critikon Dinamap Pro 2000 monitor and an appropriately-sized cuff. Subjects sat for two minutes prior to each measurement. Two measurements were made, and the first was discarded to guard against a potential white-coat effect. Weight was measured using a Detecto CN 20 scale. Shoes were removed prior to weighing. Blood samples were sent to an associated hospital laboratory for analysis.

### Recruitment and Selection of Participants

Patients aged 18 or older were invited to participate during routine visits to their endocrinologists at a diabetes and endocrine clinic associated with a large metropolitan hospital and medical school. Interested patients were required to have a diagnosis of primary hypothyroidism and to be stable (serum TSH within normal limits for at least three months) on L-T4 replacement alone. These patients were then screened for fatigue (using the PFS) and symptoms of depression or anxiety (using the GHQ-30).

To be eligible, patients had to have a minimum score of 1 on the PFS and no higher than a 45 on the GHQ-30 (scores > 45 on the GHQ-30 increase the probability of clinical depression) using Likert scoring at baseline. Sampling was carried out in such a way as to ensure that equal numbers from low (PFS = 1–5) and high fatigue (PFS = 6–10) strata were represented. Quota sampling was used to determine if effect sizes would vary depending on whether subjects had low fatigue versus high fatigue scores at baseline. To control for possible confounding, patients

with alcoholism, drug dependence, bipolar disorder, schizophrenia, dementia, brain injury, sleep disorders, chronic fatigue syndrome, and autoimmune diseases were excluded. Also excluded were patients with a history of erratic blood glucose levels. Finally, patients with unstable angina, myocardial infarction within the last year, or a history of ventricular dysrhythmias were also excluded.

### Research Design and Interventions

A randomized, double-blind, two-period, crossover design was selected because it offered the best chance of detecting significant differences. Reasons for this were 1) differences between treatments were expected to be small since subjects were being compared when they were receiving L-T4 or a near-equivalent amount of L-T4 plus L-T3 and 2) this design decreases within-group variability since each subject serves as his or her own control. Like independent sample experiments, this design also controls for threats to internal validity such as Hawthorne or placebo effects and effects due to time. A previous study<sup>6</sup> found no carry-over effects using this design. Finally, the half-life of L-T3<sup>21</sup> is short enough to allow it to be washed out five days after treatment ends.

Thirty subjects were recruited by the primary investigator (PI). The PI gave each subject a recruitment number which was based on the order in which the subject was recruited and a letter (L or H) which reflected the level of fatigue (low or high) that each subject reported at baseline. The pharmacy staff then utilized the statistical program Minitab Version-13 to randomly select recruitment numbers from each of the two fatigue strata. This process resulted in the random assignment of fourteen subjects to sequence one (half from the low and half from the high fatigue stratum). The remaining subjects (16) were then assigned to sequence two. The allocation sequence was concealed from the investigators, clinicians, and subjects by the pharmacy staff until all subjects completed the study. To accomplish double blinding, the pharmacy staff made up identically-looking L-T4 and L-T3 capsules specifically for this study. All psychometric measurements were taken at the University Clinical Research Center (UCRC) by the PI. Biometric measurements were made by the UCRC staff. To increase compliance with measurements and to decrease attrition, all measurement appointments were made at the subjects' convenience. Thus, blood samples were not drawn at the same time for each subject.

Prior to starting on study medications, subjects underwent baseline measurements. Study medications were then mailed to each subject by the pharmacy staff according to random assignment. Subjects randomized to sequence one received their normal dose of L-T4 (Levoxyl<sup>®</sup>, King Pharmaceuticals<sup>™</sup>) enclosed in one capsule plus *Lactobacillus acidophilus* in a placebo capsule. Since subjects in sequence one did not receive L-T3 (i.e., exogenous T3) until the second six weeks, a washout period for sequence one was not an issue.

Subjects in sequence two received their normal dose of L-T4 minus 50 µg in one capsule plus 10 µg of L-T3 (Cytomel<sup>®</sup>, King Pharmaceuticals<sup>™</sup>) in another capsule. For subjects in sequence two it was assumed that the L-T3 received exogenously during period one would wash out of the serum after five half-lives (i.e., about five days) into period two. Thus for sequence two, measurements at the end of the second period should reflect the fact that no exogenous L-T3 was in the serum during the preceding five-week period.

In substitution paradigms, doses are based on ratios of L-T3 to L-T4 commonly proposed to be equivalent. This is done so that subjects receive as close to the same amount of thyroid hormone on combined L-T3/LT4 treatment that they receive on L-T4 treatment alone. The only difference between treatments is hoped to be the manner in which thyroid activity is replaced (i.e., L-T3 given while on substitution treatment). In this trial 50 µg of L-T4 were substituted with 10 µg of L-T3 (as opposed to 12.5 µg of L-T3 used in the 1999 trial). This

substitution ratio (5:1) was used because the investigators wanted to avoid over-replacement with L-T3 as an alternative explanation for study results.

At the end of the first six-week period, subjects returned to the UCRC; and all measurements were repeated. At that time, subjects were crossed over by the pharmacy staff to receive the other treatment offered and given identical instructions. Subjects then returned to the UCRC at the end of the last six-week period at which time all measurements were repeated one final time. Regardless of sequence allocation, all subjects were asked to return to the UCRC after taking study medications for at least one week during each period in order to check how each subject was doing, to have each fill out a side-effect questionnaire, and to measure BP and pulse.

This protocol was approved by the Committee for the Protection of Human Subjects at the University of Texas Health Science Center at Houston. The study was reviewed and approved by the UCRC's Scientific Advisory Committee and by the Office of Clinical Research at the participating hospital. Subjects signed informed consents and authorizations for disclosure of protected health information.

### Sample Size and Power

Sample size was not predetermined. Instead the UCRC biostatistician performed a limited exploratory analysis while keeping the investigators blind. This was done after 11 subjects completed the study in order to provide an estimate of the variance of the difference between treatments on the fatigue variable. In light of the estimate provided, a sample size of 30 was selected. A power analysis (with alpha set at 0.05) was performed after the study was completed and revealed very high power ( $> 0.99$ ) to detect an effect as small as 0.5 on the PFS scale which measures fatigue on a scale of 0 to 10.

### Statistical Analysis

Tests for carry-over and treatment effects were performed (by the PI) on all primary and secondary variables using a t test for carry-over effects and a t test for treatment effects respectively as given in Rosner.<sup>22</sup> This same statistic was also used to test for treatment effects in the low and high fatigue subgroups. It is noteworthy that the t test for assessing treatment effects estimates the mean difference between treatments, and this estimate is not always equal to the actual difference between means. Minitab Version-13 was used to manage and analyze data. All tests were two-tailed because at the time the study was planned there was not enough evidence to clearly support any one-tailed result. Type I error was set at 0.05.

## RESULTS

### Participant Flow

The recruitment period was January 2002 to July 2003. A total of 39 patients were eligible for inclusion and were invited to participate. Of these, nine refused. Most refusers were female (8 or 89%). They were older than the study population – mean age 55.4 years ( $\pm 13.5$ ). Six (67%) were white, and three (33%) were black. Reasons given were none ( $n = 4$ ), not having enough time ( $n = 2$ ), feeling fearful of side effects ( $n = 2$ ), and living too far ( $n = 1$ ). Thirty subjects were enrolled; and out of these, 27 completed the trial. Two subjects were lost to follow-up. One left due to a recent diagnosis of bladder cancer and was on standard treatment plus placebo at the time. Another gave no reason and was also on standard treatment plus placebo. The last subject had interventions discontinued by the study endocrinologist due to feeling like fainting while gardening just a few days after starting treatment. This subject was on substitution therapy at the time. Data from all subjects including subjects who did not complete the trial were included in the analysis by intention to treat.

## Baseline Demographics and Clinical Characteristics

Table 1 shows the study sample characteristics (n = 30) and the characteristics of the subjects who completed the trial (n = 27) broken down by sequence. The majority of subjects enrolled were female (83% or 25). The mean age of the baseline sample was 47.5 years ( $\pm 12.9$ ). The mean dose of L-T4 prior to beginning the trial was 121  $\mu\text{g}/\text{d}$  ( $\pm 26$ ). Mean baseline serum TSH was 1.9  $\mu\text{U}/\text{ml}$  ( $\pm 1.7$ ).

## Treatment Effects

There were no carry-over effects ( $p = 0.34$ ). Table 2 shows treatment effects on fatigue and symptoms of depression. With regard to fatigue, there was no significant difference ( $p = 0.09$ ) between substitution treatment and standard L-T4 treatment ( $-0.9 \pm 0.57$ ) as measured by the PFS. This result occurred despite a power  $> 0.99$  to detect differences on this variable. Likewise, when fatigue was measured using a VAS model, there was no significant difference ( $p = 0.053$ ) between treatments ( $-12.0 \pm 6.0$ ). This trend toward lack of significance continued with respect to symptoms of depression. There were no significant differences between treatments regardless of whether symptoms of depression were measured using psychometric instruments (i.e., BDI-II and GHQ-30) or VAS models. Treatment effects were even less evident on measures of working memory. Here too, no result was statistically significant as the estimated differences were very close to zero. Table 3 shows these results. Overall, seven subjects preferred standard treatment. Eight felt no difference. Twelve subjects preferred substitution treatment.

Table 4 shows the effects of the treatments on biometric variables and symptoms of hypothyroidism. Mean serum TSH was not significantly different ( $p = 0.16$ ) between treatments although it was 2.7  $\mu\text{U}/\text{ml}$  higher during substitution treatment. Free thyroxine index and total T4 were lower (0.7 and 3.0  $\mu\text{g}/\text{dl}$  respectively) on substitution; and these changes were very significant ( $< 0.001$  in both cases). Conversely, total serum T3 was higher by about 20.5  $\text{ng}/\text{dl}$  while the subjects were on substitution therapy which was also significant ( $p = 0.004$ ). There were no significant differences between treatments with regard to symptoms of hypothyroidism (i.e., constipation, coldness, dry skin, and sleepiness).

## Subgroup Analysis and Adverse Events

Analysis of the fatigue variable (as measured by the PFS) in the prespecified subgroups of low and high fatigue at baseline was consistent with the results of the primary and secondary outcome analyses. In both, the low fatigue (n = 13) and the high fatigue (n = 14) strata, the estimated difference between treatments was not significant ( $p = 0.17$  and  $p = 0.18$  respectively) although standard errors were large relative to effect sizes. As expected, the effect size was larger in the high fatigue stratum ( $-1.26 \pm 1.0$ ) than that seen in the low fatigue stratum ( $-0.64 \pm 0.5$ ). Side effects reported on both treatments were comparable except for headaches, difficulty sleeping, and muscle weakness which were at least 10 percent greater on substitution treatment. Only one adverse event occurred (in an 82 year old subject on substitution treatment) during this trial where the subject reported feeling faint while gardening, and it was not serious.

## Overall Data Characteristics

In this trial, additional information can be gleaned by examining the overall data. Table 5 provides a comparison of means at baseline, the end of the first six weeks, and the end of the second six weeks by sequence for major variables. By comparing baseline means against the standard plus placebo means in each sequence, the degree to which placebo effects occurred can be estimated since subjects were on the same treatment at each point. One must keep in mind however that the effect of time is also a factor here.

On self-report instruments (i.e., PFS, symptom scales, BDI-II, and GHQ-30) there were placebo effects in both sequences as would be expected. However, there was an exaggerated placebo effect in sequence one as compared to sequence two; and therefore, it is likely that any positive effect of substitution treatment was obscured somewhat in this trial. Consequently, it is likely that the negative difference in efficacy (i.e., subjects feeling worse) due to substitution was due at least in part to an exaggerated placebo effect in sequence one. Therefore, examination of the overall data provides further support for the conclusion that there were no significant differences between treatments. Finally, a mean learning effect of 3.8 was noted for the Working Memory IQ at each subsequent testing period.

## DISCUSSION

This study attempted to replicate the findings of Bunevicius, Kazanavicius, Zalinkevicius, and Prange<sup>6</sup> while extending those findings to fatigue. The crossover design used in this study was similar to that used in the above trial except that we added an extra week of treatment to each period allowing L-T3 to wash out of the serum during the first week of period two for sequence two and the effect of L-T4 to predominate during the last five weeks of the same sequence. This appears to have been successful as a general carry-over or sequence effect was prevented (i.e., was not statistically significant). Moreover, to avoid over-replacement as an alternative explanation for results, we used a substitution ratio of 1:5 (versus 1:4 used in the 1999 trial). Additionally, we focused on fatigue as our primary endpoint because it is a common and early manifestation of hypothyroidism and one might expect a decrement in fatigue levels in hypothyroid patients given a treatment that is superior to the standard. Finally our design allowed for assessment of any differential effect on fatigue given the level of fatigue present at baseline. Results were expected to generalize to patients with primary hypothyroidism seen at endocrinology clinics.

With regard to fatigue, symptoms of depression, working memory, weight, heart rate, and BP, there were no statistically significant differences between treatments. Further, in the low and high fatigue subgroups, there also were no significant differences in fatigue levels between treatments. Thus, we were unable to replicate the findings of the trial reported in 1999.

However because the sample size in our study was determined around the primary fatigue outcome (as measured by the PFS) and because there were large treatment variances relative to means on the GHQ-30 and on most symptoms measured using VAS models, it is possible that more significant findings would have been detected if our sample had been larger. On the other hand, it should be emphasized that there was ample power to detect significant differences in the primary fatigue variable.

Other explanations for the lack of replication are as follows. One explanation is that the original trial utilized higher daily replacement doses of L-T4 (i.e., 175 µg/d versus 121 µg/d in our study). This was done because most subjects in the 1999 trial had thyroid cancer and were being treated with enough L-T4 to suppress TSH. Thus, the subjects in that trial received greater thyroid hormone activity overall (i.e., suppressive versus replacement therapy) than subjects in our study. Additionally, the original trial used a higher substitution dose of L-T3 (1:4 ratio) than we did (1:5).

Not only did subjects in the 1999 trial receive greater, daily thyroid hormone doses, but they were also different in that a little more than half were athyreotic secondary to treatment for thyroid cancer. In contrast, the majority of subjects in our study had an autoimmune etiology.

Also, subjects in sequence one of our study showed an exaggerated placebo effect that might have obscured any positive effect of substitution treatment. Another potential explanation for the lack of replication is that in our study all measurements (including blood draws) were taken

at the subjects' convenience; and consequently, circadian rhythmicities in the outcomes and differing L-T3 peak times might have contributed to within treatment variances and a lack of significant findings. Additionally, daily dosing of L-T3 did not allow steady-state serum levels of L-T3 to be attained given its half-life of about one day.<sup>21</sup> Another consideration is that while the half-life of L-T3 is short, the biological effects are longer and might require a longer period of time to completely wash out. On the other hand, one must keep in mind that a t test for carry-over effect was not significant ( $p = 0.34$ ).

Finally, in our study subjects were on different brands of L-T4 at commencement and switched to Levoxyl<sup>®</sup> for this trial. This brings up the question<sup>23</sup> of bioequivalence between generic and brand L-T4 products. Our decision to switch subjects to only one product was based on a fairly recent report<sup>24</sup> suggesting bioequivalence of the L-T4 products in question according to Food and Drug Administration criteria.

The PI performed a post hoc analysis to explore the potential for under-treatment (i.e., TSH > 5.00) and over-treatment (TSH < 0.32) during this trial. At no point in the study did any subject have a suppressed TSH (i.e., < 0.01). At baseline, two subjects had TSH values just over 5.00. When subjects were changed from their own brand of L-T4 to Levoxyl,<sup>®</sup> three additional subjects developed TSH values just over 5.00. However, the same number of subjects (four) with TSH < 0.32 at baseline remained below 0.32 on Levoxyl.<sup>®</sup> Thus overall, a change in L-T4 brands did not appear to have a significant impact.

In contrast, there was an additional increase in the number of subjects (three) whose TSH changed from normal to > 5.0 or from < 0.32 to the normal range (one) when they were changed from Levoxyl<sup>®</sup> to substitution treatment. This suggests that perhaps thyroid activity was somewhat less in the substitution arm versus the L-T4 arm of our study. Finally, the commonly accepted potency of L-T3 being three to eight times that of L-T4<sup>5</sup> might be inaccurate, it is certainly not exact. In fact, until the exact potency is known, there is likely to be continued questions about whether substitution regimens are really equivalent.

The apparent lack of success in using L-T3 to improve hypothyroid treatment to date does not entirely rule out the possibility that L-T3 may have a positive effect in hypothyroidism. There are several reasons for this (some alluded to already). First, questions about the relative potency of L-T3 to L-T4 remain. Second, there are issues involving more exact replication of the molar T3 to T4 ratio secreted by the thyroid and the lack of a slow-release L-T3 formulation discussed in a recent commentary.<sup>25</sup> Third, there may be one or more subgroups yet to be identified that do respond favorably to substitution treatment – for example, athyreotic subjects in the 1999 trial. Fourth, the substitution trials to date (including this one) are not without limitations.

Indeed, when we take into consideration that in most studies to date TSH levels increased during substitution treatment (significantly in one study), the question of whether subjects are being made hypothyroid to some degree on substitution regimens should be entertained. This speculation should be addressed in future research. However, an additive paradigm instead of a substitution paradigm aiming at equivalence of treatments may prove more useful. By “additive paradigm” we mean *adding physiologic amounts of L-T3 to doses of L-T4 that are equal (or very near equal) to the patient's baseline dose while ensuring that thyrotropin is not suppressed*. An additive paradigm would allow for getting away from the thorny issue of determining equivalent doses of active and inactive hormones and offers the advantage of avoiding making subjects hypothyroid – either by subtracting too much L-T4 during combined L-T3/L-T4 treatment or by some other, yet unidentified, mechanism.

In conclusion, with regard to the outcomes measured in this study, we did not find evidence to support the hypothesis that substitution of L-T3 at a 1:5 ratio for a portion of daily L-T4 produces better outcomes in a primary (mostly autoimmune) hypothyroid population than



treatment with the original amount of L-T4 alone; and therefore, we cannot recommend this treatment for the average patient. Additionally, this conclusion is consistent with the majority of evidence in the literature. This study adds to the knowledge base by extending outcomes to fatigue – a common and early symptom of hypothyroidism, by utilizing a valid and reliable measure of fatigue and other psychometric variables, and by attempting to identify a subgroup for which substitution therapy works.

### Acknowledgements

This study was supported in part by the National Institutes of Health, National Center for Research Resources, General Clinical Research Center Grant M01 RR002558.

Dr. Rodriguez's clinical research training during this study was supported by the American Nurses Association, Ethnic Minority Fellowship Program, funded by Substance Abuse and Mental Health Services Administration Grant 2T06 5M151555.

This study was conducted by Dr. Rodriguez in partial fulfillment of requirements for the Doctor of Science in Nursing degree at the University of Texas Health Science Center at Houston.

Levoxy<sup>®</sup> (levothyroxine sodium) and Cytomel<sup>®</sup> (liothyronine sodium) were donated by King Pharmaceuticals<sup>™</sup>, Inc., 501 Fifth Street, Bristol, Tennessee 37620.

### References

1. Canaris GJ, Manowitz NR, Mayor G, Ridgway EC. The Colorado Thyroid Disease Prevalence Study. *Arch Intern Med* 2000;160(4):526–34. [PubMed: 10695693]
2. Taylor S, Kapur M, Adie R. Combined thyroxine and triiodothyronine for thyroid replacement therapy. *BMJ* 1970;2:270–1. [PubMed: 5420176]
3. Saravanan P, Chau WF, Roberts N, Vedhara K, Greenwood R, Dayan CM. Psychological well-being in patients on 'adequate' doses of L-thyroxine: results of a large, controlled community-based questionnaire study. *Clin Endocrinol (Oxf)* 2002;57:577–85. [PubMed: 12390330]
4. Smith RN, Taylor SA, Massey JC. Controlled clinical trial of combined triiodothyronine and thyroxine in the treatment of hypothyroidism. *BMJ* 1970;4:145–8. [PubMed: 4097650]
5. Greenspan, FS. The thyroid gland. In: Greenspan FS, Strewler GJ. eds. *Basic and clinical endocrinology*. 5<sup>th</sup> ed. Stamford, CT: Appleton & Lange, 1997:204.
6. Bunevicius R, Kazanavicius G, Zalinkevicius R, Prange A. Effects of thyroxine as compared with thyroxine plus triiodothyronine in patients with hypothyroidism. *N Engl J Med* 1999;340:424–9. [PubMed: 9971866]
7. Toft AD. Thyroid hormone replacement – one hormone or two? *N Engl J Med* 1999;340:469–70. [PubMed: 9971874]
8. Cassio A, Cacciari E, Cicognani A, et al. Treatment for congenital hypothyroidism: thyroxine alone or thyroxine plus triiodothyronine? *Pediatrics* 2003;111(5 Pt 1):1055–60. [PubMed: 12728088]
9. Bunevicius R, Jakubonien N, Jurkevicius R, Cernicat J, Lasas L, Prange AJ. Thyroxine vs thyroxine plus triiodothyronine in treatment of hypothyroidism after thyroidectomy for Graves' disease. *Endocrine* 2002;18(2):129–33. [PubMed: 12374459]
10. Walsh JP, Shiels L, Lim EM, et al. Combined thyroxine/liothyronine treatment does not improve well-being, quality of life, or cognitive function compared to thyroxine alone: a randomized controlled trial in patients with primary hypothyroidism. *J Clin Endocrinol Metab* 2003 Oct;88(10):4543–50. [PubMed: 14557419]
11. Zulewski H, Muller B, Exer P, Miserez AR, Staub JJ. Estimation of tissue hypothyroidism by a new clinical score: evaluation of patients with various grades of hypothyroidism and controls. *J Clin Endocrinol Metab* 1997;82:771–776. [PubMed: 9062480]
12. Sawka AM, Gerstein HC, Marriott MJ, MacQueen GM, Joffe RT. Does a combination regimen of thyroxine (T4) and 3,5,3'-triiodothyronine improve depressive symptoms better than T4 alone in patients with hypothyroidism? results of a double-blind, randomized, controlled trial. *J Clin Endocrinol Metab* 2003 Oct;88(10):4551–5. [PubMed: 14557420]

13. Clyde PW, Harari AE, Getka EJ, Shakir KM. Combined levothyroxine plus liothyronine compared with levothyroxine alone in primary hypothyroidism: a randomized controlled trial. *JAMA* 2003 Dec 10;290(22):2952–8. [PubMed: 14665656]
14. Escobar-Morreale HF, Botella-Carretero JI, Gomez-Bueno M, Galan JM, Barrios V, Sancho J. Thyroid hormone replacement therapy in primary hypothyroidism: a randomized trial comparing L-thyroxine plus liothyronine with L-thyroxine alone. *Ann Intern Med* 2005 Mar 15;142(6):412–24. [PubMed: 15767619]
15. Wartofsky L. Diseases of the thyroid. In: Fauci AS, Braunwald E, Isselbacher KJ, et al. eds. *Harrison's principles of internal medicine*. 14<sup>th</sup> ed. New York, NY: McGraw-Hill, 1998:2012–35.
16. Piper BF, Dibble SL, Dodd MJ, Weiss MC, Slaughter RE, Paul SM. The Revised Piper Fatigue Scale: psychometric evaluation in women with breast cancer. *Oncol Nurs Forum* 1998;25(4):677–84. [PubMed: 9599351]
17. Beck AT, Steer RA, Brown GK. *Beck Depression Inventory – Second Edition manual*. San Antonio, TX: The Psychological Corporation, 1996.
18. Goldberg D, Williams P. *User's guide to the General Health Questionnaire*. Berkshire, Great Britain: NFER-NELSON, 1988.
19. The Psychological Corporation. *Wechsler Adult Intelligence Scale-Third Edition Wechsler Memory Scale-Third Edition: technical manual*. San Antonio: Author, 1997.
20. Walz, CF, Strickland, OL, Lenz, ER. *Measurement in Nursing Research*. 2<sup>nd</sup> ed. Philadelphia, PA: FA Davis Company, 1991.
21. Nicoloff JT, Low JC, Dussault JH, Fisher DA. Simultaneous measurement of thyroxine and triiodothyronine peripheral turnover kinetics in man. *J Clin Invest* 1972;51:473–83. [PubMed: 4110897]
22. Rosner B. *Fundamentals of biostatistics*. 4<sup>th</sup> ed. Belmont, CA: Duxbury Press, 1995:329–35.
23. Copeland PM. Two cases of therapeutic failure associated with levothyroxine brand interchange. *Ann Pharmacother* 1995;29:482–5. [PubMed: 7655130]
24. Dong BJ, Hauck WW, Gambertoglio JG, et al. Bioequivalence of generic and brand-name levothyroxine products in the treatment of hypothyroidism. *JAMA* 1997 April 16;277(15):1205–1213. [PubMed: 9103344]
25. Cooper DS. Combined T4 and T3 therapy – back to the drawing board. *JAMA* 2003 Dec 10;290(22):3002–4. [PubMed: 14665664]

**Table 1**  
Sample and Sequence Demographics and Means ( $\pm$  SD)<sup>\*</sup> at Baseline

	Sample (N = 30)	Sequence 1 (n = 12)	Sequence 2 (n = 15)
Gender			
Female	25 (83%)	11 (92%)	13 (87%)
Male	5 (17%)	1 (8%)	2 (13%)
Age (years)	47.5 (12.9)	48.1 (9.5)	43.7 (12.6)
Ethnicity			
White	23 (73%)	9 (75%)	11 (74%)
Hispanic	4 (17%)	2 (17%)	2 (13%)
Black	3 (10%)	1 (8%)	2 (13%)
Etiology			
Autoimmune	23 (77%)	10 (84%)	12 (80%)
Treatment with I <sup>131</sup>	4 (13%)	1 (8%)	1 (7%)
Thyroidectomy	3 (10%)	1 (8%)	2 (13%)
Baseline levothyroxine dose ( $\mu$ g/d)	121.0 (26.0)	118.0 (29.0)	121.0 (26.0)
Serum thyrotropin ( $\mu$ U/ml) <sup>†</sup>	1.9 (1.7)	1.7 (1.2)	1.8 (2.0)
Free thyroxine index <sup>†</sup>	3.2 (0.1)	3.2 (0.5)	3.3 (0.3)
Total serum thyroxine <sup>†</sup> ( $\mu$ g/dl)	10.9 (2.0)	11.0 (2.5)	11.2 (1.2)
Total serum triiodothyronine <sup>†</sup> (ng/dl)	79.0 (18.0)	76.2 (21.2)	79.4 (16.0)
Piper Fatigue Scale Total Score	4.9 (1.9)	4.4 (2.0)	5.0 (2.0)
General Health Questionnaire-30	30.2 (9.0)	30.1 (7.3)	32.3 (9.9)

\* Standard deviations.

<sup>†</sup> Normal reference ranges: thyrotropin (0.32 – 5.00), free thyroxine index (1.1 – 4.4), total serum thyroxine (4.5 – 12.5), and total triiodothyronine (50 – 148).

**Table 2**  
Effect of Treatments on Fatigue and Symptoms of Depression

Outcome Measure	Treatment			t <sup>†</sup>	p <sup>‡</sup>
	Standard	Substitution	Estimated Difference <sup>*</sup>		
Piper Fatigue Scale (Total)					
M	3.7	4.6	-0.9	-1.72	0.09
SD	2.1	2.3	0.5		
Fatigue <sup>§</sup>					
M	40.0	52.0	-12.0	-2.05	0.053
SD	25.0	26.0	6.0		
Beck Depression Inventory-II					
M	8.4	10.4	-2.3	-1.36	0.16
SD	7.6	7.5	1.7		
General Health Questionnaire -30					
M	25.3	29.3	-4.7	-1.46	0.14
SD	12.0	13.8	3.2		
Feeling Sad <sup>§</sup>					
M	26.0	32.0	-7.0	-0.90	0.26
SD	30.0	35.0	7.8		
Feeling Nervous <sup>§</sup>					
M	24.0	28.0	-3.0	-0.40	0.36
SD	24.0	27.0	6.9		
Feeling Irritable <sup>§</sup>					
M	30.0	36.0	-5.0	-0.78	0.29
SD	27.0	30.0	6.5		
Forgetfulness <sup>§</sup>					
M	36.0	41.0	-5.0	-0.69	0.31
SD	27.0	30.0	7.4		
Problems Concentrating <sup>§</sup>					
M	29.0	41.0	-12.0	-1.54	0.12
SD	27.0	35.0	7.6		
Slow Thinking <sup>§</sup>					
M	32.0	39.0	-7.0	-0.85	0.20
SD	28.0	31.0	7.8		
Attention Problems <sup>§</sup>					
M	28.0	39.0	-11.0	-1.46	0.14
SD	28.0	34.0	7.6		

\* The mean on standard treatment minus the mean on substitution treatment as estimated by the sum of the mean difference for each sequence divided by two. The error estimate associated with this difference is a standard error of measurement.

<sup>†</sup> Denotes t tests for a one-factor crossover design.

<sup>‡</sup> Denotes p-values associated with the estimated difference between treatment means.

<sup>§</sup> Denotes symptom was measured using a visual analogue scaling model. Higher scores mean the subject felt worse.

**Table 3**  
Effect of Treatments on Working Memory

Outcome Measure	Treatment			$t^{\ddagger}$	$p^{\S}$
	Standard	Substitution	Estimated Difference*		
L-N Sequencing <sup>§</sup>					
M	12.8	12.5	0.2	0.53	0.34
SD	3.1	3.1	0.4		
S-S Forward <sup>§</sup>					
M	9.3	9.4	-0.1	-0.30	0.38
SD	2.3	2.1	0.3		
S-S Backward <sup>§</sup>					
M	8.6	8.7	-0.1	-0.36	0.37
SD	2.1	2.1	0.3		
Digit Span Forward					
M	10.7	11.0	-0.4	-1.00	0.24
SD	2.5	2.3	0.4		
Digit Span Backward					
M	8.0	8.0	-0.1	-0.19	0.39
SD	3.5	3.2	0.4		
Working Memory IQ <sup>//</sup>					
M	117.5	116.8	0.4	0.16	0.39
SD	21.4	20.6	2.4		

\* The mean on standard treatment minus the mean on substitution treatment as estimated by the sum of the mean difference for each sequence divided by two. The error estimate associated with this difference is a standard error of measurement.

<sup>‡</sup> Denotes t tests for a one-factor crossover design.

<sup>§</sup> Denotes p-values associated with the estimated difference between treatment means.

<sup>§</sup> L-N = Letter Number Sequencing. S-S = Spatial Span tests.

<sup>//</sup> Letter Number Sequencing and Spatial Span tests are subtests of the Working Memory IQ score. Digit Span tests are optional and do not influence the Working Memory IQ score. Higher scores on all subtests mean the subject performed better.

**Table 4**  
Effect of Treatments on Biometric Variables and Symptoms of Hypothyroidism

Outcome Measure	Treatment			$t^{\dagger}$	$p^{\ddagger}$
	Standard	Substitution	Estimated Difference*		
Weight (lbs)					
M	174.2	171.9	1.6	0.87	0.27
SD	42.3	40.8	1.8		
Heart rate (beats/minute)					
M	77.0	77.0	-0.5	-0.26	0.38
SD	9.0	12.0	1.8		
Systolic blood pressure (mmHg)					
M	117.0	119.0	-2.7	-0.86	0.27
SD	17.0	15.0	3.1		
Diastolic blood pressure (mmHg)					
M	68.0	69.0	-1.6	-0.99	0.24
SD	9.0	7.0	1.6		
Serum thyrotropin ( $\mu$ U/ml) <sup>§</sup>					
M	2.7	5.6	-2.7	-1.33	0.16
SD	2.8	10.6	2.0		
Free thyroxine index <sup>§</sup>					
M	3.1	2.4	0.7	5.66	<0.001
SD	0.5	0.6	0.1		
Total serum thyroxine <sup>§</sup> ( $\mu$ g/dl)					
M	10.9	7.9	3.0	6.41	<0.001
SD	2.7	2.4	0.5		
Total serum triiodothyronine <sup>§</sup> (ng/dl)					
M	80.0	99.0	-20.5	-3.30	0.004
SD	18.0	27.0	6.2		
Constipation <sup>//</sup>					
M	32.0	35.0	-3.6	-0.54	0.34
SD	33.0	32.0	6.6		
Feeling Cold <sup>//</sup>					
M	21.0	25.0	-5.2	-1.11	0.21
SD	23.0	24.0	4.6		
Dry Skin <sup>//</sup>					
M	41.0	44.0	-4.2	-0.68	0.31
SD	27.0	33.0	6.2		
Feeling Sleepy <sup>//</sup>					
M	40.0	47.0	-8.8	-1.35	0.18
SD	28.0	33.0	6.5		

\* The mean on standard treatment minus the mean on substitution treatment as estimated by the sum of the mean difference for each sequence divided by two. The error estimate associated with this difference is a standard error of measurement.

<sup>†</sup> Denotes t tests for a one-factor crossover design.

<sup>‡</sup> Denotes p-values associated with the estimated difference between treatment means.

<sup>§</sup> Normal reference ranges: thyrotropin (0.32 – 5.00), free thyroxine index (1.1 – 4.4), total serum thyroxine (4.5 – 12.5), and total serum triiodothyronine (50 – 148).

<sup>//</sup> Denotes symptoms were measured using a visual analogue scaling model. Higher scores mean the subject felt worse.

**Table 5**  
Selected Means ( $\pm$  SD) \* at Baseline, Six, and Twelve Weeks by Sequence

Variables Measured	Baseline	Standard + Placebo	Substitution Treatment
Sequence 1 (n = 12)			
Piper Fatigue Scale Total	4.4 (2.0)	3.0 (1.9)	4.8 (1.7)
Fatigue <sup>†</sup>	48.3 (27.1)	34.0 (22.2)	57.6 (23.2)
Beck Depression Inventory-II	14.2 (8.8)	5.2 (6.4)	9.9 (7.8)
General Health Questionnaire-30	30.1 (7.3)	20.0 (8.7)	31.8 (13.4)
Working Memory IQ <sup>‡</sup>	111.1 (15.4)	114.5 (18.6)	116.9 (19.1)
Thyrotropin ( $\mu$ U/ml) <sup>‡</sup>	1.7 (1.2)	2.9 (3.4)	3.3 (4.0)
Total serum thyroxine <sup>‡</sup> ( $\mu$ g/dl)	11.0 (2.5)	10.8 (3.0)	7.6 (1.9)
Total serum triiodothyronine <sup>‡</sup> (ng/dl)	76.2 (21.2)	73.5 (19.1)	104.1 (25.0)
Variables Measured	Baseline	Substitution Treatment	Standard + Placebo
Sequence 2 (n = 15)			
Piper Fatigue Scale Total	5.0 (2.0)	4.3 (2.6)	4.2 (2.2)
Fatigue <sup>†</sup>	51.5 (28.6)	46.8 (28.7)	45.5 (26.7)
Beck Depression Inventory-II	14.2 (10.2)	10.8 (7.5)	10.9 (7.8)
General Health Questionnaire-30	32.3 (9.9)	27.3 (14.3)	29.6 (12.9)
Working Memory IQ <sup>‡</sup>	110.7 (19.0)	116.7 (22.4)	119.9 (23.8)
Thyrotropin ( $\mu$ U/ml) <sup>‡</sup>	1.8 (2.0)	7.6 (13.8)	2.5 (2.5)
Total serum thyroxine <sup>‡</sup> ( $\mu$ g/dl)	11.2 (1.2)	8.2 (2.7)	10.9 (2.7)
Total serum triiodothyronine <sup>‡</sup> (ng/dl)	79.4 (16.0)	94.5 (29.4)	86.5 (16.1)

\* Standard deviations.

<sup>†</sup> Denotes symptom measured using a visual analogue scaling model. Higher scores mean the subject felt worse.

<sup>‡</sup> Normal reference ranges: thyrotropin (0.32 – 5.00), total serum thyroxine (4.5 – 12.5), and total serum triiodothyronine (50 – 148).