Effect of Thyroid Function Variations Within the Laboratory Reference Range on Health Status, Mood, and Cognition in Levothyroxine-Treated Subjects

Mary H. Samuels,[1] Irina Kolobova,[2] Anne Smeraglio,[3] Meike Niederhausen,[1,4,5]
Jeri S. Janowsky,[6] and Kathryn G. Schuff[1]

Background: There has been recent debate within the thyroid field regarding whether current upper limits of the thyrotropin (TSH) reference range should be lowered. This debate can be better informed by investigation of whether variations in thyroid function within the reference range have clinical effects. One important target organ for thyroid hormone is the brain, but little is known about variations in neurocognitive measures within the reference range for thyroid function.

Methods: This was a cross-sectional study of 132 otherwise healthy hypothyroid subjects receiving chronic replacement therapy with levothyroxine (LT4) who had TSH levels across the full span of the laboratory reference range (0.34–5.6 mU/L). Subjects underwent detailed tests of health status, mood, and cognitive function, with an emphasis on memory and executive functions.

Results: Subjects with low-normal (≤2.5 mU/L) and high-normal (>2.5 mU/L) TSH levels did not differ on most tests of health status, mood, or cognitive function, and there were no correlations between TSH, free T4, or free T3 levels and most outcomes. There was, however, a suggestion that thyroid function affected performance on the Iowa Gambling Task, which mimics real life decision-making. Subjects with low-normal TSH levels made more advantageous decisions than those with high-normal TSH levels.

Conclusions: Variations in thyroid function within the laboratory reference range do not appear to have clinically relevant effects on health status, mood, or memory in LT4 treated subjects. However, decision making, which encompasses many executive functions, may be affected. Unless further studies strengthen this finding, these data do not support narrowing the TSH reference range.

Introduction

S ERUM THYROTROPIN (TSH) levels provide a highly sensitive measure of thyroid function, but there is increasing debate over the optimal reference TSH range. Recent population-based data suggest that the current upper limit of the TSH reference range is skewed by subjects with occult mild hypothyroidism, leading to recommendations that it be lowered (1). However, other analyses show that the upper TSH reference range increases normally with age, suggesting that age-based TSH reference ranges should be used (2). This debate has enormous public health implications, since high-normal TSH levels are common. Lowering the upper limit of the TSH reference range would instantly label tens of millions of people

with thyroid disease, and would affect levothyroxine (LT4) doses in millions more with existing thyroid disease (1).

This debate can be better informed by knowledge of the clinical consequences of variations in thyroid function within the reference range. Such consequences would strengthen the argument that the TSH reference range should be more narrowly defined. In fact, a number of studies have reported clinically relevant health consequences based on variations in thyroid function within the reference range. These outcomes have mainly been cardiovascular risk factors and events, metabolic parameters, bone density, and fracture risk [reviewed in Taylor *et al.* (3)].

The brain is another important target organ for thyroid hormone, but less is known regarding how psychological or

---

[1]Division of Endocrinology, Diabetes, and Clinical Nutrition; [4]Biostatistics and Design Program, OHSU-PSU School of Public Health; Oregon Health and Science University, Portland, Oregon.
[2]Department of Human Development and Family Science, East Carolina University, Greenville, North Carolina.
[3]Stanford University School of Medicine, Department of Internal Medicine, Stanford, California.
[5]Department of Mathematics, University of Portland, Portland, Oregon.
[6]Saturday Academy, Portland, Oregon.

cognitive function varies across TSH or thyroid hormone reference ranges. In euthyroid subjects without thyroid disease, depression, anxiety, or cognitive decrements have been linked to variations in TSH or free thyroxine (fT4) levels (4–17). However, the correlations have shown improvements or decrements with higher or lower thyroid function, depending on the report. Other recent large population-based studies found no correlations between normal and near-normal TSH levels and depression, anxiety, or cognitive tests (18–21). These latter studies circumvented biases inherent in studies of selected subjects, but they often utilized relatively insensitive screening tests for global cognitive function, which may miss subtle but important effects in specific, relevant cognitive domains.

Hypothyroid patients receiving LT4 therapy represent a special subset of ''euthyroid'' subjects. Despite normal TSH levels, many patients continue to complain of impaired health status, mood, and cognitive function. This often leads patients to request to increase their levothyroxine (LT4) doses or try alternate thyroid preparations. However, it is not clear that targeting the lower end of the TSH reference range will achieve patients' desired outcomes. The observational literature is divergent, and there is only one controlled trial addressing this, which did not show significant effects (11,22–28).

To address this issue, we recruited subjects with primary hypothyroidism receiving replacement doses of LT4, who then underwent extensive testing for health status, mood, and cognitive function. Rather than using less sensitive global screening tests, we employed intensive, sensitive measures that targeted two specific cognitive domains: executive function and memory. Our decision to focus on memory was based on our previous data and other studies suggesting that memory is preferentially affected in subjects with mild thyroid dysfunction (21,29,30), as well as animal studies that support a major role for LT4 in brain areas that mediate memory (31,32). Our decision to focus on executive function was based on the relative lack of information regarding thyroid effects on this cognitive domain, although studies and clinical observations suggest that this critical cognitive process is also affected [reviewed by Samuels (33)]. We hypothesized that LT4 treated subjects with lower TSH levels within the reference range have better health status, mood, memory, and executive function compared with subjects with higher TSH levels within the reference range. We also correlated these outcomes with free T4 and free triiodothyronine (fT3) levels, since some studies report associations with thyroid hormone, rather than TSH, levels.

Experimental subjects

One hundred thirty-two subjects receiving LT4 therapy for hypothyroidism, with TSH levels within the laboratory reference range, were recruited as a convenience sample between February 2009 and August 2012 from the authors' clinics, through review of electronic health records, by flyers, and with mailings. Twelve were male and 120 were female. They were aged 21–70 years and were receiving LT4 for primary hypothyroidism ($n = 102$); hypothyroidism following ^{131}I therapy for Graves' disease ($n = 13$); postpartum thyroiditis leading to permanent hypothyroidism ($n = 4$); or a history of thyroid surgery ($n = 13$). All were diagnosed as adults and had received LT4 therapy for 5 months to 50 years (mean 12 years).

All subjects had past elevated TSH levels. Levothyroxine doses were stable for at least 3 months (mean time on current LT4 dose 2 years). None of the subjects had any acute or chronic illness or were on medications that affect thyroid hormone levels, mood, or cognition. Stable doses of oral contraceptive or estrogen therapy were allowed. Testing was done during the first 14 days after onset of menstrual bleeding or an oral contraceptive cycle in premenopausal women.

Materials and Methods

Experimental design

The protocol was approved by the Oregon Health and Science University (OHSU) Institutional Review Board, and subjects gave written informed consent.

Screening visit. Subjects were screened for general health, medicines, thyroid status, and mood or cognitive disorders by history, physical examination and laboratory testing. General intelligence was estimated by the Wechsler Adult Intelligence–Revised (WAIS-R) Vocabulary subtest (34).

Testing visit. Within six weeks of the screening visit, subjects returned for a three- to four-hour testing visit. Subjects refrained from taking their LT4 dose that morning. Serum TSH, fT4, and fT3 levels were obtained at the beginning of each visit. Sixty-one of the samples were collected between 0700 and 0859 hours, 59 between 0900 and 1159 hours, and 12 between 1200 and 1359 hours, due to scheduling limitations for some subjects. Subjects self-completed the following measures of health status and mood:

A. *The Thirty-Six Item Short Form Health Survey (SF-36)*, a questionnaire about general health (35). Higher scores on the summary scales and subscales reflect better health status and well-being.

B. *The Profile of Mood States (POMS)*, a questionnaire about mood (36). Higher scores on POMS subscales reflect mood decrements, except for the vigor subscale, where higher scores reflect improved mood.

C. *The Affective Lability Scale (ALS)*, a questionnaire on which subjects rate their agreement with statements regarding the tendency of their moods to shift between baseline and anger, depression, elation, and anxiety (37). Higher scores indicate increased lability of mood.

Cognitive tests were administered by a single experienced research assistant. Based on existent literature and our previous studies, we did not utilize a battery of general cognitive measures, but rather sensitive measures targeted to specific domains likely to be affected by altered thyroid status:

Test of declarative memory

A. *Paragraph Recall Test (verbal memory)*. Subjects were read a brief story and verbally recalled it immediately and after a 30 minute interval. The outcome measure was the total number of story elements recalled at each interval (38).

Tests of executive function

A. *Attention/concentration: The Letter Cancellation Test.* This test consisted of a sheet of paper with six lines of 52 letters in random sequence. The subject was instructed

to circle two specified target letters each time they appeared in the sequence, as quickly as possible. The score was the number of errors and the time taken to complete the task (39).

B. *Cognitive flexibility: The Trail Making Test.* In this task, the subject connected circles as quickly as possible without lifting the pen from the paper. In Part A, the circles were numbered and the subject was instructed to draw lines to connect the numbers in ascending order (1 to 2, 2 to 3, etc.). In Part B, the circles included both numbers and letters; the subject again drew lines to connect the circles in an ascending pattern, but also had to alternate between numbers and letters (1-A-2-B-3-C, etc.). Subjects were scored on the number of errors and time to complete the task (40).

C. *Decision making: The Iowa Gambling Task.* This task consisted of four decks of cards on a computer screen, shown face down. The subject had to choose cards from any deck, and each card resulted in the gain or loss of money. The subject was unaware that two of the decks were advantageous (small gains but smaller losses), while the other two decks were disadvantageous (large gains but even larger losses). The subject's choices were classified as advantageous (X) or disadvantageous (Y), with a net score of X – Y, over five trials of 100 cards each. A net score of zero is chance performance. This task assesses real-life decision making and responses to rewards and punishments (41).

D. *Working memory tests.*

i. *N-Back Test.* A series of letters was presented one at a time on a computer screen. Subjects responded each time a letter appeared that they had seen on the previous screen (one-back). The task was repeated with an increase in memory load by imposing intervening letters while the subjects had to hold in mind letters that had appeared two-back and then three-back. Outcome measures were the total number correct (on target) and the total number incorrect (off target) for each condition (42).

ii. *Subject Ordered Pointing.* Subjects were presented with a series of computer screens which had abstract drawings on them (6, 8, 10, or 12 per screen). Each screen in a set showed the same array of abstract drawings, but in a different spatial arrangement. The subject was instructed to indicate one drawing per screen using computer keys. They were to avoid choosing the same drawing on subsequent screens in the set. Subjects erred when they chose a drawing that had been previously chosen. Each set was repeated three times. The outcome measure was the total number of errors across each screen set (43). Similar to the N-Back test, the subject must hold the drawings in mind and inhibit responses to previously selected drawings.

Tests of motor learning and motor memory:

A. *Pursuit Rotor.* Subjects held a photosensitive wand to maintain contact with a 2 cm light disk rotating on a turntable (Model 30014, Lafayette Instrument Company). Two blocks of eight 20 second trials were administered, with a 20 second rest after each trial, and a 60 second rest period after four trials. After a 30 minute interval, the two blocks were repeated. The outcome measure was the mean total time the stylus remained on target (44).

B. *Motor Sequence Learning Test.* The subject memorized two keypress sequences on a computer, each associated with a letter of the alphabet. A "T" was associated with the sequence 1-3-2 and an "H" was associated with 3-1-2. As soon as the T or H appeared on the screen, the subject performed the appropriate sequence as quickly as possible. Subjects performed 10 blocks of 18 trials each. The outcome was the average total movement time (time from character presentation to completion of the keypress sequence) (45).

Analytic methods

TSH was measured by immunochemiluminometric assay (Beckman Coulter); functional sensitivity 0.02 mU/L, reference range 0.34–5.60 mU/L, interassay coefficient of variability (CV) 5% at 0.70 mU/L. Free T4 was measured by direct equilibrium dialysis (Quest Diagnostics); sensitivity 0.08 ng/dL, reference range 0.8–2.7 ng/dL, interassay CV 6.8% at 0.3 ng/dL and 1.6% at 3.8 ng/dL. Free T3 was measured by tracer dialysis (Quest Diagnostics); sensitivity 25 pg/dL, reference range 210–440 pg/dL, interassay CV 4%. TSH levels were measured at the time of testing, with no change in assay characteristics during the period of the study. Free T4 and fT3 levels were batched and analyzed at the end of the study. All samples were run in duplicate.

Statistical methods

Subjects were divided into two groups based on serum TSH levels. The Low-Normal TSH group was defined as subjects with a TSH between 0.34 (the lower limit of the assay reference range) and 2.50 mU/L. The High-Normal TSH group was defined as subjects with a TSH between 2.51 and 5.60 mU/L (the upper limit of the assay reference range). The cut-off of 2.50 mU/L for the two groups was decided based on recent debate within the thyroid field over whether the TSH reference range should be restricted to an upper limit of 2.5 mU/L to achieve a Gaussian distribution of TSH levels in healthy populations and exclude subjects with possible incipient hypothyroidism (46). There were 85 subjects in the Low-Normal TSH group and 47 in the High-Normal TSH group. Seven of the 12 men were in the Low-Normal TSH group and 5 were in the High-Normal TSH group.

Health status, mood, and cognitive outcomes were compared between the two groups. Subscales of each measure were analyzed together using linear repeated measures analyses (R version 3.2.1) (47) using the non linear mixed effects (NLME) package lme function (48) or, for binary outcomes, generalized estimating equations (geepack package geeglm function) (49). These methodologies allow for correlation between subscale measures for each subject. For continuous outcomes, a compound symmetric covariance structure was used to analyze the data. These models included adjustments for age, WAIS-R vocabulary score, years in school, body mass index, estrogen status, duration of time on LT4, duration of time at current LT4 dose, and LT4 dose (μg/kg). Analysis of binary outcomes used compound symmetric covariance matrices and was unadjusted due to the more limited nature of the data.

An initial assessment of interaction between group and subscale was obtained for each set of subscales. Likelihood ratio

tests were conducted to determine whether models with the interaction were significant at level 0.10, in which case a comparison of groups was conducted for each subscale. If the addition of the interaction was not significant, the comparison of groups was conducted for the set of scales as a whole (the main effect of group was analyzed, dropping the interaction from the model). To limit the effect of multiple comparisons, the original plan was to conduct follow-up comparisons of groups for each subscale individually only if evidence of a group effect was observed at level 0.05. However, since none of these tests were significant, follow-up comparisons for all individual subscales were tested to confirm the lack of significance.

In addition, we examined relationships between outcomes and TSH, fT4, and fT3 as continuous variables using the same repeated measures methodology but substituting, in separate models, the selected hormone for the categorical group variable of low-normal and high-normal TSH.

Results

Clinical parameters and thyroid function tests

The low-normal and high-normal TSH groups were well matched for age, WAIS-R Vocabulary score, years in school, sex, estrogen status, body mass index, duration of LT4 treatment, and duration at current LT4 dose (Table 1). As would be predicted, LT4 doses were higher in the low-normal TSH group compared with the high-normal TSH group (1.51 ± 0.05 vs. $1.26 \pm 0.06 \mu g/kg/day$, $p = 0.002$). By design, all subjects had TSH levels within the reference range, with mean TSH levels lower in the low-normal TSH group than the high-normal TSH group (1.35 ± 0.07 vs. 3.60 ± 0.11 mU/L, $p < 0.0001$). Mean fT4 and fT3 levels were similar in the two groups. No subject had a fT4 level outside the reference range. Seventy subjects had low fT3 levels, between 118 and 209 pg/dL (reference

range 210–440 pg/dL). Forty of these subjects were in the Low-Normal TSH group (47% of this group) and 30 were in the high-normal TSH group (64% of this group).

Health status and mood: SF-36, POMS, ALS

There were no significant differences between the low-normal TSH and high-normal TSH groups in SF-36, POMS, or ALS overall scales or subscales (Table 2). Analyzing TSH, fT4, and fT3 as continuous variables across both groups (Table 3), the POMS anger subscale was negatively correlated with increasing fT4 levels ($p = 0.04$), although the magnitude of the correlation was small (0.5 unit decrease in POMS-Anger for each 1.0 ng/dL increase in fT4 level). There were no other significant correlations between TSH, fT4, or fT3 and health status or mood measures.

Cognitive tests

The interaction between thyroid groups and the Iowa Gambling Task (IGT) net scores was significant ($p = 0.049$), indicating that the low-normal TSH and high-normal TSH groups exhibited different trends across the five trials of their IGT net scores. Both groups started with disadvantageous decision making with the first deck of cards (Net-1), with the low-normal TSH group significantly worse than the high-normal TSH group ($p = 0.02$). Both groups learned to make advantageous decisions with subsequent decks, but the low-normal TSH group improved more with each deck than the high-normal TSH group. The low-normal TSH group showed a pattern of increased learning to choose advantageously (better decision making), whereas the high-normal TSH group plateaued earlier and did not show consistent improvement across trials (Fig. 1). There were no significant differences between the two groups for other cognitive outcomes (Table 4).

TABLE 1. DEMOGRAPHIC, CLINICAL, AND THYROID FUNCTION VARIABLES IN THE LOW-NORMAL THYROTROPIN AND HIGH-NORMAL THYROTROPIN GROUPS

	Low-normal TSH mean ± SEM n = 85	High-normal TSH mean ± SEM n = 47	p-Value
Age (years)	48.9 ± 1.2	49.6 ± 1.9	0.74
WAIS-R vocabulary subtest	10.77 ± 0.21	10.98 ± 0.33	0.59
Years in school	16.2 ± 0.3	15.6 ± 0.5	0.34
Sex	92% Female	89% Female	0.76
	8% Male	11% Male	
Estrogen status	8% Male	11% Male	0.89
	38% Pre-none	43% Pre-none	
	11% Pre-on	9% Pre-on	
	39% Post-none	36% Post-none	
	5% Post-on	2% Post-on	
BMI (kg/m^2)	27.5 ± 0.6	26.9 ± 0.9	0.61
LT4 time (years)	12.6 ± 1.1	11.4 ± 1.2	0.47
LT4 dose time (years)	1.66 ± 0.23	2.06 ± 0.40	0.38
LT4 dose ($\mu g/kg$)	**1.51 ± 0.05**	**1.26 ± 0.06**	**0.002**
TSH (mU/L)	**1.35 ± 0.07**	**3.60 ± 0.11**	**<0.0001**
Free T4 (ng/dL)	1.67 ± 0.04	1.59 ± 0.05	0.20
Free T3 (pg/dL)	215 ± 5	208 ± 6	0.40

Significant differences between groups are shown in bold with corresponding *p*-values.
BMI, body mass index; T3, triiodothyronine; LT4 dose time, duration of current levothyroxine (LT4) dose; LT4 time, duration of LT4 therapy; Pre-none, premenopausal, no hormone treatment; Pre-on, premenopausal on hormone treatment; Post-none, postmenopausal, no hormone treatment; Post-on, postmenopausal on hormone treatment; TSH, thyrotropin; WAIS-R, Wechsler Adult Intelligence–Revised.

TABLE 2. HEALTH STATUS AND MOOD IN THE LOW-NORMAL TSH AND HIGH-NORMAL TSH GROUPS

Measure	Low-normal TSH mean ± SEM n = 85	High-normal TSH mean ± SEM n = 47	p-Value
Short Form-36			
MCS (mental component summary)	45.6 ± 0.9	45.5 ± 0.9	0.74
PCS (physical component summary)	52.3 ± 0.7	52.8 ± 0.9	0.84
GH (general health)	77.8 ± 1.8	78.4 ± 2.3	0.91
MH (mental health)	57.8 ± 1.1	57.1 ± 1.3	0.44
VT (vitality)	58.2 ± 2.2	55.0 ± 2.8	0.25
BP (bodily pain)[a]	29% high	29% high	0.97
PF (physical functioning)[a]	41% high	44% high	0.66
RP (role physical)[a]	69% high	73% high	0.52
SF (social functioning)[a]	60% high	64% high	0.59
RE (role emotional)[a]	78% high	76% high	0.82
Profile of Mood States[b]			
A (anger)	4.3 ± 0.5	4.6 ± 0.6	0.38
C (confusion)	6.5 ± 0.4	6.2 ± 0. 4	0.97
D (depression)	4.4 ± 0.6	5.5 ± 0.7	0.11
F (fatigue)	6.6 ± 0.6	7.3 ± 0.7	0.17
T (tension)	6.4 ± 0.4	6.6 ± 0.5	0.60
V (vigor)	16.0 ± 0.7	14.8 ± 0.8	0.35
Affective Lability Scale			
Bipolar	0.7 ± 0.1	0.6 ± 0.1	0.55
Depression	1.0 ± 0.1	1.0 ± 0.1	0.78
Elation	0.8 ± 0.1	0.8 ± 0.1	0.72
Anger[c]	57% score >0	70% score >0	0.19
Anxiety[c]	81% score >0	83% score >0	0.79
Anxiety depression[c]	73% score >0	78% score >0	0.55

p-Values for continuous outcomes were adjusted for age, years of education, WAIS-R, BMI, estrogen status, LT4 time, LT4 dose time, and LT4 dose (μg/kg).

[a]For these variables, the distributions within each group were highly skewed. The highest observed values of each subscale were used as the cut-points for producing a dichotomous measure. For BP, the highest observed value was 90, whereas for the other scales, the highest observed value was 100.

[b]Profile of Mood States values were natural log-transformed prior to analysis because the raw data were skewed. All values were increased by one before the transformation due to the presence of zeros. The vigor subscale was analyzed separately since it was the only scale for which higher values represented improved mood.

[c]These scores were compared as the proportion positive between the groups, since the measures were skewed and contained a large proportion of zeros.

Analyzing TSH, fT4, and fT3 as continuous variables across both groups (Table 5), the IGT Net-1 (baseline) score was positively correlated with TSH levels, indicating better performance (1.65 unit increase for every 1 mU/L increase in TSH level, p = 0.02). There were no correlations with the N-Back number correct target or incorrect nontarget, except the three-back number incorrect nontarget was negatively correlated with fT4 levels (p = 0.04). Further, the magnitude of the correlation was small (1.05 unit decrease in three-back number incorrect nontarget for every 1 ng/dL increase in fT4 level). There were no correlations with accuracy on the Trail Making Test, although the time to complete the test was positively correlated with fT3 levels (p = 0.049). This indicated worse performance, although again the magnitude of the correlation was small (0.26 second increase in time for each 10 pg/dL increase in fT3 level). There were no other significant correlations between TSH, fT4, or fT3 and cognitive measures.

Discussion

In this cohort of LT4 treated subjects who had TSH levels across the full span of the laboratory reference range, we found little evidence that variations in thyroid function were correlated with health status, mood, or memory. This was true whether the data were analyzed as dichotomous variables (low-normal vs. high-normal TSH) or as continuous variables. There was a suggestion in both analyses of the IGT that higher thyroid function (lower TSH levels) within the reference range was associated with better "real-life" decision making. However, given the numbers of measures and the weak effects, additional study would be needed to verify this finding.

The published literature regarding these neurocognitive outcomes in LT4 treated subjects has been divergent and inconclusive. Some studies have shown that LT4 treated subjects have decreased psychological well-being or cognitive measures compared with control groups, while other studies have reported no differences in these measures (11,22–27). Only four studies further investigated whether outcomes varied by TSH or thyroid hormone levels within the reference range in LT4 treated subjects, again with divergent findings (11,23,25,27). Our study extends these limited published data in LT4 treated euthyroid subjects. We did not find clinically significant associations between

Table 3. Correlations Between Thyroid Hormone Levels and Health Status and Mood Measures in LT4 Treated Subjects

Measure	fT4		fT3		TSH	
	Coefficient	p-Value	Coefficient	p-Value	Coefficient	p-Value
Short Form-36						
MCS (mental component summary)	0.48	0.82	0.16	0.28	−0.55	0.35
PCS (physical component summary)	−0.58	0.73	−0.13	0.24	0.77	0.09
GH (general health)	3.14	0.45	−0.19	0.49	0.59	0.63
MH (mental health)	2.71	0.35	−0.03	0.89	−1.22	0.11
VT (vitality)	2.65	0.61	0.36	0.30	−0.90	0.54
BP (bodily pain)[a]	−7%	0.89	−2%	0.66	12.44%	0.44
PF (physical functioning)[a]	109%	0.14	−3%	0.39	4.84%	0.74
RP (role physical)[a]	16%	0.79	5%	0.30	17.89%	0.29
SF (social functioning)[a]	6%	0.91	−4%	0.26	4.21%	0.77
RE (role emotional)[a]	−10%	0.85	4%	0.39	−5.83%	0.71
Profile of Mood States[b]						
A (anger)	**−0.50**	**0.04**	−0.001	0.94	0.09	0.20
C (confusion)	0.09	0.41	0.008	0.28	0.01	0.86
D (depression)	−0.10	0.69	0.01	0.44	0.11	0.12
F (fatigue)	−0.22	0.31	−0.007	0.65	0.03	0.65
T (tension)	0.05	0.70	−0.00006	0.995	0.06	0.11
V (vigor)	0.17	0.14	0.003	0.72	−0.04	0.16
Affective Lability Scale						
Bipolar	0.07	0.63	−0.003	0.73	−0.06	0.14
Depression	0.03	0.85	−0.005	0.67	−0.02	0.57
Elation	−0.07	0.67	0.005	0.61	−0.02372	0.55
Anger[c]	27%	0.63	0.05%	0.99	2.24%	0.88
Anxiety[c]	−44%	0.35	−0.7%	0.88	−6.81%	0.69
Anxiety depression[c]	−14%	0.79	3%	0.52	14.89%	0.40

Correlations were analyzed by repeated measures methodology using separate models for each hormone. Positive coefficients indicate that the measure increased with increasing hormone levels, while negative coefficients indicate that the measure decreased with increasing hormone levels.

For continuous measures, the magnitude of the coefficient indicates the estimated change in the measure with a 1 unit increase in free T4 (fT4) or TSH, or a 10 unit increase in fT3.

For binary measures, coefficients were transformed to estimate the percent change in the predicted odds of the measure for a 1 unit increase in fT4 or TSH, or a 10 unit increase in fT3. The transformed coefficients are estimates of the risk ratios associated with the 1 or 10 unit increase in the respective hormone level.

p-Values for continuous outcomes were adjusted for age, years of education, WAIS-R, BMI, estrogen status, LT4 time, LT4 dose time, and LT4 dose (μg/kg). Significant coefficients are shown in bold with corresponding p-values.

[a]For these variables, the distributions within each group were highly skewed. The highest observed values of each subscale were used as the cut-points for producing a dichotomous measure. For BP, the highest observed value was 90, whereas for the other scales, the highest observed value was 100.

[b]Profile of Mood States values were natural log–transformed prior to analysis because the raw data were skewed. All values were increased by one before the transformation due to the presence of zeros. The vigor subscale was analyzed separately since it was the only scale for which higher values represented improved mood. The magnitude of the coefficient indicates the estimated change in the natural log of the measure plus one with a 1 unit (10 units for fT3) increase.

[c]These scores were compared as the proportion positive between the groups, since the measures were skewed and contained a large proportion of zeros.

FIG. 1. Net scores for each card deck for the Iowa Gambling Task in the low-normal and high-normal thyrotropin (TSH) groups. Mean Net-1 score was significantly lower in the low-normal TSH group (p = 0.02). Mean Net-4 and Net-5 scores were significantly higher in the low-normal TSH group (p = 0.01, p = 0.04 respectively).
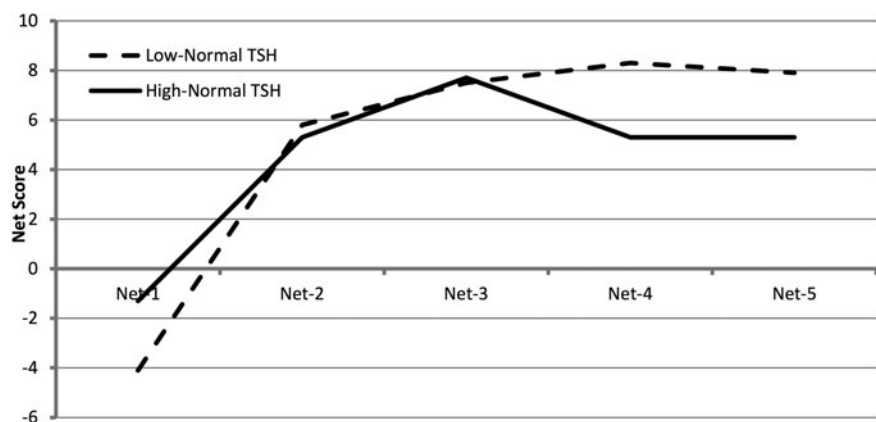
TABLE 4. COGNITIVE MEASURES IN THE LOW-NORMAL TSH AND HIGH-NORMAL TSH GROUPS

	Test	Low-normal TSH mean ± SEM	High-normal TSH mean ± SEM	p-Value
Declarative memory	**Paragraph Recall Test**			
	Immediate	12.6 ± 0.3	12.8 ± 0.5	0.41
	30 min delay	11.1 ± 0.4	11.3 ± 0.6	0.45
Executive function	**Letter Cancellation Test**			
	Time (sec)	100 ± 2	106 ± 3	0.15
	% with no errors	20%	19%	0.91
	Trail Making Test			
	Time (sec)	24.2 ± 0.7	24.6 ± 1.3	0.54
	ABC time (sec)	61.8 ± 2.5	60.0 ± 2.6	0.48
	% With errors	11%	15%	0.47
	% With ABC errors	32%	28%	0.53
	Iowa Gambling Task			
	Net-1	**−4.1 ± 0.9**	**−1.3 ± 1.4**	**0.02**
	Net-2	5.8 ± 0.9	5.3 ± 1.3	0.70
	Net-3	7.5 ± 1.0	7.7 ± 1.5	0.99
	Net-4	**8.3 ± 1.1**	**5.3 ± 1.6**	**0.01**
	Net-5	**7.9 ± 1.2**	**5.3 ± 1.5**	**0.04**
Executive function (working memory)	**N-Back** No. correct on target			
	1-Back[a]	74%	81%	0.30
	2-Back[a]	42%	43%	0.92
	3-Back	10.9 ± 0.3	10.4 ± 0.3	0.24
	N-Back No. incorrect/nontarget			
	1-Back[a]	28%	13%	0.06
	2-Back[a]	33%	30%	0.78
	3-Back	3.5 ± 0.2	3.8 ± 0.3	0.77
	Subject Ordered Pointing No. of errors			
	6	0.6 ± 0.05	0.5 ± 0.1	0.79
	8	1.1 ± 0.1	1.1 ± 0.1	0.87
	10	1.3 ± 0.1	1.5 ± 0.1	0.29
	12	1.7 ± 0.1	1.7 ± 0.2	0.60
Motor learning	**Pursuit Rotor Trial** Time on target (sec)			
	1	34.0 ± 1.3	34.4 ± 1.6	0.93
	2	36.4 ± 1.3	37.3 ± 1.5	0.60
	3	37.5 ± 1.4	37.4 ± 1.6	0.88
	4	38.7 ± 1.4	38.6 ± 1.6	0.92
	Motor Sequence Learning Test			
	Total movement time (sec)	1123 ± 40	1073 ± 34	0.39

Individual tests are grouped by cognitive subdomains (first column). p-Values for continuous outcomes were adjusted for age, years of education, WAIS-R, BMI, estrogen status, LT4 time, LT4 dose time, and LT4 dose (μg/kg).

Significant differences between the two groups are shown in bold with corresponding p-values.

[a]These values were calculated as proportion of subjects for which each measure was ≥15 (for correct on target) or >0 (for incorrect/nontarget), since there were floor effects. The N-Back p-values for comparing groups do not include three-back, since this variable was not calculated as a proportion.

ABC, Trail Making Test Part B.

thyroid function and health status or mood. Most relevant, ours is the first report that correlates thyroid function in this population with sensitive tests of two cognitive domains, memory and executive function, which map to brain areas known to be responsive to thyroid hormone (30–32).

There were no significant differences in tests of memory between the low-normal and high-normal TSH groups, including the domains of declarative memory, working memory, and motor learning/motor memory. We did find some

associations with memory in our analysis across both groups using TSH and thyroid hormones as continuous variables, but the magnitude of the associations was small and of limited clinical relevance.

Executive functions have not been extensively studied in thyroid disease, since rodent models do not adequately represent complex executive function in humans, and many laboratory-based measures of executive function are labor intensive and/or relatively insensitive to "real-world" daily life

Table 5. Correlations Between Thyroid Hormone Levels and Cognitive Measures in LT4 Treated Subjects

	Test	fT4		fT3		TSH	
		Coefficient	p-Value	Coefficient	p-Value	Coefficient	p-Value
Declarative memory	**Paragraph Recall Test**						
	Immediate	1.07	0.20	−0.005	0.93	0.19	0.38
	30 min delay	0.89	0.31	0.02	0.71	0.20	0.40
Executive function	**Letter Cancellation Test**						
	Time (sec)	−4.97	0.33	−0.04	0.90	2.04	0.13
	% with no errors	15%	0.81	−8%	0.13	5%	0.77
	Trail Making Test						
	Time (sec)	1.67	0.40	**0.26**	**0.05**	0.31	0.56
	ABC time (sec)	−0.54	0.92	0.08	0.83	−1.04	0.46
	% with errors	−23%	0.72	1%	0.82	25%	0.27
	% with ABC errors	−41%	0.32	6%	0.17	−14%	0.33
	Iowa Gambling Task						
	Net-1	0.04	0.99	−0.03	0.88	**1.65**	**0.02**
	Net-2	−1.68	0.50	0.08	0.63	−0.44	0.50
	Net-3	−1.46	0.59	−0.07	0.71	0.25	0.73
	Net-4	−0.94	0.75	0.02	0.94	−1.05	0.17
	Net-5	−2.10	0.48	−0.09	0.67	−1.11	0.16
Executive function (working memory)	**N-Back** No. correct on target						
	1-Back[a]	66%	0.38	−6%	0.13	24%	0.20
	2-Back[a]	−9%	0.84	0.5%	0.90	16%	0.29
	3-Back	−0.27	0.66	0.01	0.77	−0.18	0.30
	N-Back No. incorrect/nontarget						
	1-Back[a]	−66%	0.08	3%	0.56	**−30%**	**0.05**
	2-Back[a]	6%	0.92	−1%	0.82	−8%	0.59
	3-Back	**−1.05**	**0.04**	−0.03	0.37	0.004	0.98
	Subject Ordered Pointing No. of errors						
	6	0.11	0.32	0.001	0.85	−0.02	0.50
	8	−0.04	0.83	−0.002	0.90	−0.0004	0.99
	10	0.17	0.46	−0.001	0.94	0.03	0.67
	12	−0.12	0.70	0.01	0.73	−0.06	0.43
Motor learning	**Pursuit Rotor Trial** Time on target (sec)						
	1	0.52	0.88	0.05	0.83	−0.09	0.93
	2	1.74	0.60	0.01	0.96	0.41	0.65
	3	0.04	0.99	0.07	0.77	−0.18	0.85
	4	3.03	0.37	−0.16	0.49	−0.04	0.97
	Motor Sequence Learning Test						
	Total movement time (sec)	−59.05	0.45	7.20	0.16	−2.49	0.90

Individual tests are grouped by cognitive subdomains (first column).

Correlations were analyzed by repeated measures methodology using separate models for each hormone.

Positive coefficients indicate that the measure increased with increasing hormone levels, while negative coefficients indicate that the measure decreased with increasing hormone levels.

For continuous cognitive measures, the magnitude of the coefficient indicates the estimated change in the measure with a 1 unit increase in fT4 or TSH, or a 10 unit increase in fT3.

For binary cognitive measures, coefficients were transformed to estimate the percent change in the predicted odds of the measure for a 1 unit increase in fT4 or TSH, or a 10 unit increase in fT3. The transformed coefficients are estimates of the risk ratios associated with the 1 or 10 unit increase in the respective hormone level.

p-Values for continuous outcomes were adjusted for age, years of education, WAIS-R, BMI, estrogen status, LT4 time, LT4 dose time, and LT4 dose (μg/kg). Significant coefficients are shown in bold with corresponding p-values.

[a]These values were calculated as proportion of subjects for which each measure was ≥15 (for correct target) or >0 (for incorrect/nontarget), since there were floor effects. The N-Back p-values for comparing groups do not include 3-Back since this variable was not calculated as a proportion.

scenarios. For these reasons, most of the published studies referenced above did not include sophisticated tests of executive functions. We utilized five tests that measure different aspects of this executive cognitive domain. Two of them are relatively simple and widely used tests which measure attention and concentration (Letter Cancellation Test) (39) and attention and cognitive flexibility (Trail Making Test) (40). Two of them are sensitive measures of working memory (N-Back and Subject Ordered Pointing) (42,43). In concordance with the few available reports in non-LT4 treated (7,17) and LT4 treated euthyroid subjects (23,26,27), we did not find associations between thyroid function and these four measures.

A novel strength of our study is the inclusion of an additional test of executive function, the IGT, as a representation of real-world decision making. In the IGT, individuals experience rewards and punishments for selecting cards from decks that provide either high immediate rewards and larger punishments or smaller immediate rewards and smaller punishments. Advantageous decision making relies on shifting choices away from disadvantageous card decks toward advantageous card decks (41). We found that the low-normal TSH group initially made poor decisions compared to the high-normal TSH group (Net-1), but then the low-normal TSH group outperformed the high-normal TSH group with larger gains across subsequent trials. There was also a clinically relevant relationship between TSH and performance across all subjects. The pattern we observed in the low-normal TSH group (consistent improvement across trials) appears similar to that reported in healthy control subjects (49), and better than the plateau observed in the high-normal TSH group. To our knowledge, this is the first report that utilized such a ''real-life'' measure of decision making in LT4 treated subjects (27). However, these results are preliminary, given our limited sample size and multiple comparisons. In addition, it is difficult to explain why the low-normal TSH group would have worse scores initially but then improve to a greater extent; there were no systematic differences, such as season or time of day, in testing between the two groups. It is possible that this represents regression to the mean, since the low-normal group had more room to improve. For these reasons, we interpret this finding as preliminary, and its major importance is to suggest a focus for future studies on thyroid status and executive function.

Complementing our observational data, there are two interventional trials of LT4 therapy in subjects with normal TSH levels. Walsh et al. varied LT4 doses in a blinded fashion in LT4 treated subjects to achieve low-normal or high-normal TSH levels and did not find any effects on hypothyroid symptoms, quality of life, or cognitive function, although executive function was only tested with the Trail-Making Test (28). Pollock et al. administered LT4 or placebo in a blinded fashion to euthyroid (non-LT4 treated) subjects who had hypothyroid-type symptoms, as well as asymptomatic controls, and found no improvement in psychological measures in either group, although cognitive tests were not done (50).

An interesting side finding in our study was the high prevalence of low serum fT3 levels in many LT4 treated subjects with normal TSH levels, even in the low-normal TSH group. This has been described in previous reports [reviewed in Jonklaas et al. (51)], and has led to suggestions that hypothyroid subjects may benefit from L-triiodothyronine (LT3) treatment. However, studies of LT3 add-on or monotherapy in hypothyroidism have not shown improvements in quality of life, mood, or cognitive outcomes, and LT3 is not recommended for standard treatment of hypothyroidism [reviewed in Jonklaas et al. (51)]. Our data showing lack of correlation between fT3 levels and these outcomes is concordant with this recommendation.

Despite our novel emphasis on executive function, our study also has limitations. Our sample size is limited, and it is possible that we missed small effects. However, the small magnitude of effects we report suggest that clinically meaningful alterations for each measure are unlikely, with the possible exception of the IGT. In addition, two studies reported that subjects with subclinical hypothyroidism (mean TSH = 14.7 and 19.4) performed more poorly than euthyroid subjects on the N-Back test. The numbers of subjects in each study was quite small ($n = 11$ and 16), suggesting that the sensitive tests we employed are responsive to small changes in thyroid function and do not require large samples sizes for effects to be apparent (30,52). We performed a relatively large number of correlations for our sample size, although we accounted for this in our analysis, and it is likely that some of our minor findings were due to chance. We did not include a control group of non-LT4 treated euthyroid subjects due to resource constraints, so we cannot state that our subjects had similar outcomes as healthy subjects. However, our results are similar to population norms and our previous studies in healthy control subjects for our test measures (53,54). Most of our subjects were women, and although we included sex as a covariate, it is unclear whether our results would be similar in a larger group of men. Subjects also tended to be younger and slimmer than the overall U.S. population, and our results may not be generalizable to older or heavier subjects receiving LT4 for hypothyroidism. Subjects were heterogeneous in terms of underlying thyroid diagnosis and length of LT4 treatment, although this reflects the reality of recruiting subjects for intensive clinical studies and provides relevance for clinical practice. We attempted to collect blood samples at a consistent time of day to avoid circadian variations in TSH levels, but this was not always possible due to scheduling limitations. In healthy subjects with typical sleep–wake cycles, serum TSH levels generally decrease slightly between 0700 and 0900 hours, and then remain stable until the evening (55). Thus, there may have been slight variation in TSH levels in our study due to sampling time that could limit our results.

It is a common clinical observation that some otherwise healthy patients with hypothyroidism continue to complain of fatigue, poor mood, inability to concentrate, and vague cognitive difficulties (often described as ''brain fog'') despite normal TSH levels. Published studies have attempted to document this (11,22–27), but it can be difficult to map these subjective complaints to specific objective domains. Our test measures most relevant for these complaints include subsets of the SF-36, POMS and ALS for fatigue and mood, as well as the Letter Cancellation and Trail Making Tests (attention, concentration, and cognitive flexibility). We did not find any correlations between thyroid function and these measures. The relevance of the IGT results to these complaints is less clear, as patients may have trouble elucidating specific deficits in executive function beyond phrases like ''brain fog.''

Also of note, psychological well-being in subjects with normal TSH levels seems to depend on whether they have a diagnosed thyroid condition, suggesting that self-knowledge of a thyroid disorder impairs well-being regardless of the TSH

level (11,54). All of our subjects knew they had hypothyroidism, and may have been less satisfied with their health status and therefore more likely to volunteer for a research study. We did not query them regarding dissatisfaction with their LT4 treatment, but it would be fruitful in future studies to correlate this parameter to self-reported health status and mood.

In summary, we found no differences in measures of health status, mood, memory or executive functions in LT4 treated subjects based on whether their TSH levels were above or below 2.50 mU/L, a level that has been suggested as a target for LT4 therapy (46). We also found few correlations between continuous measures of TSH, fT4, or fT3 levels within the reference range and health status, mood, or cognition. These findings augment the limited body of literature that suggests that variations in thyroid function within the reference range do not adversely affect these neurocognitive measures in a clinically significant way. On the other hand, our preliminary findings with the IGT raise the intriguing possibility that complex "real life" decision making may be affected by small variations in thyroid function. Our results suggest future directions for research that include more sensitive and specific validated tests that encompass symptoms typically reported by patients rather than general tests of health status and mood; studies that specifically target dissatisfied patients; and studies that examine the complexities of executive function in more depth than has traditionally occurred. In the absence of further data, complaints of quality of life, mood, or cognitive decrements should not be used as the sole reason to alter thyroid hormone doses in treated hypothyroid patients in attempts to achieve lower TSH levels within the reference range.

Acknowledgments

Author Disclosure Statement

No competing financial interests exist.

References

1. Spencer CA, Hollowell JG, Kazarosyan M, Braverman LE 2007 National Health and Nutrition Examination Survey III thyroid-stimulating hormone (TSH)-thyroperoxidase antibody relationships demonstrate that TSH upper reference limits may be skewed by occult thyroid dysfunction. J Clin Endocrinol Metab **92:**4236–4240.

2. Surks MI, Hollowell JG. 2007 Age-specific distribution of serum thyrotropin and antithyroid antibodies in the US population: implications for the prevalence of subclinical hypothyroidism. J Clin Endocrinol Metab **92:**4575–4582.

3. Taylor PN, Razvi S, Pearce SH, Dayan CM. 2013 Clinical review: a review of the clinical consequences of variation in thyroid function within the reference range. J Clin Endocrinol Metab **98:**3562–3571.

4. Wahlin A, Wahlin TB, Small BJ, Bäckman L. 1998 Influences of thyroid stimulating hormone on cognitive functioning in very old age. J Gerontol B Psychol Sci Soc Sci **53:**P234–239.

5. Prinz PN, Scanlan JM, Vitaliano PP, Moe KE, Borson S, Toivola B, Merriam GR, Larsen LH, Reed HL. 1999 Thyroid hormones: positive relationships with cognition in healthy, euthyroid older men. J Gerontol A Biol Sci Med Sci **54:**M111–116.

6. Volpato S, Guralnik JM, Fried LP, Remaley AT, Cappola AR, Launer LJ. 2002 Serum thyroxine level and cognitive decline in euthyroid older women. Neurology **58:**1055–1061.

7. Van Boxtel MP, Menheere PP, Bekers O, Hogervorst E, Jolles J. 2004 Thyroid function, depressed mood, and cognitive performance in older individuals: the Maastricht Aging Study. Psychoneuroendocrinology **29:**891–898.

8. Forman-Hoffman V, Philibert RA. 2006 Lower TSH and higher T4 levels are associated with current depressive syndrome in young adults. Acta Psychiatr Scand **114:**132–139.

9. Joffe RT, Levitt AJ. 2008 Basal thyrotropin and major depression: relation to clinical variables and treatment outcome. Can J Psychiatry **53:**833–838.

10. Hogervorst E, Huppert F, Matthews FE, Brayne C. 2008 Thyroid function and cognitive decline in the MRC Cognitive Function and Ageing Study. Psychoneuroendocrinology **33:**1013–1022.

11. Panicker V, Evans J, Bjøro T, Asvold BO, Dayan CM, Bjerkeset O. 2009 A paradoxical difference in relationship between anxiety, depression and thyroid function in subjects on and not on T4: findings from the HUNT study. Clin Endocrinol (Oxf) **71:**574–580.

12. Livner A, Wahlin A, Bäckman L. 2009 Thyroid stimulating hormone and prospective memory functioning in old age. Psychoneuroendocrinology **34:**1554–1559.

13. Grigorova M and Sherwin BB. Thyroid hormones and cognitive functioning in healthy, euthyroid women: a correlational study. Horm Behav 2012;**61:**617–622.

14. Beydoun MA, Beydoun HA, Kitner-Triolo MH, Kaufman JS, Evans MK, Zonderman AB. 2013 Thyroid hormones are associated with cognitive function: moderation by sex, race, and depressive symptoms. J Clin Endocrinol Metab **98:**3470–3481.

15. Medici M, Direk N, Visser WE, Korevaar TI, Hofman A, Visser TJ, Tiemeier H, Peeters RP. 2014 Thyroid function within the normal range and the risk of depression: a population-based cohort study. J Clin Endocrinol Metab **99:**1213–1219.

16. Moon JH, Park YJ, Kim TH, Han JW, Choi SH, Lim S, Park do J, Kim KW, Jang HC. 2014 Lower-but-normal serum TSH level is associated with the development or progression of cognitive impairment in elderly: Korean Longitudinal Study on Health and Aging (KLoSHA). J Clin Endocrinol Metab **99:**424–432.

17. Beydoun MA, Beydoun HA, Rostant OS, Dore GA, Fanelli-Kuczmarski MT, Evans MK, Zonderman AB. 2015 Thyroid hormones are associated with longitudinal cognitive change in an urban adult population. Neurobiol Aging **36:**3056–3066.

18. Gussekloo J, van Exel E, de Craen AJ, Meinders AE, Frolich M, Westendorp RG. 2004 Thyroid status, disability and cognitive function, and survival in old age. JAMA **292:**2591–2599.

19. Roberts LM, Pattison H, Roalfe A, Franklyn J, Wilson S, Hobbs FD, Parle JV 2006 Is subclinical thyroid dysfunction

in the elderly associated with depression or cognitive dysfunction? Ann Intern Med **145:**573–581.

20. van de Ven AC, Netea-Maier RT, de Vegt F, Ross HA, Sweep FC, Kiemeney LA, Hermus AR, den Heijer M. 2012 Is there a relationship between fatigue perception and the serum levels of thyrotropin and free thyroxine in euthyroid subjects? Thyroid **22:**1236–1243.

21. Booth T, Deary IJ, Starr JM 2013 Thyroid stimulating hormone, free thyroxine and cognitive ability in old age: the Lothian Birth Cohort Study 1936. Psychoneuroendocrinology **38:**597–601.

22. Saravanan P, Chau WF, Roberts N, Vedhara K, Greenwood R, Dayan CM 2002 Psychological well-being in patients on 'adequate' doses of l-thyroxine: results of a large, controlled community-based questionnaire study. Clin Endocrinol (Oxf) **57:**577–585.

23. Wekking EM, Appelhof BC, Fliers E, Schene AH, Huyser J, Tijssen JG, Wiersinga WM 2005 Cognitive functioning and well-being in euthyroid patients on thyroxine replacement therapy for primary hypothyroidism. Eur J Endocrinol **153:**747–753.

24. Escobar-Morreale HF, Botella-Carretero JI, Gomez-Bueno M, Galan JM, Barrios V, Sancho J 2005 Thyroid hormone replacement therapy in primary hypothyroidism: a randomized trial comparing l-thyroxine plus liothyronine with l-thyroxine alone. Ann Intern Med **142:** 412–424.

25. Saravanan P, Visser TJ, Dayan CM 2006 Psychological well-being correlates with free thyroxine but not free 3,5,3'-triiodothyronine levels in patients on thyroid hormone replacement. J Clin Endocrinol Metab **91:**3389–3393.

26. Kramer CK, von Muhlen D, Kritz-Silverstein D 2009 Treated hypothyroidism, cognitive function, and depressed mood in old age: the Rancho Bernardo Study. Eur J Endocrinol **161:**917–921.

27. Samuels MH, Kolobova I, Smeraglio A, Peters D, Janowsky JS, Schuff KG 2014 The effects of levothyroxine replacement or suppressive therapy on health status, mood, and cognition. J Clin Endocrinol Metab **99:**843–851.

28. Walsh JP, Ward LC, Burke V, Bhagat CI, Shiels L, Henley D, Gillett MJ, Gilbert R, Tanner M, Stuckey BG 2006 Small changes in thyroxine dosage do not produce measurable changes in hypothyroid symptoms, well-being, or quality of life: results of a double-blind, randomized clinical trial. J Clin Endocrinol Metab **91:** 2624–2630.

29. Samuels MH, Schuff KG, Carlson NE, Carello P, Janowsky JS 2007 Health status, mood, and cognition in experimentally induced subclinical hypothyroidism. J Clin Endocrinol Metab **92:**2545–2551.

30. Zhu DF, Wang ZX, Zhang DR, Pan ZL, He S, Hu XP, Chen XC, Zhou JN 2006 fMRI revealed neural substrate for reversible working memory dysfunction in subclinical hypothyroidism. Brain **129:**2923–2930.

31. Sinha AK, Pickard MR, Kim KD, Ahmed MT, al Yatama F, Evans IM, Elkins RP 1994 Perturbation of thyroid hormone homeostasis in the adult and brain function. Acta Med Austriaca **21:**35–43.

32. Broedel O, Eravci M, Fuxius S, Smolarz T, Jeitner A, Grau H, Stoltenburg-Didinger G, Plueckhan H, Meinhold H,Baumgartner A 2003 Effects of hyper- and hypothyroidism on thyroid hormone concentrations in regions of the rat brain. Am J Physiol Endocrinol Metab **285:**470–480.

33. Samuels MH. 2014 Thyroid disease and cognition. Endocrinol Metab Clin N Am **43:**529–543.

34. Spreen O, Strauss EA 1998 A Compendium of Neuropsychological Tests: Administration, Norms, and Commentary. Oxford University Press, New York, pp 90–102.

35. Spreen O, Strauss EA 1998 A Compendium of Neuropsychological Tests: Administration, Norms, and Commentary. Oxford University Press, New York, pp 612–616.

36. Spreen O, Strauss EA 1998 A Compendium of Neuropsychological Tests: Administration, Norms, and Commentary. Oxford University Press, New York, pp 644–646.

37. Harvey PD, Greenberg BR, Serper MR 1989 The affective lability scales: development, reliability, and validity. J Clin Psychol **45:**786–793.

38. Lezak M, Howieson DB, Loring DW 1995 Neuropsychological Assessment. Oxford University Press, New York, pp 444–450.

39. Byrd DA, Touradji P, Tang MX, Manly JJ 2004 Cancellation test performance in African American, Hispanic, and White elderly. J Int Neuropsychol Soc **10:**401–411.

40. Lezak MD, Howieson DB, Loring DW 1995 Neuropsychological Assessment. Oxford University Press, New York, pp 371–374.

41. Singh V, Khan A 2009 Heterogeneity in choices on Iowa Gambling Task: preference for infrequent-high magnitude punishment. Mind Soc **8:**43–57.

42. Lezak MD, Howieson DB, Loring DW 1995 Neuropsychological Assessment. Oxford University Press, New York, pp 363–364.

43. Spreen O, Strauss EA 1998 A Compendium of Neuropsychological Tests: Administration, Norms, and Commentary. Oxford University Press, New York, pp 208–212.

44. Van Gorp WG, Altshuler L, Theberge DC, Mintz J 1999 Declarative and procedural memory in bipolar disorder. Biol Psychiat **46:**525–531.

45. Spreen O, Strauss E 1998. A Compendium of Neuropsychological Tests: Administration, Norms and Commentary. Oxford University Press, New York, pp 1042–1043.

46. Wartofsky L, Dickey RA 2005 The evidence for a narrower thyrotropin reference range is compelling. J Clin Endocrinol Metab **90:**5483–5488.

47. R Core Team 2015 R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available at: www.R-project.org/ (accessed February 1, 2016).

48. Pinheiro J, Bates D, DebRoy S, Sarkar D; R Core Team 2015 nlme: Linear and nonlinear mixed effects models. R package, version 3.1–120. Available at: http://CRAN.R-project.org/package=nlme (accessed February 1, 2016).

49. Højsgaard, S, Halekoh, U, Yan J 2006 The R package geepack for generalized estimating equations. J Stat Softw **15:**1–11.

50. Pollock MA, Sturrock A, Marshall K, Davidson KM, Kelly CJ, McMahon AD, McLaren EH 2001 Thyroxine treatment in patients with symptoms of hypothyroidism but thyroid function tests within the reference range: randomised double blind placebo controlled crossover trial. BMJ **323:**891–895.

51. Jonklaas J, Bianco AC, Bauer AJ, Burman KD, Cappola AR, Celi FS, Cooper DS, Kim BW, Peeters RP, Rosenthal MS, Sawka AM; American Thyroid Association Task Force on

Thyroid Hormone Replacement 2014 Guidelines for the treatment of hypothyroidism: prepared by the American Thyroid Association Task Force on thyroid hormone replacement. Thyroid **24:**1670–1751.

52. Yin JJ, Liao LM, Luo DX, Xu K, Ma SH, Wang ZX, Le HB, Huang RR, Cai ZL, Zhang J 2013 Spatial working memory impairment in subclinical hypothyroidism: an fMRI study. Neuroendocrinology **97:**260–270.

53. Samuels MH, Schuff KG, Carlson NE, Carello P, Janowsky JS 2007 Health status, psychological symptoms, mood and cognition in L-thyroxine treated hypothyroid subjects. Thyroid **17:**249–258.

54. Samuels MH, Kolobova I, Smeraglio A, Peters D, Janowsky JS, Schuff KG 2014 The effects of levothyroxine replacement or suppressive therapy on health status, mood and cognition. J Clin Endocrinol Metab **99:**843–851.

55. Roelfsema F, Veldhuis JD 2013 Thyrotropin secretion patterns in health and disease. Endocr Rev **34:**619–657.

Address correspondence to:
Mary H. Samuels, MD
Division of Endocrinology, Diabetes and Clinical Nutrition
Oregon Health & Science University
3181 Southwest Sam Jackson Park Road
Portland, OR 97239

E-mail: samuelsm@ohsu.edu