Neurobiology of Aging 36 (2015) 3056–3066

Contents lists available at ScienceDirect

Neurobiology of Aging

journal homepage: www.elsevier.com/locate/neuaging

Thyroid hormones are associated with longitudinal cognitive change in an urban adult population

May A. Beydoun [a,*], Hind A. Beydoun [b], Ola S. Rostant [a], Greg A. Dore [a], Marie T. Fanelli-Kuczmarski [c], Michele K. Evans [a,1], Alan B. Zonderman [a,1]

[a] NIH Biomedical Research Center, National Institute on Aging, NIA/NIH/IRP, Baltimore, MD, USA
[b] Graduate Program in Public Health, Eastern Virginia Medical School, Norfolk, VA, USA
[c] Department of Behavioral Health and Nutrition, University of Delaware, Newark, DE, USA

ARTICLE INFO

Article history:
Received 12 May 2015
Received in revised form 30 July 2015
Accepted 1 August 2015
Available online 8 August 2015

Keywords:
Thyroid hormones
Cognitive function
Longitudinal studies
Aging

ABSTRACT

Recent evidence indicates that thyroid hormones may be closely linked to cognition among adults. We investigated associations between thyroid hormones and longitudinal cognitive change, within and outside of reference ranges, stratifying by sex and race. This longitudinal study used data from the Healthy Aging in Neighborhoods of Diversity Across the Lifespan study, set in Baltimore City, MD, 2004–2013, on adults aged 30–64 years at baseline visit, with a length of follow-up between visits 1 and 2 ranging from <1 to 8 years; mean ± standard deviation: 4.64 ± 0.93. The final analytic sample sizes ranged from 1486 to 1602 participants with 1.6–1.7 visits per participant (total visits: 2496–2757), depending on the cognitive test. Eleven cognitive test scores spanning domains of learning or memory, language or verbal, attention, visuospatial and/or visuoconstruction, psychomotor speed, executive function, and mental status were used. Mixed-effects regression models were conducted, interacting time of follow-up with several thyroid exposures. Whites performed better than African Americans, with only 4 cognitive test scores of 11 declining significantly over time. Importantly, above reference range thyroid stimulating hormone (vs. reference range, thyroid stimulating hormone, above reference range [TSHarr]) was linked to faster rates of decline on the digits span backwards test, reflecting working memory (TSHarr × time $\gamma \pm$ standard error: -0.14 ± 0.05, $p = 0.006$) and clock-command, at test of visuospatial and/or visuoconstruction abilities (TSHarr × Time $\gamma \pm$ standard error: -0.10 ± 0.04, $p = 0.004$). The latter finding was replicated when comparing normal thyroid function to "subclinical hypothyroidism". Within-reference ranges, a higher thyroid stimulating hormone was related to faster decline on the clock-command test scores in women. In sum, higher baseline thyroid stimulating hormone was associated with faster cognitive decline over-time among urban US adults, specifically in domains of working memory and visuospatial and/or visuoconstruction abilities.

Published by Elsevier Inc.

1. Introduction

Cognitive impairment, a principal cause for functional disability among the elderly, can lead to dementing illness over time mainly in the form of Alzheimer's disease (AD). In fact, the prevalence of AD is expected to rise, reaching 100 million worldwide by 2050, with 1 in 85 persons potentially living with AD (Alzheimer's Association, 2009). Thus, it is important to uncover some of the modifiable risk factors that would prevent or delay cognitive impairment, the hallmark of AD and other dementing illnesses.

Among those modifiable factors, hormonal influence on cognition is increasingly gaining interest among researchers in the field. Altered thyroid function is well-known to co-occur with psychological and cognitive changes in adults (Samuels, 2014). However, it is uncertain which type of disordered function affects cognition, to what extent, among which subgroups, and for which domains of cognition. With advances in the neurosciences, it is now possible to use validated neurocognitive tests reflecting specific cognitive domains and mapped directly to specific brain regions (Samuels, 2014). Moreover, 4 categories of thyroid dysfunction are commonly studied in the literature, based on laboratory testing of free thyroxine (fT_4), tri-iodothyronine (T_3), and thyroid stimulating hormone (TSH) levels (Samuels, 2014). Those can be summarized as

May A Beydoun had full access to the data used in this article and completed all the statistical analyses.

* Corresponding author at: NIH Biomedical Research Center, National Institute on Aging, NIA/NIH/IRP, 251 Bayview Blvd., Suite 100, Room #: 04B118, Baltimore, MD 21224, USA. Tel.: 410 558 8648; fax: 410 558 8236.
E-mail address: baydounm@mail.nih.gov (M.A. Beydoun).
[1] Cosenior authors.

0197-4580/$ – see front matter Published by Elsevier Inc.
http://dx.doi.org/10.1016/j.neurobiolaging.2015.08.002

follow: (1) overt hypothyroidism: low-serum fT_4-coupled with elevated serum TSH; (2) overt thyrotoxicosis: high-serum fT_4 and/or T_3 and suppressed TSH level; (3) subclinical hypothyroidism (elevated TSH, normal fT_4); and (4) subclinical thyrotoxicosis (suppressed TSH, normal fT_4, and T_3) (Samuels, 2014). Despite the common use of those groupings for clinical purposes, thyroid function and dysfunction is often thought of as a continuum, thus the importance of examining effects of each of the hormonal factors separately. Some of the domains commonly affected by thyroid dysfunction include memory, executive function, and attention or concentration. Many of those cognitive deficits may be completely or partially reversed by administration of levothyroxine (L-T_4) (Bono et al., 2004).

Emerging evidence from animal studies and clinical observations suggests that thyroid hormones are crucial to a well-functioning central nervous system and that those hormones may play a role for structural and functional development of the brain early on, including brain areas that regulate mood and cognition (Koromilas et al., 2010). In fact, hypothyroidism causes a condition termed pseudodementia, a progressive nondegenerative cognitive impairment characterized by slower thought processes (Dugbartey, 1998). Studies also show that thyroid hormones continue to modulate the function of the adult brain, which explains the tight regulation of thyroid hormone transport into the brain, region-specific T_4 to T_3 conversion as well as T_3 receptor levels (Ceballos et al., 2009). Epidemiological studies indicated that thyroid dysfunction whether hypothyroidism or hyperthyroidism (overt or subclinical) increases the risk of cognitive impairment, (Beydoun et al., 2013; Bono et al., 2004; Correia et al., 2009; Miller et al., 2006; Munte et al., 2001) although the evidence is still sparse (Almeida et al., 2007; Ceresini et al., 2009; de Jongh et al., 2011; Formiga et al., 2014; Joffe et al., 2013; Kramer et al., 2009; Parle et al., 2010; Samuels et al., 2007; Wijsman et al., 2013). It is less well-known how thyroid hormone fluctuations within normal ranges can affect cognitive outcomes in the general population, particularly when studies have examined cognitive performance among middle-aged adults (Beydoun et al., 2012, 2013; Grigorova and Sherwin, 2012; van Boxtel et al., 2004).

Limited research has systematically tested the associations between thyroid hormones (both outside and within normal ranges) and cognitive change over-time in a large sample of middle-age adults. Thus, we describe the relationships between variations in thyroid hormones and longitudinal cognitive change in a large socioeconomically diverse biracial population of adult men and women. Because of the strong evidence of differential thyroid function by sex as well as by race (Aoki et al., 2007), we stratified part of the analysis by those 2 sociodemographic factors.

2. Materials and methods

2.1. Database

Initiated in 2004, the Healthy Aging in Neighborhoods of Diversity across the Life Span (HANDLS) study is an ongoing prospective cohort study that used area probability sampling to recruit a socioeconomically diverse and representative sample of African Americans and Whites (30–64 years) residing in Baltimore, Maryland (Evans et al., 2010). Written informed consent was obtained from all participants who were provided with a protocol booklet in layman's terms and a video explaining all study procedures including future recontacts. Materials' approval was completed by MedStar Institutional Review Board. The present study used longitudinal HANDLS data from baseline and the first follow-up examination (visit 2 ended in 2013). Time between examination visits 1 (wave 1) and 2 (wave 3) ranged from <1 to ~8 years, with a mean of 4.64 ± 0.93 years.

2.2. Study subjects

Initially, 3720 participants were recruited, of whom 2630 had baseline complete data on one of the measure of mental status (Mini-Mental State Examination [MMSE]). Of those, 2045 had non-missing dietary data that were used to compute the 2010-Healthy Eating Index (2010-HEI), while 2077 had complete data on CES-D total score. In addition, thyroid hormone exposures were available for ~2500 participants of whom 2296–2381 were within the reference ranges. Available and reliable cognitive data varied by cognitive test ranging from N = 2088 for California Verbal Learning Test-free delayed recall (CVLT-DFR) to 2700 for clock-command test at visit 1. At the follow-up visit, those sample sizes were reduced to a range of 1846 (CVLT-DFR) to 2139 (animal fluency). When combining waves in the final analytic models, samples of participants with complete data on outcomes at either visit, as well as exposures and covariates at baseline were reduced to a range of N = 1486–1602 with a mean repeat of 1.6–1.7 visits per participant and a total number of visits ranging from 2496 to 2757. As is discussed in further details in the Section 2.6, possible sample selectivity was corrected by using a 2-stage Heckman selection approach (Heckman, 1979).

2.3. Cognitive assessment

Cognitive assessment consisted of 7 tests with 11 test scores covering 7 domains (Mental status, attention, learning or memory, executive function, visuospatial or visuoconstruction ability, psychomotor speed, language or verbal): the MMSE, the California Verbal Learning Test immediate (list A) and delayed free recall, digit span forward and backward tests (DS-F and DS-B), the Benton visual retention test, animal fluency test (AF), Brief Test of Attention, Trails A and B, and the Clock Drawing Test (See Appendix I for full description of tests and scores). Only individual test scores were used in the analysis rather than cognitive domains. All participants were judged capable of informed consent and were probed for their understanding of the protocol. Although no formal dementia diagnoses were performed, all participants were administered mental status tests, which they completed successfully. In every case, low mental status performance was due to poor literacy skills with no other signs of dementia.

2.4. Thyroid hormone assessment

Several assays for thyroid hormone assessment were completed at Quest Diagnostics laboratories (http://www.questdiagnostics.com/home.html). First, immunochemiluminometric (ICMA) TSH assays (TSH-ICMA; ADVIA Centaur XP, Siemens) were conducted with a 0.01–0.02 mU/L sensitivity (Ross, 1988). Reference range for TSH among adults aged 20+ years is 0.4–4.5 mU/L (http://www.questdiagnostics.com/testcenter/TestDetail.action?ntc=899), with an interindividual coefficient of variation of 32%. Total thyroxine (tT_4) was measured using ICMA (AU 5400, Beckman Coulter) with a 0.8-μg/dL sensitivity and a reference range of 4.8–10.4 μg/dL (http://www.questdiagnostics.com/testcenter/TestDetail.action?ntc=17733). Measurements of free thyroxine (fT_4) concentration were also conducted using ICMA (ADVIA Centaur XP, Siemens), nondialysis, with a sensitivity of 0.1 ng/dL, and a reference range of 0.8–1.8 ng/dL (http://www.questdiagnostics.com/testcenter/TestDetail.action?ntc=866). Tri-iodothyronine (T_3) percent uptake (T_{3pu}) is used to estimate thyroxin binding globulin (TBG) availability, a protein carrying most of serum T_3 and T_4. TBG is known to have an inverse relationship with T_{3pu}, with a lower TBG (or higher

T_{3pu}) suggestive of possible hyperthyroidism or thyrotoxicosis. T_3 (% uptake) was also measured by ICMA (AU 5400, Beckman Coulter) and had a reference range of 24%–37% (Baskin et al., 2002). Using fT_4 and TSH criteria, thyroid dysfunction status was defined as follows: (1) overt hypothyroidism: low-serum fT_4 coupled with elevated serum TSH; (2) overt thyrotoxicosis: high-serum fT_4 and suppressed TSH level; (3) subclinical hypothyroidism (elevated TSH, normal fT_4); (4) subclinical thyrotoxicosis (suppressed TSH and normal fT_4); and (5) other type of dysfunction which were compared to (N) normal TSH and fT_4 levels. The distribution of reference ranges, abnormal values and thyroid dysfunction groups are presented in Table 1.

2.5. Covariates

Many variables were included in the analyses namely age, sex, self-reported race (White vs. African American), marital status, educational attainment (<high school (HS); HS, >HS), poverty income ratio (<125% for "poor"), measured body mass index (BMI, kg/m^2), current drug use ("opiates, marijuana, or cocaine" vs. none), smoking status ("current" vs. "never or former") and the Wide Range Achievement Test (WRAT) letter and word reading subtotal score to measure literacy (see Appendix I) The 20-item Center for Epidemiologic Studies Depression Scale (CES-D) scale was used to assess affective, depressed mood. Baseline CES-D total score was used in our analyses, with CES-D $\geq$ 16 labeled as "elevated depressive symptoms" (see Appendix I). Moreover, overall dietary quality as measured by the HEI-2010 based on two 24-hour recalls administered in HANDLS baseline visit was also included in the analyses, due to its potentially confounding effect between thyroid hormonal function and cognitive performance and/or decline (Beydoun et al., 2014; Fontana et al., 2006; van de Rest et al., 2015). The steps for calculating HEI-2010 are provided by the National Cancer Institute's Applied Research Web site (http://appliedresearch.cancer.gov/tools/hei/tools.html) as well as the HANDLS Web site (http://handls.nih.gov/06Coll-dataDoc.htm). Worth of noting that total and component HEI-2010 scores were calculated for each recall day (day 1 and day 2) and then averaged to obtain the mean HEI-2010 total and component scores, thus combining both days. In the present study, only total HEI-2010 score was considered. Use of antidepressants was included as a covariate in a sensitivity analysis.

2.6. Statistical analysis

Stata release 13.0 was used to conduct all analyses. First, using survey commands, we applied medical research vehicle examination sampling weights in the descriptive parts of the analysis, to obtain population estimates of means, proportions, and regression coefficients, given unequal probability of sampling from the target Baltimore city population (Lohr, 1999). Means across binary variables were compared using linear regression models that accounted for those sampling weights (svy:reg), whereas design-based F-tests were conducted to test associations between categorical variables using cross-tabulations between those variables while accounting for those same weights (svy:tab).

Second, mixed-effects linear regression models on 11 continuous cognitive test score(s) comparing above and below reference ranges to the reference range of thyroid hormones were conducted (models 1–4). Interactions by sex or race were not tested, given the expected lower statistical power for those categorical exposures, when compared to continuous exposures.

Third, similar mixed-effects regression on 11 continuous cognitive test score(s) were used to examine associations between the 4 continuous thyroid hormone exposure variables within-reference ranges (also termed models 1–4) and cognitive performance over-time, controlling for potential confounders. Moderating effects of sex and race were tested by adding interaction terms to the multivariable mixed-effects regression models (3-way interactions time $\times$ exposure $\times$ sex or time $\times$ exposure $\times$ race) and stratifying by sex or race or both, although separately, when interactions with sex and/or race are deemed significant.

Finally, thyroid dysfunction status was examined in a similar way by comparing the 4 categories of dysfunction to the normal category defined by fT_4 and TSH levels (see Section 2.4). Appendix II describes the mixed-effects regression modeling approach used in detail.

To minimize potential selection bias in mixed-effects regression models (due to the nonrandom selection of participants with complete data from the target study population), a 2-stage Heckman selection model was constructed, by running a probit model to compute an inverse mills ratio at the first stage (derived from the predicted probability of being selected, conditional on the covariates in the probit model, mainly baseline age, sex, race, poverty status, and education), as was done in an earlier study (Heckman, 1979). This inverse mills ratio was then entered as covariate in the mixed-effects regression model at the second stage, as was done in a previous study (Beydoun et al., 2013). Because of possible collinearity between the inverse mills ratio and the common covariates entered in both the mixed-effects regression model and the probit model, poverty status was eliminated from the mixed-effects regression in a sensitivity analysis. In a second sensitivity analysis, use of antidepressants was included as an additional covariate in the mixed-effects regression models.

In all analyses, a type I error of 0.05 was considered for main effects whereas a $p < 0.10$ was deemed significant for interaction terms, (Selvin, 2004), before correcting for multiple testing. A familywise Bonferroni procedure was used to correct for multiple testing by accounting only for cognitive tests and assuming that hormonal exposures related to separate substantive hypotheses (Hochberg and Tamhane, 1987). Therefore, for main effects, $p < 0.004$ (0.05/11) was considered significant. Because of their lower statistical power compared to main effects, interaction terms had their critical p-values reduced to (0.10/11 = 0.009).

3. Results

Table 1 displays baseline (visit 1) characteristics among participants with complete and dependable MMSE scores, by sex and race. Compared to men, women had lower income, education and literacy, were less likely to be married, to be current smokers or illicit drug users, and had an overall higher BMI and CES-D total score. Both mean fT_4 and T_{3pu} within the reference range were lower in women than in men, while both proportions >reference and <reference for TSH were higher in women. Racial differences were also noted for sociodemographic, lifestyle, and thyroid hormonal exposures, although no difference by race was observed in terms of CES-D scores or BMI. Specifically, compared to Whites, African Americans were less likely to be married, had lower income, education, and literacy but were more likely to be currently smoking or using illicit drugs, to have poorer quality diet, and had a lower TSH level within the reference range. African Americans were also more likely to have suboptimal TSH values (4.2% in African Americans vs. 1.2% in Whites), with the reverse being observed for above-reference range values (1.8% in African Americans vs. 5.9% in Whites). An above-reference range tT_4 was more likely in African Americans (11.2% vs. 4.6%), who were also more likely to have a suboptimal T_3 %uptake (7.1% vs. 3.3%). Thyroid function status varied by sex and race, with a significantly larger proportion of Whites fitting the "subclinical hypothyroidism" compared to African Americans. In contrast, "subclinical thyrotoxicosis" was more

Table 1
Selected baseline (visit 1) study participant characteristics by sex and race or ethnicity for HANDLS participants with complete and reliable baseline MMSE scores (n = 2630)[a]

	All	Men	Women	Whites	African Americans	p[b]	
	N = 2630	45.4 ± 1.7 (N = 1142)	54.5 ± 1.7 (N = 1488)	36.2 ± 1.5 (N = 1118)	63.8 ± 1.5 (N = 1512)	Men versus women	Whites versus African Americans
Age at baseline	47.0 ± 0.3 (N = 2630)	47.2 ± 0.4 (N = 1142)	46.9 ± 0.4 (N = 1487)	46.7 ± 0.4 (N = 1118)	47.2 ± 0.4 (N = 1512)	0.63	0.41
Married, %	35.0 ± 1.7 (N = 2447)	38.9 ± 2.5 (N = 1061)	31.8 ± 2.2 (N = 1386)	44.9 ± 2.3 (N = 1018)	29.6 ± 2.2 (N = 1429)	0.03	<0.001
Education, %							
<HS	4.3 ± 0.6	4.5 ± 0.8	4.1 ± 0.7	5.5 ± 0.9	3.7 ± 0.7	0.70	<0.001
HS	52.8 ± 1.7	54.0 ± 2.4	51.7 ± 2.3	40.2 ± 2.0	59.9 ± 2.4		
>HS	38.5 ± 1.7	36.6 ± 2.4	40.1 ± 2.3	46.6 ± 2.1	33.9 ± 2.3		
Missing	4.4 ± 0.8 (N = 2630)	4.9 ± 1.3 (N = 1142)	4.1 ± 1.0 (N = 1488)	7.7 ± 1.0 (N = 1118)	2.6 ± 1.1 (N = 1512)		
Literacy (WRAT score)	43.2 ± 0.3 (N = 2616)	43.0 ± 0.4 (N = 1136)	43.3 ± 0.3 (N = 1480)	46.7 ± 0.3 (N = 1114)	41.2 ± 0.3 (N = 1502)	0.48	<0.001
PIR <125, %	19.6 ± 1.0 (N = 2630)	16.4 ± 1.0 (N = 1142)	22.2 ± 1.5 (N = 1488)	12.3 ± 0.8 (N = 1118)	23.7 ± 1.5 (N = 1512)	0.003	<0.001
Current smoking status, %							
Currently smoking	43.7 ± 1.7	49.7 ± 2.4	38.7 ± 2.3	36.0 ± 2.0	46.3 ± 2.3	0.007	<0.001
Missing	4.9 ± 1.7 (N = 2630)	3.7 ± 1.0 (N = 1142)	6.0 ± 1.3 (N = 1488)	3.6 ± 0.7 (N = 1118)	5.7 ± 1.2 (N = 1512)		
Current use of illicit drugs, %							
Used any type	61.8 ± 1.6	72.3 ± 2.0	53.0 ± 2.3	55.5 ± 2.1	65.3 ± 2.2	<0.001	<0.001
Missing	7.8 ± 0.8 (N = 2630)	6.6 ± 1.0 (N = 1142)	8.7 ± 2.3 (N = 1488)	11.0 ± 1.3 (N = 1117)	6.0 ± 1.0 (N = 1511)		
Body mass index, kg m^{-2}	29.7 ± 0.3 (N = 2630)	28.1 ± 0.3 (N = 1142)	31.1 ± 0.4 (N = 1488)	29.2 ± 0.3 (N = 1118)	30.0 ± 0.4 (N = 1512)	<0.001	0.10
HEI-2010 total score	43.8 ± 0.4 (N = 2045)	43.0 ± 0.5 (N = 875)	44.4 ± 0.6 (N = 1170)	45.1 ± 0.6 (N = 865)	43.0 ± 0.5 (N = 1180)	0.07	0.008
Depressive symptoms							
CES-D score	10.5 ± 0.3	9.6 ± 0.3	11.2 ± 0.4	9.9 ± 0.4	10.7 ± 0.4	<0.001	0.96
CES-D score ≥16 (EDS), %	22.1 ± 1.5 (N = 2077)	17.8 ± 2.1 (N = 892)	25.6 ± 2.2 (N = 1187)	23.5 ± 2.1 (N = 823)	21.5 ± 2.0 (N = 1255)	0.010	0.48
Antidepressant use, %	12.3 ± 0.1 (N = 2399)	8.1 ± 0.1 (N = 1045)	15.7 ± 0.2 (N = 1354)	18.2 ± 0.2 (N = 1015)	9.1 ± 0.1 (N = 1384)	<0.001	<0.001
Thyroid hormones, within reference range[c]							
TSH, mU/L	1.73 ± 0.03 (N = 2296)	1.68 ± 0.04 (N = 1018)	1.78 ± 0.05 (N = 1278)	1.90 ± 0.04 (N = 986)	1.63 ± 0.05 (N = 1310)	0.12	<0.001
Free T_4, μg/dL	1.13 ± 0.01 (N = 2381)	1.14 ± 0.01 (N = 1030)	1.12 ± 0.01 (N = 1351)	1.14 ± 0.01 (N = 1048)	1.12 ± 0.01 (N = 1333)	0.038	0.07
Total T_4, ng/dL	7.53 ± 0.04 (N = 2232)	7.51 ± 0.07 (N = 963)	7.55 ± 0.05 (N = 1269)	7.57 ± 0.05 (N = 995)	7.51 ± 0.06 (N = 1237)	0.63	0.52
T_3, %uptake	30.50 ± 0.10 (N = 2310)	30.94 ± 0.16 (N = 984)	30.13 ± 0.13 (N = 1326)	30.52 ± 0.12 (N = 1042)	30.48 ± 0.15 (N = 1268)	<0.001	0.85
Thyroid hormones, above or below reference range							
TSH, mU/L							
<0.4	3.0 ± 0.5	2.7 ± 0.7	3.3 ± 0.7	1.2 ± 0.4	4.2 ± 0.8	0.013	<0.001
>4.5	3.4 ± 0.4 (N = 2497)	1.9 ± 0.4 (N = 1079)	4.6 ± 0.7 (N = 1418)	5.9 ± 0.8 (N = 1089)	1.8 ± 0.4 (N = 1408)		
Free T_4, μg/dL							
<0.8	4.6 ± 0.6	5.5 ± 1.2	3.9 ± 0.7	5.0 ± 0.9	4.4 ± 0.9	0.12	0.78
>1.8	0.3 ± 0.1 (N = 2502)	0.0 ± 0.0 (N = 1082)	0.4 ± 0.1 (N = 1420)	0.2 ± 0.1 (N = 1089)	0.3 ± 0.1 (N = 1413)		
Total T_4, ng/dL							
<4.8	3.3 ± 0.6	4.2 ± 1.1	2.5 ± 0.7	3.7 ± 0.9	3.0 ± 0.8	0.042	0.002
>10.4	8.7 ± 1.1 (N = 2504)	6.2 ± 1.2 (N = 1082)	10.8 ± 1.8 (N = 1422)	4.6 ± 0.9 (N = 1089)	11.2 ± 1.7 (N = 1415)		
T_3, %uptake							
<24%	5.6 ± 0.8	5.2 ± 1.0	6.0 ± 1.3	3.3 ± 1.0	7.1 ± 1.3	0.09	0.026
>37%	1.7 ± 0.4 (N = 2504)	2.6 ± 0.7 (N = 1082)	0.9 ± 0.3 (N = 1422)	1.7 ± 0.7 (N = 1089)	1.7 ± 0.4 (N = 1415)		
Thyroid function categories							
Normal	89.8 ± 0.9	90.7 ± 1.4	89.0 ± 1.2	89.1 ± 1.2	90.2 ± 1.2	0.017	<0.001
Overt hypothyroidism	0.7 ± 0.2	0.5 ± 0.2	0.8 ± 0.2	1.1 ± 0.3	0.5 ± 0.2		
Overt thyrotoxicosis	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.0 ± 0.0	0.1 ± 0.1		
Subclinical hypothyroidism	2.7 ± 0.4	1.4 ± 0.3	3.8 ± 0.7	4.9 ± 0.8	1.4 ± 0.4		
Subclinical thyrotoxicosis	2.8 ± 0.5	2.5 ± 0.7	3.2 ± 0.7	1.1 ± 0.4	3.9 ± 0.8		
Other thyroid dysfunction	3.9 ± 0.6 (N = 2497)	4.9 ± 1.1 (N = 1079)	3.1 ± 0.6 (N = 1418)	3.9 ± 0.9 (N = 1089)	4.0 ± 0.8 (N = 1408)		

Key: CES-D, Center for Epidemiologic Studies-Depression; EDS, elevated depressive symptoms; HANDLS, Healthy Aging in Neighborhoods of Diversity across the Lifespan; HS, high school; MMSE, Mini-Mental State Examination; PIR, poverty income ratio; SEP, standard error of the percentage; T_3, tri-iodothyronine; T_4, thyroxine; TSH, thyroid stimulating hormone; WRAT, Wide Range Achievement Test.

[a] Values are weighted mean ± standard error of the mean or percent ± SEP.

[b] p-Value was based on linear regression models when row variable is continuous (svy:reg) and design-based F-test when row variable is categorical (svy:tab).

[c] TSH, free T_4 (fT_4), and total T_4 (tT_4) values outside the reference range were excluded in this analysis (See Section 2 for reference ranges).

prevalent in African Americans. Both types of dysfunctions were more prevalent in women than in men.

Table 2 shows that in addition to persistent racial differences in cognitive performance across the 2 visits with poorer performance found in African Americans, only 4 of 11 cognitive tests changed between visits, with consistent indication of over-time cognitive decline in 3 of the 4. In particular, verbal and visual memory scores declined in both sexes and racial groups, while MMSE scores reflecting mental status improved over time possibly due to learning, particularly among Whites.

Table 3 displays associations between the 4 thyroid hormone exposures (comparing suboptimal and above-reference levels to within-reference range [ARRVRR]) and longitudinal cognitive change in 4 separate models, based on multiple mixed-effects regression analyses. After correction for multiple testing (type I error corrected to 0.004 for main effects), suboptimal tT_4 was associated with better performance in AF (model 3, below reference range vs. reference, $\gamma \pm$ standard error of the estimate [SEE]: $+2.08 \pm 0.70$, $p = 0.003$) at baseline. When examining cognitive change (type I error corrected to 0.009), none of the associations survived multiple testing correction. However, when comparing participants above-reference ranges to those within (model 1, ARRVRR × time), above-reference range TSH was linked to faster rates of decline on DS-B, a test of working memory ($\gamma \pm$ SEE: -0.14 ± 0.05, $p = 0.006$) and clock-command, at test of visuospatial and visuoconstruction abilities ($\gamma \pm$ SEE: -0.10 ± 0.04, $p = 0.004$).

Fig. 1A and B show predictive margins from 2 mixed-effects regression models whereby the outcomes were DS-B and clock-command test scores and the key predictor was thyroid function status, controlling for the same covariates as in Table 3. In both models, "sub-clinical hypothyroidism" (category C, see Section 2) compared to the "normal" thyroid function category (thyroid_st_CN) was linked to a faster rate of cognitive decline over-time ($p < 0.009$ for time × thyroid_st_CN). In particular, subclinical hypothyroidism was associated with 14%–15% poorer cognitive performance on DS-B after 5 years compared to baseline and ~7% poorer performance on clock-command compared to baseline. The corresponding decline for "normal" thyroid function was <1% in both cases.

Table 2
Cognitive performance test scores at visits 1 and 2, by sex and race or ethnicity for HANDLS participants with complete and reliable baseline MMSE scores[a]

	All	Men	Women	Whites	African Americans
Mini-Mental State Exam, total score					
Visit 1	27.83 ± 0.07 (N = 2630)	27.71 ± 0.10 (N = 1142)	27.94 ± 0.09 (N = 1488)	28.43 ± 0.07 (N = 1118)	27.50 ± 0.10[b] (N = 1512)
Visit 2	28.04 ± 0.06 (N = 1934)	27.96 ± 0.10 (N = 775)	28.10 ± 0.08 (N = 1159)	28.65 ± 0.06 (N = 767)	27.70 ± 0.09[b] (N = 1167)
p (visit 2–visit 1)	0.028	0.08	0.19	0.022	0.12
California Verbal Learning Test, list A					
Visit 1	24.99 ± 0.26 (N = 2172)	23.53 ± 0.39 (N = 939)	26.26 ± 0.34[c] (N = 1233)	26.96 ± 0.36 (N = 895)	23.93 ± 0.35[b] (N = 1277)
Visit 2	20.08 ± 0.26 (N = 1976)	18.73 ± 0.37 (N = 817)	21.12 ± 0.37[c] (N = 1159)	22.55 ± 0.40 (N = 781)	18.72 ± 0.33[b] (N = 1195)
p (visit 2–visit 1)	<0.001	<0.001	<0.001	<0.001	<0.001
CVLT, free delayed recall					
Visit 1	7.34 ± 0.12 (N = 2088)	6.83 ± 0.17 (N = 900)	7.78 ± 0.16[c] (N = 1188)	8.36 ± 0.16 (N = 863)	6.79 ± 0.15[b] (N = 1225)
Visit 2	5.82 ± 0.13 (N = 1846)	5.34 ± 0.19 (N = 759)	6.18 ± 0.17[c] (N = 1087)	7.20 ± 0.20 (N = 719)	5.09 ± 0.15[b] (N = 1127)
p (visit 2–visit 1)	<0.001	<0.001	<0.001	<0.001	<0.001
Benton visual retention test					
Visit 1	5.66 ± 0.16 (N = 2594)	5.20 ± 0.23 (N = 1129)	6.05 ± 0.23[c] (N = 1465)	5.01 ± 0.18 (N = 1108)	6.03 ± 0.23[b] (N = 1486)
Visit 2	7.65 ± 0.18 (N = 2085)	7.20 ± 0.26 (N = 861)	7.99 ± 0.24[c] (N = 1224)	6.24 ± 0.22 (N = 816)	8.42 ± 0.25[b] (N = 1269)
p (visit 2–visit 1)	<0.001	<0.001	<0.001	<0.001	<0.001
Brief Test of Attention					
Visit 1	6.72 ± 0.08 (N = 2247)	6.57 ± 0.11 (N = 980)	6.84 ± 0.12 (N = 1267)	7.47 ± 0.09 (N = 930)	6.30 ± 0.12[b] (N = 1317)
Visit 2	6.64 ± 0.09 (N = 1907)	6.54 ± 0.14 (N = 789)	6.72 ± 0.12 (N = 1118)	7.21 ± 0.10 (N = 772)	6.33 ± 0.13[b] (N = 1135)
p (visit 2–visit 1)	0.55	0.89	0.48	0.06	0.84
Animal fluency					
Visit 1	19.19 ± 0.20 (N = 2695)	19.79 ± 0.29 (N = 1177)	18.68 ± 0.27[c] (N = 1518)	21.25 ± 0.30 (N = 1136)	18.02 ± 0.25[b] (N = 1559)
Visit 2	19.46 ± 0.24 (N = 2139)	20.06 ± 0.38 (N = 895)	18.99 ± 0.30[c] (N = 1244)	21.66 ± 0.32 (N = 838)	18.28 ± 0.31[b] (N = 1300)
p (visit 2–visit 1)	0.38	0.57	0.45	0.35	0.52
Digits span, forward					
Visit 1	7.42 ± 0.07 (N = 2579)	7.48 ± 0.11 (N = 1127)	7.37 ± 0.10 (N = 1452)	8.06 ± 0.10 (N = 1092)	7.06 ± 0.10[b] (N = 1487)
Visit 2	7.50 ± 0.09 (N = 1971)	7.55 ± 0.14 (N = 829)	7.46 ± 0.12 (N = 1142)	8.24 ± 0.12 (N = 760)	7.10 ± 0.12[b] (N = 1211)
p (visit 2–visit 1)	0.52	0.71	0.59	0.24	0.79
Digits span, backward					
Visit 1	5.79 ± 0.07 (N = 2561)	5.78 ± 0.11 (N = 1121)	5.79 ± 0.10 (N = 1440)	6.69 ± 0.10 (N = 1091)	5.26 ± 0.10[b] (N = 1470)
Visit 2	5.78 ± 0.08 (N = 1965)	5.75 ± 0.12 (N = 824)	5.80 ± 0.10 (N = 1141)	6.70 ± 0.12 (N = 755)	5.30 ± 0.10[b] (N = 1210)
p (visit 2–visit 1)	0.96	0.87	0.93	0.93	0.80
Clock-command					
Visit 1	8.79 ± 0.04 (N = 2700)	8.87 ± 0.06 (N = 1179)	8.73 ± 0.06 (N = 1521)	9.04 ± 0.05 (N = 1144)	8.65 ± 0.06[b] (N = 1556)
Visit 2	8.78 ± 0.05 (N = 2104)	8.82 ± 0.06 (N = 873)	8.75 ± 0.07 (N = 1231)	9.04 ± 0.05 (N = 829)	8.64 ± 0.07[b] (N = 1275)
p (visit 2–visit 1)	0.87	0.61	0.80	0.97	0.90
Trail Making Test, part A					
Visit 1	34.86 ± 0.59 (N = 2557)	34.97 ± 0.70 (N = 1096)	34.77 ± 0.91 (N = 1461)	29.58 ± 0.50 (N = 1094)	37.94 ± 0.89[b] (N = 1463)
Visit 2	36.48 ± 1.39 (N = 1874)	37.29 ± 1.64 (N = 763)	35.88 ± 2.10 (N = 1111)	29.89 ± 0.71 (N = 774)	40.06 ± 2.11[b] (N = 1100)
p (visit 2–visit 1)	0.28	0.19	0.63	0.72	0.36
Trail Making Test, part B					
Visit 1	138.77 ± 4.57 (N = 2556)	143.11 ± 7.55 (N = 1096)	135.22 ± 5.55 (N = 1460)	92.56 ± 3.80 (N = 1094)	165.71 ± 6.77[b] (N = 1462)
Visit 2	127.87 ± 5.79 (N = 1728)	130.30 ± 8.89 (N = 705)	126.06 ± 7.64 (N = 1023)	77.22 ± 2.30 (N = 724)	155.91 ± 8.59[b] (N = 1004)
p (visit 2–visit 1)	0.14	0.27	0.33	0.001	0.37

Key: HANDLS, Healthy, Aging in Neighborhoods of Diversity across the Life Span; MMSE, Mini-Mental State Examination.
[a] Most cognitive test scores were in the direction of higher score = better performance, except for BVRT (total errors), and Trail Making Test both parts (expressed in seconds).
[b] $p < 0.05$ for null hypothesis of no difference in means of cognitive test scores by race within each visit, Wald test from svy:reg command.
[c] $p < 0.05$ for null hypothesis of no difference in means of cognitive test scores by sex within each visit, Wald test from svy:reg command.

Table 3
Longitudinal cognitive change by thyroid hormonal status: mixed-effects linear regression models[a,b]

	Intercept		Time		Below reference range versus reference (BRRVRR)		(BRRVRR) × time		Above reference range versus reference (ARRVRR)		(ARRVRR) × time	
	γ ± SEE	*p*	γ ± SEE	*p*	γ ± SEE	*p*	γ ± SEE	*p*	γ ± SEE	*p*	γ ± SEE	*p*
Mini-Mental State Exam, total score												
Model 1: TSH (N = 1580; N' = 2592)	**26.8 ± 0.2**	**<0.001**	+0.09 ± 0.05	0.08	−0.05 ± 0.25	0.83	−0.01 ± 0.06	0.83	+0.26 ± 0.19	0.18	−0.03 ± 0.05	0.62
Model 2: fT_4 (N = 1583; N' = 2597)	**26.9 ± 0.2**	**<0.001**	+0.09 ± 0.05	0.10	+0.05 ± 0.20	0.78	**−0.10 ± 0.06**	**0.08**	−1.02 ± 0.64	0.11	−0.14 ± 0.16	0.38
Model 3: tT_4 (N = 1585; N' = 2598)	**26.9 ± 0.2**	**<0.001**	+0.09 ± 0.05	0.10	−0.11 ± 0.25	0.66	−0.00 ± 0.07	0.97	−0.22 ± 0.16	0.16	−0.00 ± 0.04	0.91
Model 4: T_{3pu} (N = 1585; N' = 2598)	**26.9 ± 0.2**	**<0.001**	+0.09 ± 0.05	0.10	−0.30 ± 0.18	0.10	+0.01 ± 0.05	0.92	+0.05 ± 0.31	0.87	−0.07 ± 0.08	0.39
California Verbal Learning, list A												
Model 1: TSH (N = 1515; N' = 2376)	**25.8 ± 0.7**	**<0.001**	**−1.28 ± 0.16**	**<0.001**	−1.41 ± 0.92	0.13	+0.26 ± 0.21	0.21	−0.56 ± 0.70	0.43	+0.12 ± 0.17	0.47
Model 2: fT_4 (N = 1518; N' = 2382)	**25.7 ± 0.7**	**<0.001**	**−1.25 ± 0.16**	**<0.001**	−1.22 ± 0.69	0.08	+0.06 ± 0.18	0.72	−2.91 ± 2.35	0.22	−0.71 ± 0.58	0.22
Model 3: tT_4 (N = 1518; N' = 2381)	**25.7 ± 0.7**	**<0.001**	**−1.25 ± 0.16**	**<0.001**	+0.30 ± 0.91	0.75	+0.01 ± 0.22	0.97	+0.10 ± 0.58	0.86	−0.11 ± 0.14	0.41
Model 4: T_{3pu} (N = 1518; N' = 2381)	**25.7 ± 0.7**	**<0.001**	**−1.25 ± 0.16**	**<0.001**	−0.71 ± 0.65	0.27	−0.17 ± 0.16	0.29	+0.69 ± 1.19	0.56	+0.06 ± 0.27	0.56
CVLT, free delayed recall												
Model 1: TSH (N = 1486; N' = 2275)	**7.96 ± 0.33**	**<0.001**	**−0.40 ± 0.08**	**<0.001**	+0.40 ± 0.44	0.36	−0.06 ± 0.11	0.60	+0.11 ± 0.33	0.73	+0.03 ± 0.08	0.69
Model 2: fT_4 (N = 1489; N' = 2281)	**7.96 ± 0.33**	**<0.001**	**−0.40 ± 0.08**	**<0.001**	**−0.87 ± 0.32**	**0.007**	+0.12 ± 0.09	0.18	−0.88 ± 1.11	0.42	−0.34 ± 0.29	0.25
Model 3: tT_4 (N = 1489; N' = 2280)	**7.95 ± 0.33**	**<0.001**	**−0.40 ± 0.08**	**<0.001**	−0.27 ± 0.43	0.52	+0.12 ± 0.11	0.27	+0.27 ± 0.28	0.34	−0.12 ± 0.07	0.07
Model 4: T_{3pu} (N = 1489; N' = 2280)	**7.96 ± 0.33**	**<0.001**	**−0.39 ± 0.08**	**<0.001**	−0.09 ± 0.31	0.55	−0.09 ± 0.08	0.26	+0.20 ± 0.56	0.72	+0.01 ± 0.13	0.90
Benton visual retention test												
Model 1: TSH (N = 1594; N' = 2674)	**8.90 ± 0.52**	**<0.001**	**+0.40 ± 0.13**	**<0.001**	−0.73 ± 0.69	0.29	+0.21 ± 0.16	0.18	+0.57 ± 0.53	0.28	−0.14 ± 0.13	0.29
Model 2: fT_4 (N = 1597; N' = 2680)	**8.94 ± 0.51**	**<0.001**	**+0.38 ± 0.13**	**0.003**	+0.08 ± 0.53	0.88	−0.08 ± 0.14	0.57	+1.08 ± 1.72	0.53	−0.15 ± 0.41	0.72
Model 3: tT_4 (N = 1599; N' = 2682)	**8.93 ± 0.51**	**<0.001**	**+0.37 ± 0.13**	**0.003**	−0.10 ± 0.66	0.89	−0.08 ± 0.17	0.64	+0.30 ± 0.42	0.49	+0.14 ± 0.10	0.18
Model 4: T_{3pu} (N = 1599; N' = 2682)	**8.89 ± 0.52**	**<0.001**	**+0.37 ± 0.13**	**0.003**	+0.55 ± 0.49	0.27	+0.13 ± 0.12	0.30	+0.59 ± 0.84	0.48	−0.03 ± 0.21	0.87
Brief Test of Attention												
Model 1: TSH (N = 1546; N' = 2496)	**6.48 ± 0.24**	**<0.001**	−0.09 ± 0.06	0.12	−0.07 ± 0.32	0.82	+0.06 ± 0.06	0.57	**−0.50 ± 0.24**	**0.034**	+0.06 ± 0.06	0.30
Model 2: fT_4 (N = 1549; N' = 2502)	**6.42 ± 0.24**	**<0.001**	−0.08 ± 0.06	0.17	**−0.56 ± 0.24**	**0.018**	+0.06 ± 0.07	0.35	+0.01 ± 0.80	0.99	−0.05 ± 0.22	0.80
Model 3: tT_4 (N = 1549; N' = 2501)	**6.43 ± 0.24**	**<0.001**	−0.09 ± 0.06	0.14	**−0.77 ± 0.31**	**0.013**	**+0.19 ± 0.08**	**0.023**	−0.03 ± 0.19	0.88	+0.08 ± 0.05	0.09
Model 4: T_{3pu} (N = 1549; N' = 2501)	**6.42 ± 0.24**	**<0.001**	−0.08 ± 0.06	0.19	+0.09 ± 0.23	0.68	−0.05 ± 0.06	0.38	−0.14 ± 0.38	0.71	−0.02 ± 0.09	0.80
Animal fluency												
Model 1: TSH (N = 1599; N' = 2749)	**+17.4 ± 0.6**	**<0.001**	−0.08 ± 0.12	0.51	−0.03 ± 0.73	0.96	−0.24 ± 0.04	0.10	+0.74 ± 0.55	0.18	−0.02 ± 0.12	0.86
Model 2: fT_4 (N = 1602; N' = 2755)	**+17.4 ± 0.6**	**<0.001**	−0.08 ± 0.12	0.47	+0.11 ± 0.56	0.85	−0.19 ± 0.12	0.14	−1.86 ± 1.83	0.31	−0.37 ± 0.38	0.33
Model 3: tT_4 (N = 1604; N' = 2757)	**+17.4 ± 0.6**	**<0.001**	−0.08 ± 0.12	0.48	**+2.08 ± 0.70**	**0.003**	−0.13 ± 0.16	0.41	+0.53 ± 0.44	0.23	−0.05 ± 0.09	0.61
Model 4: T_{3pu} (N = 1604; N' = 2757)	**+17.5 ± 0.6**	**<0.001**	−0.09 ± 0.12	0.43	−0.64 ± 0.51	0.21	+0.09 ± 0.11	0.41	+0.03 ± 0.88	0.98	−0.23 ± 0.18	0.22
Digits span, forward												
Model 1: TSH (N = 1594; N' = 2627)	**+6.81 ± 0.22**	**<0.001**	+0.02 ± 0.05	0.71	−0.13 ± 0.29	0.65	−0.01 ± 0.06	0.91	+0.08 ± 0.22	0.70	+0.01 ± 0.05	0.77
Model 2: fT_4 (N = 1597; N' = 2632)	**+6.81 ± 0.22**	**<0.001**	+0.02 ± 0.05	0.68	−0.42 ± 0.22	0.06	+0.01 ± 0.05	0.80	−1.05 ± 0.76	0.17	−0.15 ± 0.19	0.41
Model 3: tT_4 (N = 1598; N' = 2632)	**+6.82 ± 0.22**	**<0.001**	+0.02 ± 0.05	0.71	−0.01 ± 0.28	0.98	−0.06 ± 0.07	0.36	−0.06 ± 0.18	0.75	+0.01 ± 0.04	0.80
Model 4: T_{3pu} (N = 1598; N' = 2632)	**+6.85 ± 0.22**	**<0.001**	+0.02 ± 0.05	0.76	−0.35 ± 0.21	0.09	+0.05 ± 0.05	0.28	+0.55 ± 0.35	0.11	−0.11 ± 0.08	0.13
Digits span, backward												
Model 1: TSH (N = 1593; N' = 2612)	+1.73 ± 4.59	0.71	+0.97 ± 1.08	0.37	+0.24 ± 0.27	0.38	**−0.13 ± 0.06**	**0.038**	−0.07 ± 0.21	0.74	**−0.14 ± 0.05**	**0.006**
Model 2: fT_4 (N = 1596; N' = 2617)	+1.87 ± 4.59	0.68	+0.98 ± 1.09	0.37	−0.37 ± 0.21	0.07	+0.05 ± 0.05	0.36	+0.42 ± 0.72	0.56	−0.20 ± 0.19	0.28
Model 3: tT_4 (N = 1597; N' = 2617)	+1.20 ± 4.60	0.26	+1.12 ± 1.09	0.31	−0.25 ± 0.27	0.36	−0.01 ± 0.07	0.92	+0.03 ± 0.17	0.85	**−0.08 ± 0.04**	**0.06**
Model 4: T_{3pu} (N = 1597; N' = 2617)	+1.37 ± 4.58	0.77	+0.88 ± 1.08	0.42	−0.33 ± 0.19	0.09	+0.01 ± 0.05	0.83	−0.05 ± 0.33	0.89	**+0.18 ± 0.08**	**0.018**
Clock, command												
Model 1: TSH (N = 1597; N' = 2745)	**+8.82 ± 0.13**	**<0.001**	**−0.08 ± 0.04**	**0.031**	+0.22 ± 0.18	0.22	−0.01 ± 0.05	0.87	+0.07 ± 0.13	0.61	**−0.10 ± 0.04**	**0.004**
Model 2: fT_4 (N = 1600; N' = 2751)	**+8.82 ± 0.13**	**<0.001**	**−0.09 ± 0.04**	**0.012**	+0.01 ± 0.14	0.94	+0.01 ± 0.04	0.72	+0.39 ± 0.44	0.37	−0.03 ± 0.12	0.80

(*continued on next page*)

Table 3 (*continued*)

	Intercept		Time		Below reference range versus reference (BRRVRR)		(BRRVRR) × time		Above reference range versus reference (ARRVRR)		(ARRVRR) × time	
	γ ± SEE	*p*	γ ± SEE	*p*	γ ± SEE	*p*	γ ± SEE	*p*	γ ± SEE	*p*	γ ± SEE	*p*
Model 3: fT_4 (N = 1602; N' = 2753)	**+8.81 ± 0.13**	**<0.001**	**−0.09 ± 0.04**	**0.013**	−0.20 ± 0.17	0.23	**+0.12 ± 0.05**	**0.015**	−0.02 ± 0.11	0.87	+0.01 ± 0.03	0.85
Model 4: T_{3pu} (N = 1602; N' = 2753)	**+8.82 ± 0.12**	**<0.001**	**−0.09 ± 0.04**	**0.016**	−0.03 ± 0.12	0.83	−0.05 ± 0.04	0.20	−0.24 ± 0.21	0.25	+0.02 ± 0.06	0.77
Trail Making Test, part A												
Model 1: TSH (N = 1563; N' = 2639)	**+35.1 ± 3.9**	**<0.001**	**+2.22 ± 1.13**	**0.049**	−6.16 ± 5.03	0.22	+1.26 ± 1.41	0.37	−3.81 ± 3.86	0.32	+0.24 ± 1.13	0.84
Model 2: fT_4 (N = 1566; N' = 2645)	**+34.70 ± 3.9**	**<0.001**	**+2.27 ± 1.12**	**0.044**	−1.71 ± 3.87	0.66	+0.81 ± 1.22	0.51	−8.50 ± 12.3	0.49	+0.09 ± 3.61	0.98
Model 3: fT_4 (N = 1568; N' = 2648)	**+35.00 ± 3.86**	**<0.001**	**+2.24 ± 1.12**	**0.045**	+0.20 ± 4.90	0.97	**+2.73 ± 1.53**	**0.07**	−4.45 ± 3.09	0.15	+0.30 ± 0.91	0.74
Model 4: T_{3pu} (N = 1568; N' = 2648)	**+35.03 ± 3.88**	**<0.001**	**+2.27 ± 1.13**	**0.044**	−3.11 ± 3.56	0.38	−0.62 ± 1.07	0.56	−4.63 ± 5.92	0.43	+0.05 ± 1.77	0.98
Trail Making Test, part B												
Model 1: TSH (N = 1551; N' = 2546)	**+208.2 ± 52.7**	**<0.001**	+2.82 ± 12.4	0.82	+38.0 ± 20.2	0.06	+0.14 ± 4.03	0.97	−11.1 ± 15.5	0.47	−0.46 ± 3.35	0.89
Model 2: fT_4 (N = 1554; N' = 2551)	**+205.2 ± 52.9**	**<0.001**	+2.11 ± 12.4	0.87	−13.0 ± 15.5	0.40	+2.53 ± 3.42	0.46	+16.4 ± 49.9	0.74	−2.22 ± 11.49	0.85
Model 3: fT_4 (N = 1556; N' = 2554)	**+206.8 ± 52.9**	**<0.001**	+3.45 ± 12.4	0.78	−15.7 ± 19.6	0.42	+0.32 ± 4.46	0.94	+8.9 ± 15.5	0.47	+2.59 ± 2.66	0.33
Model 4: T_{3pu} (N = 1556; N' = 2554)	**+202.2 ± 53.0**	**<0.001**	+3.14 ± 12.4	0.80	+9.34 ± 14.4	0.52	−1.83 ± 3.13	0.56	+36.2 ± 24.1	0.13	−3.96 ± 5.50	0.47

Key: ARRVRR, above reference range versus reference range; BRRVRR, below reference range versus reference range; BVRT, Benton Visual Retention Test; CES-D, Center for Epidemiologic Studies-Depression; fT_4, free thyroxine; HANDLS, Healthy Aging in Neighborhoods of Diversity across the Life Span; HEI, Healthy Eating Index; MMSE, Mini-Mental State Examination; N, number of participants; N', number, of visits; SEE, standard error of the estimate; T_{3pu}, %uptake of tri-iodothyronine; TSH, thyroid stimulating hormone; fT_4, total thyroxine; WRAT, Wide Range Achievement Test.

The numbers are bolded to highlight *p*-values that were <0.10 for interaction terms and <0.05 for main effects.

[a] Multiple mixed-effects linear regression models adjusted for baseline age, sex, race or ethnicity, marital status, education, WRAT total score, poverty income ratio, current smoking status, current use of illicit drugs, body mass index, and 2010-HEI.

[b] Most cognitive test scores were in the direction of higher score = better performance, except for BVRT (total errors) and Trail Making Test both parts (expressed in seconds).

Within-reference range (Table 4), the higher the TSH level, a faster rate of decline was noted in clock-command scores among women (model 1: TSH × time γ ± SEE: −0.03 ± 0.01, $p = 0.008$; $p = 0.009$ for 3-way interaction TSH × time × male). Although other statistically significant 3-way interactions were detected, none of the stratum-specific effects survived correction for multiple testing. Despite the positive relationship between suboptimal fT_4 and AF test scores at baseline (Table 3) within-reference ranges, both higher fT_4 and tT_4 levels were marginally but positively related to AF at baseline (models 2–3, thyroid hormone within-reference range, $0.004 < p < 0.05$). T_{3pu} (model 4) was not associated with cognitive performance or decline outside or within-reference ranges.

In a sensitivity analysis, poverty status was removed from the main mixed-effects regression model, allowing it to be an instrumental variable to compute the inverse mills ratio. Key results were not altered. A second sensitivity analysis in which antidepressant use was included as an additional covariate in the models indicated that antidepressant use was not an important confounder in the relationship between thyroid hormones, particularly within normal ranges, and cognitive performance or decline (data not shown).

4. Discussion

The present study examined associations between thyroid hormones (within and outside normal ranges) and over-time longitudinal change in cognitive performance among middle-aged US adults, using several domains of cognition and stratifying by sex and race. Several key findings emerged. Whites performed consistently better than African-American on all cognitive tests, with only tests of mental status (MMSE), verbal memory (California Verbal Learning Test-List A and delayed free recall), and visuomotor and/or visuoconstructional abilities (Benton visual retention test) declining significantly over-time. Importantly, when examining cognitive change (type I error corrected to 0.009) in relation to below reference range versus within-reference range hormonal status, none of the associations survived multiple testing correction. However, when comparing participants ARRVRR, above-reference range TSH was linked to faster rates of decline on DS-B, a test of working memory ($p = 0.006$), and clock-command, at test of visuospatial and visuoconstruction abilities ($p = 0.004$). This finding was replicated when comparing normal thyroid function to "subclinical hypothyroidism". Within-reference ranges, the higher the TSH level, the faster was the rate of decline on the clock-command test scores in women.

Our previous cross-sectional analysis of HANDLS data (Beydoun et al., 2013) uncovered stratum-specific associations between thyroid hormones within normal ranges and cognitive performance which were not thoroughly reported in our present study. Moreover, 2 cognitive test scores (card rotation and identical pictures) were not measured at follow-up visits, thus precluding longitudinal analyses. Although a similar trend was detected in cross-sectional results, most of these associations at baseline did not pass correction for multiple testing, given the slightly different samples selected between studies (Beydoun et al., 2013).

At least 9 previous cohort studies examined longitudinal relationships between thyroid hormones and cognitive performance (Booth et al., 2013; de Jong et al., 2006, 2009; de Jongh et al., 2011; Forti et al., 2012; Gussekloo et al., 2004; Hogervorst et al., 2008; Tan et al., 2008; Volpato et al., 2002). Of those selected studies, 6 indicated significant (de Jong et al., 2006, 2009; Forti et al., 2012; Hogervorst et al., 2008; Tan et al., 2008; Volpato et al., 2002) and 3 indicated nonsignificant findings (Booth et al., 2013; de Jongh et al., 2011; Gussekloo et al., 2004). Although many of those studies used a single cognitive test score or dementia and/or AD diagnosis as the outcome, a number of findings are notable. For

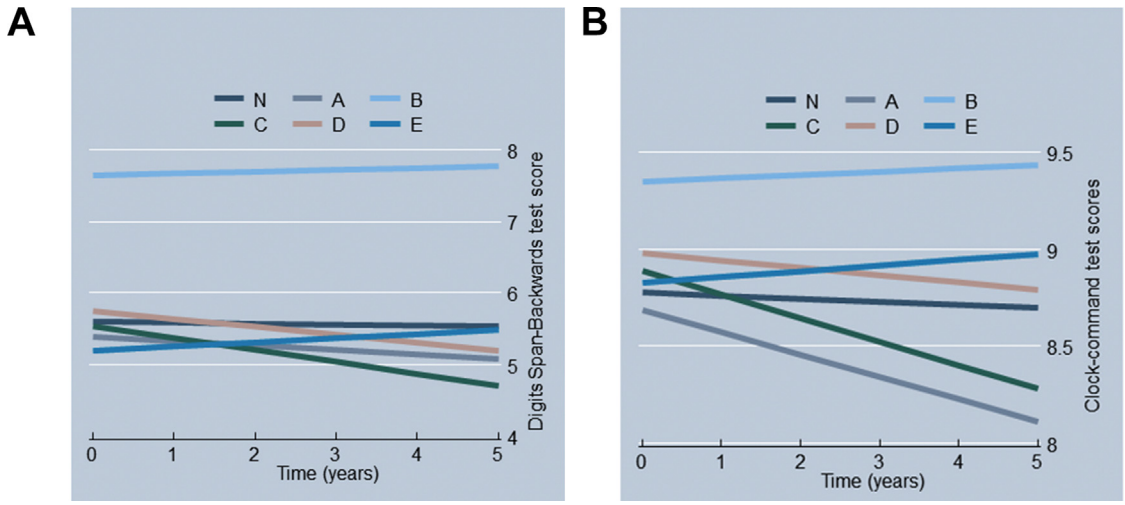

Fig. 1. (A) Predictive margins of digits span-backwards tests scores over-time from mixed-effects regression model by thyroid function status. (B) Predictive margins of clock-command test scores over-time from mixed-effects regression model by thyroid function status. Abbreviations: A, overt hypothyroidism; B, overt thyrotoxicosis; C, subclinical hypothyroidism; D, subclinical thyrotoxicosis; E, other dysfunction; N, normal.

instance, a large cohort study of older adults (age $\geq$ 65 years, n = 1047) observed that both higher TSH and fT_4 within normal ranges were associated with poorer performance and decline on the MMSE (Hogervorst et al., 2008). The latter study suggested that thyroxine can generate oxidative stress and damage neurons and concluded that treatment with thyroxine when thyroid disease is absent is not recommended and that optimal levels of thyroxine in the elderly is possibly lower than previously indicated (Hogervorst et al., 2008). Thus, further large cohort studies are needed to assess whether fT_4 levels indeed have a curvilinear relationship with cognitive function or decline among euthyroid individuals, whereby normal high fT_4 may result in worse cognitive outcomes.

In another study that failed to show an association between thyroid hormones and incident dementia (age: 60–90 years, n = 1077), higher fT_4 was shown to be associated with greater atrophy in the hippocampus and amygdala regions of the brain (de Jong et al., 2006). Those findings are comparable to ours, particularly the cross-sectional inverse relationship between fT_4 and performance in the domain of language or verbal fluency and the longitudinal association between higher TSH and faster rates of decline in the domains of working memory and visuospatial and/or visuoconstruction abilities. Similarly, higher TSH was associated with increased risk of vascular dementia but not AD or mild cognitive impairment in another large cohort study (Forti et al., 2012), whereas in the Framingham study (N = 1864 cognitively intact individuals), the risk of AD incidence among women was linked to both a high (>2.1 mIU/L: hazard ratio = 2.15 [95% confidence interval: 1.31–3.52, p = 0.003]) and a low (<1.0 mIU/L: hazard ratio = 2.39 [95% confidence interval: 1.47–3.87, $p < 0.001$]) TSH level (Tan et al., 2008).

Of 7 surveyed experimental studies (Bono et al., 2004; Burmeister et al., 2001; Correia et al., 2009; Miller et al., 2006; Munte et al., 2001; Osterweil et al., 1992; Parle et al., 2010), 3 had positive findings (Bono et al., 2004; Correia et al., 2009; Munte et al., 2001), whereas the others reported mixed or null findings (Burmeister et al., 2001; Miller et al., 2006; Osterweil et al., 1992; Parle et al., 2010). Specifically, L-thyroxine replacement was shown to normalize verbal memory in 1 trial for both overt and subclinical hypothyroid groups and for spatial memory among the subclinical hypothyroid group (Correia et al., 2009). In another trial of L-thyroxine replacement conducted among 36 women, slight improvements in verbal fluency and depression scores were noted that were accompanied by an increase in serum fT_4 in parallel with TSH level reduction (Bono et al., 2004).

Moreover, a neuroanatomical basis for the link between subclinical hypothyroidism and a defect in verbal working memory and executive function in particular was provided by a recent study (Zhu et al., 2006). In fact, subjects with a mean TSH of 14.7 mU/L were shown to have an impaired verbal working memory and abnormal functional magnetic resonance imaging findings in the frontal areas of the brain which are responsible for executive function. Of those participants, a subset was treated with L-T_4 for 6 months reducing TSH to a mean of 1.35 mU/L which normalized both verbal working memory and functional magnetic resonance imaging results, reflecting increased regional brain glucose metabolism with such treatment (Zhu et al., 2006). Similarly, a more recent study by the same group (Yin et al., 2013) showed similar results. Individuals with a mean TSH of 19.4 mIU/L exhibited decreased performance on a spatial working memory task (2-back), compared with euthyroid controls. Additionally, diminished functional activity in the right dorsolateral prefrontal cortex, right parietal lobe, and the supplementary motor area and anterior cingulate cortex was observed for those with elevated TSH levels, compared with controls. After treatment with L-T_4, TSH levels, visual working memory, and blood oxygen level dependent functional magnetic resonance imaging (BOLD fMRI) responses were similar between controls and subclinical hypothyroid patients (Yin et al., 2013).

Several mechanisms may explain the associations between thyroid function and cognition. First, both T_4 and its more potent metabolite T_3 are regulated in such a way as to preserve narrow concentration ranges in the brain, independent of changes in their corresponding bloodstream levels. This indicates that minute changes in thyroid hormones within brain tissues can alter behavior significantly. Moreover, T_3 levels in brain tissue is largely determined by circulating T_4 through local enzymatic deiodination (5′D-II deiodinase), rather than through active transport of serum T_3 into the brain. Importantly, thyroid hormones in several animal studies were shown to inhibit the expression of the β-amyloid precursor protein gene (Volpato et al., 2002). Other animal studies also show that adult-onset hypothyroidism in rats can reduce granule cells in the dentate gyrus and pyramidal cells of the hippocampal CA1 region, reduce apical dentritic spine density in the hippocampal CA1 pyramidal neurons, decrease synaptic plasticity within the hippocampus, and impair learning, particularly in spatial

Table 4
Longitudinal cognitive change by thyroid hormonal level within reference range: mixed-effects linear regression models[a,b]

	Intercept		Time		THWRR		(THWRR) × time	
	γ ± SEE	*p*	γ ± SEE	*p*	γ ± SEE	*p*	γ ± SEE	*p*
Mini-Mental State Exam, total score								
Model 1: TSH (N = 1452; N' = 2376)	**+27.0 ± 0.2**	**<0.001**	+0.06 ± 0.06	0.35	−0.10 ± 0.05	0.07	+0.01 ± 0.01	0.44
Model 2: fT_4 (N = 1500; N' = 2465)	**+27.5 ± 0.4**	**<0.001**	+0.09 ± 0.09	0.35	**−0.58 ± 0.27**	**0.029**	−0.00 ± 0.07	0.99
Model 3: tT_4 (N = 1406; N' = 2314)	**+27.1 ± 0.3**	**<0.001**	−0.02 ± 0.09	0.80	−0.02 ± 0.04	0.66	+0.01 ± 0.01	0.13
Model 4: T_{3pu} (N = 1462; N' = 2403)	**+27.0 ± 0.5**	**<0.001**	+0.13 ± 0.13	0.32	−0.00 ± 0.00	0.85	−0.00 ± 0.00	0.63
CVLT, list A								
Model 1: TSH (N = 1393; N' = 2183)	**+26.8 ± 0.8**	**<0.001**	**−1.42 ± 0.19**	**<0.001**	**−0.42 ± 0.19**	**0.029**	+0.03 ± 0.05	0.47
Model 2: fT_4 (N = 1438; N' = 2255)	**+24.0 ± 1.3**	**<0.001**	**−0.98 ± 0.30**	**0.001**	+1.60 ± 0.97	0.10	−0.25 ± 0.22	0.26
Model 3: tT_4 (N = 1397; N' = 2195)	**+25.0 ± 1.3**	**<0.001**	**−1.08 ± 0.29**	**<0.001**	+0.14 ± 0.13	0.28	−0.02 ± 0.03	0.57
Model 4: T3pu (N = 1397; N' = 2195)	**+25.1 ± 1.8**	**<0.001**	**−1.31 ± 0.41**	**0.002**	+0.02 ± 0.06	0.66	+0.00 ± 0.01	0.86
CVLT, free delayed recall								
Model 1: TSH (N = 1364; N' = 2091)	**+8.25 ± 0.39**	**<0.001**	**−0.42 ± 0.09**	**<0.001**	−0.13 ± 0.09	0.17	+0.01 ± 0.02	0.56
Model 2: fT_4 (N = 1411; N' = 2158)	**+6.85 ± 0.60**	**<0.001**	−0.21 ± 0.14	0.15	**+1.05 ± 0.46**	**0.022**	−0.18 ± 0.11	0.11
Model 3: tT_4 (N = 1326; N' = 2036)	**+7.49 ± 0.59**	**<0.001**	**−0.30 ± 0.14**	**0.031**	+0.08 ± 0.06	0.21	−0.01 ± 0.01	0.50
Model 4: T_{3pu} (N = 1373 ; N' = 2107)	**+7.55 ± 0.85**	**<0.001**	**−0.46 ± 0.20**	**0.023**	+0.02 ± 0.03	0.56	+0.00 ± 0.01	0.75
Benton visual retention test								
Model 1: TSH (N = 1467; N' = 2456)	**+8.63 ± 0.61**	**<0.001**	**+0.41 ± 0.15**	**0.007**	+0.19 ± 0.14	0.19	−0.02 ± 0.03	0.51
Model 2: fT_4 (N = 1515; N' = 2545),	**+8.62 ± 0.94**	**<0.001**	+0.23 ± 0.23	0.30	+0.31 ± 0.72	0.66	+0.11 ± 0.17	0.53
fT_4 × time × male: γ ± SEE:−0.93 ± 0.35, ***p* = 0.009**								
Men	**+5.69 ± 1.49**	**<0.001**	**+1.00 ± 0.35**	**0.005**	+1.92 ± 1.15	0.10	**−0.48 ± 0.27**	**0.07**
Women	**+9.64 ± 1.25**	**<0.001**	−0.23 ± 0.30	0.45	−0.40 ± 0.93	0.67	**+0.46 ± 0.22**	**0.039**
Model 3: tT_4 (N = 1422; N' = 2393), tT_4 × time × male:	**+9.03 ± 0.94**	**<0.001**	+0.40 ± 0.22	0.07	−0.05 ± 0.10	0.59	−0.00 ± 0.02	0.91
γ ± SEE:−0.09 ± 0.04, ***p* = 0.045**								
Men	**+6.91 ± 1.37**	**<0.001**	**+0.80 ± 0.31**	**0.011**	+0.11 ± 0.15	0.45	−0.05 ± 0.03	0.13
Women	**+10.30 ± 1.27**	**<0.001**	+0.00 ± 0.30	1.00	−0.20 ± 0.13	0.13	+0.04 ± 0.03	0.16
tT_4 × time × AA: γ ± SEE:−0.13 ± 0.05, ***p* = 0.004**								
Whites	**+9.65 ± 1.29**	**<0.001**	−0.48 ± 0.31	0.12	−0.23 ± 0.14	0.09	**+0.08 ± 0.03**	**0.010**
AA	**+8.90 ± 1.32**	**<0.001**	**+1.24 ± 0.31**	**<0.001**	+0.06 ± 0.13	0.67	**−0.06 ± 0.03**	**0.06**
Model 4: T_{3pu} (N = 1478; N' = 2483)	**+8.90 ± 1.31**	**<0.001**	**+0.67 ± 0.32**	**0.040**	−0.01 ± 0.04	0.80	−0.01 ± 0.01	0.35
Brief Test of Attention								
Model 1: TSH (N = 1468; N' = 2376)	**+6.22 ± 0.43**	**<0.001**	+0.03 ± 0.10	0.75	+0.15 ± 0.33	0.65	−0.09 ± 0.08	0.26
Model 2: fT_4 (N = 1376; N' = 2229)	**+6.14 ± 0.43**	**<0.001**	+0.03 ± 0.10	0.74	+0.05 ± 0.04	0.25	−0.01 ± 0.01	0.18
Model 3: tT_4 (N = 1376; N' = 2229)	**+6.14 ± 0.43**	**<0.001**	+0.03 ± 0.10	0.75	+0.05 ± 0.04	0.25	−0.01 ± 0.01	0.18
Model 4: T_{3pu} (N = 1427; N' = 2308)	**+6.27 ± 0.61**	**<0.001**	−0.13 ± 0.15	0.40	+0.00 ± 0.02	0.89	−0.00 ± 0.00	0.64
Animal fluency								
Model 1: TSH (N = 1471; N' = 2527)	**+17.16 ± 0.65**	**<0.001**	+0.03 ± 0.14	0.80	+0.00 ± 0.15	0.99	−0.03 ± 0.03	0.34
Model 2: fT_4 (N = 1518; N' = 2618)	**+15.3 ± 1.0**	**<0.001**	+0.08 ± 0.20	0.70	**+1.70 ± 0.76**	**0.024**	−0.13 ± 0.15	0.39
Model 3: tT_4 (N = 1421; N' = 2452)	**+15.4 ± 1.0**	**<0.001**	−0.18 ± 0.20	0.38	**+0.27 ± 0.10**	**0.007**	+0.02 ± 0.02	0.27
Model 4: T_{3pu} (N = 1477; N' = 2545)	**+18.2 ± 1.4**	**<0.001**	−0.22 ± 0.29	0.45	−0.02 ± 0.04	0.61	+0.01 ± 0.01	0.43
Digits span, forward								
Model 1: TSH (N = 1466; N' = 2416)	**+6.98 ± 0.26**	**<0.001**	−0.02 ± 0.06	0.69	−0.07 ± 0.06	0.21	+0.02 ± 0.01	0.10
Model 2: fT_4 (N = 1513; N' = 2496)	**+6.39 ± 0.40**	**<0.001**	−0.04 ± 0.09	0.67	+0.41 ± 0.31	0.18	+0.05 ± 0.07	0.42
Model 3: tT_4 (N = 1598; N' = 2632)	**+6.74 ± 0.30**	**<0.001**	−0.02 ± 0.07	0.80	+0.01 ± 0.03	0.68	+0.00 ± 0.01	0.43
Model 4: T_{3pu} (N = 1472; N' = 2431)	**+6.45 ± 0.56**	**<0.001**	−0.03 ± 0.12	0.80	+0.02 ± 0.02	0.36	+0.00 ± 0.00	0.80
Digits span, backward								
Model 1: TSH (N = 1465; N' = 2404)	+1.08 ± 4.82	0.82	+0.96 ± 1.13	0.40	**−0.12 ± 0.06**	**0.030**	−0.01 ± 0.01	0.64
Model 2: fT_4 (N = 1512; N' = 2482)	+0.94 ± 4.76	0.84	+1.26 ± 1.11	0.25	+0.47 ± 0.29	0.10	−0.05 ± 0.07	0.46
Model 3: tT_4 (N = 1420; N' = 2344)	+0.09 ± 4.91	0.99	+1.48 ± 1.13	0.19	+0.05 ± 0.04	0.21	−0.00 ± 0.01	0.87
Model 4: T_{3pu} (N = 1420; N' = 2344)	+0.03 ± 4.86	1.00	+1.53 ± 1.13	0.17	+0.02 ± 0.02	0.22	+0.00 ± 0.00	0.92
Clock-command								
Model 1: TSH (N = 1470; N' = 2533), TSH × time × male:	**+8.78 ± 0.16**	**<0.001**	−0.04 ± 0.04	0.39	+0.00 ± 0.04	0.93	−0.01 ± 0.01	0.19
γ ± SEE: +0.05 ± 0.02, ***p* = 0.009**								
Men	**+8.97 ± 0.24**	**<0.001**	−0.04 ± 0.06	0.38	−0.05 ± 0.06	0.38	+0.02 ± 0.02	0.20
Women	**+8.71 ± 0.21**	**<0.001**	−0.04 ± 0.06	0.48	+0.04 ± 0.05	0.43	**−0.03 ± 0.01**	**0.008**
Model 2: fT_4 (N = 1517; N' = 2614)	**+9.10 ± 0.24**	**<0.001**	**−0.20 ± 0.06**	**0.002**	−0.25 ± 0.18	0.18	**+0.09 ± 0.05**	**0.05**

Model 3: fT_4 (N = 1423; N' = 2454)	**+8.81 ± 0.24**	**<0.001**	**−0.13 ± 0.06**	**0.037**	+0.00 ± 0.01	0.52	+0.00 ± 0.01	0.52
Model 4: T_{3pu} (N = 1478; N' = 2546)	**+9.12 ± 0.33**	**<0.001**	−0.16 ± 0.09	0.09	−0.01 ± 0.01	0.26	+0.00 ± 0.00	0.44
Trail Making Test, part A								
Model 1: TSH (N = 1436; N' = 2426)	**+32.3 ± 4.7**	**<0.001**	+2.59 ± 1.37	0.06	1.26 ± 1.06	0.23	−0.21 ± 0.31	0.48
Model 2: fT_4 (N = 1484; N' = 2511)	**+46.4 ± 7.0**	**<0.001**	+2.00 ± 2.04	0.33	−10.42 ± 5.31	0.05	+0.16 ± 1.53	0.91
Model 3: fT_4 (N = 1395; N' = 2364)	**+33.2 ± 6.8**	**<0.001**	**+3.94 ± 1.93**	**0.041**	+0.10 ± 0.20	0.36	−0.18 ± 0.10	0.36
Model 4: T_{3pu} (N = 1447; N' = 2447)	**+52.2 ± 9.9**	**<0.001**	−0.12 ± 2.93	0.97	−0.60 ± 0.31	0.05	+0.09 ± 0.09	0.34
Trail Making Test, part B								
Model 1: TSH (N = 1426; N' = 2342)	**+178.8 ± 55.5**	**0.001**	+7.90 ± 13.19	0.55	+7.12 ± 4.14	0.09	−1.11 ± 0.86	0.20
Model 2: fT_4 (N = 1472; N' = 1472)	**+228.4 ± 60.1**	**<0.001**	+0.08 ± 14.07	1.00	−15.2 ± 21.1	0.47	+3.44 ± 4.20	0.47
Model 3: fT_4 (N = 1472; N' = 2419)	**+202.2 ± 60.2**	**0.001**	+4.08 ± 14.85	0.78	−1.53 ± 2.79	0.58	−1.53 ± 2.79	0.58
Model 4: T_{3pu} (N = 1435; N' = 2362)	**+201.7 ± 68.5**	**0.003**	+8.45 ± 14.74	0.57	−1.33 ± 1.19	0.27	−0.08 ± 0.26	0.76

Key: BVRT, Benton Visual Retention Test; CVLT, California Verbal Learning Test; fT4 , free thyroxine; HEI, Healthy Eating Index; MMSE, Mini-Mental State Examination; N, number of participants; N', number of visits; SEE, standard error of the estimate; T3pu , %uptake of tri-iodothyronine; THWRR, thyroid hormone within reference range; TSH, thyroid stimulating hormone; fT_4, total thyroxine; WRAT, Wide Range Achievement Test.
The numbers are bolded to highlight *p*-values that were <0.10 for interaction terms and <0.05 for main effects.
[a] Multiple mixed-effects linear regression models adjusted for baseline age, sex, race or ethnicity, marital status, education, WRAT total score, poverty income ratio, current smoking status, current use of illicit drugs, body mass index, and 2010-HEI.
[b] Most cognitive test scores were in the direction of higher score = better performance, except for BVRT (total errors) and Trail Making Test both parts (expressed in seconds).

and memory domains (Cao et al., 2012). Other adverse effects of thyroid dysfunction include altered expression of hippocampal enzymes that regulate catecholamine, serotonin, and γ-aminobutyric acid systems (Koromilas et al., 2010).

Our study has several notable strengths. In addition to its large sample size allowing for stratified analyses by sex and race, and its longitudinal design which allows us to ascertain temporality of associations, our study also included cognitive tests that spanned many domains of cognition, controlled for key potentially confounding factors that were sociodemographic, lifestyle, and health-related. It made use of advanced multivariable techniques, including mixed-effects regression models that took into account sample selectivity. In addition, the descriptive part of the analysis also accounted for sampling weights to obtain representative estimates of means and proportions.

Despite its strengths, our study findings should be interpreted in light of key limitations. First, although major potentially confounding variables were adjusted for, residual confounding cannot be ruled out. Specifically, although many central nervous system medications aside from antidepressants may affect thyroid hormonal level, previous studies have shown that their key findings were not affected by excluding individuals who were on any type of central nervous system medication (Prinz et al., 1999). Moreover, T_3 and TBG were not directly available in the first-visit of HANDLS, which prevented us from examining their association with longitudinal cognitive change over-time in this sample of US middle-aged urban adults. Although reference ranges are indicative of normal levels of thyroid hormones, they may vary according to populations and published evidence. Furthermore, only 2 time points were available for our longitudinal analyses, which although an improvement over cross-sectional analyses, may be limited compared to having 3 or more time points. Thus, our key finding of a significant relationship between higher baseline TSH and cognitive decline in domains of working memory and visuospatial and/or visuoconstruction abilities can possibly be the result of random fluctuation in performance rather than true decline. This random fluctuation is a result of reliability in the instrument itself and may also differ across study groups. Until further studies are done with 3 or more assessment on a comparable population of urban adults, this finding needs to be interpreted with caution. Furthermore, the effect size of the association between elevated TSH and the rate of change in measures of working memory and visuospatial and/or visuoconstruction abilities may have been large in relative terms compared to the "normal" group. However, in terms of absolute decline, the effect size was smaller than anticipated, possibly due to the young age at baseline of this study population. Finally, although a large battery of neuropsychological tests was available from which cognitive domains could be extracted using factor analysis, a prior attempt to group those individual tests into distinctive domains showed that there was a lack of factorial invariance across the major variables used in HANDLS sampling design, including sex, race, age, and poverty status. For this reason, only individual test scores were used and interpreted in terms of their salient domain of cognitive performance.

In sum, our study findings indicated that thyroid hormones, particularly higher TSH, are linked to faster rate of cognitive decline over-time, particularly in domains of working memory and visuospatial and/or visuoconstruction ability. Moreover, subclinical hypothyroidism whereby higher TSH levels are coupled with normal fT_4 levels was specifically linked to decline over-time, as well as higher TSH values within normal ranges among women in the case of visuospatial and/or visuoconstruction ability. Furthermore, large cohort studies are needed to replicate those findings as well as hormone replacement interventions that examine both short-term and long-term effects of thyroid hormones on age-related cognitive decline in different domains of cognition.

Disclosure statement

The authors declare no conflict of interest.

Acknowledgements

This study was entirely supported by the National Institute on Aging, Intramural Research Program (NIA/NIH/IRP).

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.neurobiolaging.2015.08.002.

References

Almeida, C., Vaisman, M., Costa, A.J., Reis, F.A., Reuters, V., Teixeira, P., Ferreira, M., Teixeira, L.B., Araujo, G.R., Brasil, M.A., 2007. Are neuropsychological changes relevant in subclinical hypothyroidism? Arq. Bras. Endocrinol. Metabol. 51, 606–611.

Alzheimer's Association, 2009. 2009 Alzheimer's disease facts and figures. Alzheimers Dement. 5, 234–270.

Aoki, Y., Belin, R.M., Clickner, R., Jeffries, R., Phillips, L., Mahaffey, K.R., 2007. Serum TSH and total T4 in the United States population and their association with participant characteristics: National Health and Nutrition Examination Survey (NHANES 1999-2002). Thyroid 17, 1211–1223.

Baskin, H.J., Cobin, R.H., Duick, D.S., Gharib, H., Guttler, R.B., Kaplan, M.M., Segal, R.L., 2002. American Association of Clinical Endocrinologists medical guidelines for clinical practice for the evaluation and treatment of hyperthyroidism and hypothyroidism. Endocr. Pract. 8, 457–469.

Beydoun, M.A., Beydoun, H.A., Kitner-Triolo, M.H., Kaufman, J.S., Evans, M.K., Zonderman, A.B., 2013. Thyroid hormones are associated with cognitive function: moderation by sex, race, and depressive symptoms. J. Clin. Endocrinol. Metab. 98, 3470–3481.

Beydoun, M.A., Beydoun, H.A., Shroff, M.R., Kitner-Triolo, M.H., Zonderman, A.B., 2012. Serum leptin, thyroxine and thyroid-stimulating hormone levels interact to affect cognitive function among US adults: evidence from a large representative survey. Neurobiol. Aging 33, 1730–1743.

Beydoun, M.A., Gamaldo, A.A., Beydoun, H.A., Tanaka, T., Tucker, K.L., Talegawkar, S.A., Ferrucci, L., Zonderman, A.B., 2014. Caffeine and alcohol intakes and overall nutrient adequacy are associated with longitudinal cognitive performance among U.S. adults. J. Nutr. 144, 890–901.

Bono, G., Fancellu, R., Blandini, F., Santoro, G., Mauri, M., 2004. Cognitive and affective status in mild hypothyroidism and interactions with L-thyroxine treatment. Acta Neurol. Scand. 110, 59–66.

Booth, T., Deary, I.J., Starr, J.M., 2013. Thyroid stimulating hormone, free thyroxine and cognitive ability in old age: the Lothian birth cohort study 1936. Psychoneuroendocrinology 38, 597–601.

Burmeister, L.A., Ganguli, M., Dodge, H.H., Toczek, T., DeKosky, S.T., Nebes, R.D., 2001. Hypothyroidism and cognition: preliminary evidence for a specific defect in memory. Thyroid 11, 1177–1185.

Cao, L., Wang, F., Yang, Q.G., Jiang, W., Wang, C., Chen, Y.P., Chen, G.H., 2012. Reduced thyroid hormones with increased hippocampal SNAP-25 and Munc18-1 might involve cognitive impairment during aging. Behav. Brain Res. 229, 131–137.

Ceballos, A., Belinchon, M.M., Sanchez-Mendoza, E., Grijota-Martinez, C., Dumitrescu, A.M., Refetoff, S., Morte, B., Bernal, J., 2009. Importance of monocarboxylate transporter 8 for the blood-brain barrier-dependent availability of 3,5,3'-triiodo-L-thyronine. Endocrinology 150, 2491–2496.

Ceresini, G., Lauretani, F., Maggio, M., Ceda, G.P., Morganti, S., Usberti, E., Chezzi, C., Valcavi, R., Bandinelli, S., Guralnik, J.M., Cappola, A.R., Valenti, G., Ferrucci, L., 2009. Thyroid function abnormalities and cognitive impairment in elderly people: results of the Invecchiare in Chianti study. J. Am. Geriatr. Soc. 57, 89–93.

Correia, N., Mullally, S., Cooke, G., Tun, T.K., Phelan, N., Feeney, J., Fitzgibbon, M., Boran, G., O'Mara, S., Gibney, J., 2009. Evidence for a specific defect in hippocampal memory in overt and subclinical hypothyroidism. J. Clin. Endocrinol. Metab. 94, 3789–3797.

de Jong, F.J., den Heijer, T., Visser, T.J., de Rijke, Y.B., Drexhage, H.A., Hofman, A., Breteler, M.M., 2006. Thyroid hormones, dementia, and atrophy of the medial temporal lobe. J. Clin. Endocrinol. Metab. 91, 2569–2573.

de Jong, F.J., Masaki, K., Chen, H., Remaley, A.T., Breteler, M.M., Petrovitch, H., White, L.R., Launer, L.J., 2009. Thyroid function, the risk of dementia and neuropathologic changes: the Honolulu-Asia aging study. Neurobiol. Aging 30, 600–606.

de Jongh, R.T., Lips, P., van Schoor, N.M., Rijs, K.J., Deeg, D.J., Comijs, H.C., Kramer, M.H., Vandenbroucke, J.P., Dekkers, O.M., 2011. Endogenous subclinical thyroid disorders, physical and cognitive function, depression, and mortality in older individuals. Eur. J. Endocrinol. 165, 545–554.

Dugbartey, A.T., 1998. Neurocognitive aspects of hypothyroidism. Arch. Intern. Med. 158, 1413–1418.

Evans, M.K., Lepkowski, J.M., Powe, N.R., LaVeist, T., Kuczmarski, M.F., Zonderman, A.B., 2010. Healthy aging in neighborhoods of diversity across the life span (HANDLS): overcoming barriers to implementing a longitudinal, epidemiologic, urban study of health, race, and socioeconomic status. Ethn. Dis. 20, 267–275.

Fontana, L., Klein, S., Holloszy, J.O., Premachandra, B.N., 2006. Effect of long-term calorie restriction with adequate protein and micronutrients on thyroid hormones. J. Clin. Endocrinol. Metab. 91, 3232–3235.

Formiga, F., Ferrer, A., Padros, G., Contra, A., Corbella, X., Pujol, R., Octabaix Study, G., 2014. Thyroid status and functional and cognitive status at baseline and survival after 3 years of follow-up: the OCTABAIX study. Eur. J. Endocrinol. 170, 69–75.

Forti, P., Olivelli, V., Rietti, E., Maltoni, B., Pirazzoli, G., Gatti, R., Gioia, M.G., Ravaglia, G., 2012. Serum thyroid-stimulating hormone as a predictor of cognitive impairment in an elderly cohort. Gerontology 58, 41–49.

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

Gussekloo, J., van Exel, E., de Craen, A.J., Meinders, A.E., Frolich, M., Westendorp, R.G., 2004. Thyroid status, disability and cognitive function, and survival in old age. JAMA 292, 2591–2599.

Heckman, J.J., 1979. Sample selection bias as a specification error. Econometrica 47, 153–161.

Hochberg, Y., Tamhane, A.C., 1987. Multiple Comparison Procedures. Wiley, New York.

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

Joffe, R.T., Pearce, E.N., Hennessey, J.V., Ryan, J.J., Stern, R.A., 2013. Subclinical hypothyroidism, mood, and cognition in older adults: a review. Int. J. Geriatr. Psychiatry 28, 111–118.

Koromilas, C., Liapi, C., Schulpis, K.H., Kalafatakis, K., Zarros, A., Tsakiris, S., 2010. Structural and functional alterations in the hippocampus due to hypothyroidism. Metab. Brain Dis. 25, 339–354.

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

Lohr, S.L., 1999. Sampling: Design and Analysis. Duxbury-Press, Boston, MA.

Miller, K.J., Parsons, T.D., Whybrow, P.C., van Herle, K., Rasgon, N., van Herle, A., Martinez, D., Silverman, D.H., Bauer, M., 2006. Memory improvement with treatment of hypothyroidism. Int. J. Neurosci. 116, 895–906.

Munte, T.F., Radamm, C., Johannes, S., Brabant, G., 2001. Alterations of cognitive functions induced by exogenous application of thyroid hormones in healthy men: a double-blind cross-over study using event-related brain potentials. Thyroid 11, 385–391.

Osterweil, D., Syndulko, K., Cohen, S.N., Pettler-Jennings, P.D., Hershman, J.M., Cummings, J.L., Tourtellotte, W.W., Solomon, D.H., 1992. Cognitive function in non-demented older adults with hypothyroidism. J. Am. Geriatr. Soc. 40, 325–335.

Parle, J., Roberts, L., Wilson, S., Pattison, H., Roalfe, A., Haque, M.S., Heath, C., Sheppard, M., Franklyn, J., Hobbs, F.D., 2010. A randomized controlled trial of the effect of thyroxine replacement on cognitive function in community-living elderly subjects with subclinical hypothyroidism: the Birmingham Elderly Thyroid study. J. Clin. Endocrinol. Metab. 95, 3623–3632.

Prinz, P.N., Scanlan, J.M., Vitaliano, P.P., Moe, K.E., Borson, S., Toivola, B., Merriam, G.R., Larsen, L.H., Reed, H.L., 1999. Thyroid hormones: positive relationships with cognition in healthy, euthyroid older men. J. Gerontol. A Biol. Sci. Med. Sci. 54, M111–M116.

Ross, D.S., 1988. Subclinical hyperthyroidism: possible danger of overzealous thyroxine replacement therapy. Mayo Clin. Proc. 63, 1223–1229.

Samuels, M.H., 2014. Thyroid disease and cognition. Endocrinol. Metab. Clin. North Am. 43, 529–543.

Samuels, M.H., Schuff, K.G., Carlson, N.E., Carello, P., Janowsky, J.S., 2007. Health status, psychological symptoms, mood, and cognition in L-thyroxine-treated hypothyroid subjects. Thyroid 17, 249–258.

Selvin, S., 2004. Statistical Analysis of Epidemiologic Data, third ed. Oxford University Press, New York, NY.

Tan, Z.S., Beiser, A., Vasan, R.S., Au, R., Auerbach, S., Kiel, D.P., Wolf, P.A., Seshadri, S., 2008. Thyroid function and the risk of Alzheimer disease: the Framingham Study. Arch. Intern. Med. 168, 1514–1520.

van Boxtel, M.P., Menheere, P.P., 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.

van de Rest, O., Berendsen, A.A., Haveman-Nies, A., de Groot, L.C., 2015. Dietary patterns, cognitive decline, and dementia: a systematic review. Adv. Nutr. 6, 154–168.

Volpato, S., Guralnik, J.M., Fried, L.P., Remaley, A.T., Cappola, A.R., Launer, L.J., 2002. Serum thyroxine level and cognitive decline in euthyroid older women. Neurology 58, 1055–1061.

Wijsman, L.W., de Craen, A.J., Trompet, S., Gussekloo, J., Stott, D.J., Rodondi, N., Welsh, P., Jukema, J.W., Westendorp, R.G., Mooijaart, S.P., 2013. Subclinical thyroid dysfunction and cognitive decline in old age. PLoS One 8, e59199.

Yin, J.J., Liao, L.M., Luo, D.X., Xu, K., Ma, S.H., Wang, Z.X., Le, H.B., Huang, R.R., Cai, Z.L., Zhang, J., 2013. Spatial working memory impairment in subclinical hypothyroidism: an FMRI study. Neuroendocrinology 97, 260–270.

Zhu, D.F., Wang, Z.X., Zhang, D.R., Pan, Z.L., He, S., Hu, X.P., Chen, X.C., Zhou, J.N., 2006. fMRI revealed neural substrate for reversible working memory dysfunction in subclinical hypothyroidism. Brain 129 (Pt 11), 2923–2930.

APPENDIX I: Description of cognitive tests and the CES-D

Mini-Mental State Examination (MMSE)

The MMSE (1) is a brief mental status test and global cognitive functioning measuring orientation, concentration, immediate and delayed memory, language and constructional praxis. Scores range from 0 to 30, with higher scores indicating better cognitive performance.

California Verbal Learning Test (CVLT)

The CVLT (2) is a 16-item shopping list measuring verbal learning and memory. A modified version of the CVLT was used with three, rather than five, list A learning trials. Cued recall was not administered. Variables of interest in this study were total correct for List A sum across trials 1-3 and List A long-delay free recall. Scores ranged from 0 to 48 for List A sum and 0 to 16 for List A long-delay free recall. Higher scores indicate better verbal memory. The CVLT is described in detail elsewhere (2).

Benton Visual Retention Test (BVRT)

The BVRT (3) is a test of short-term figural memory and visuo-constructional abilities. Administration A, Form D was used. Two trained examiners independently scored the BVRT using a modified error scoring system, based on the BVRT Manual scoring. A consensus was achieved for discrepancies in scoring. If a consensus between the two examiners could not be reached, MKT, a research psychologist assigned the score. Scores were total errors, such that higher values indicate poorer visual memory.

Digit Span Forward and Backward (DS-F and DS-B)

The Wechsler Adult Intelligence Scale, Revised(4) Digit Span Forward and Backward are tests of attention and executive functioning, specifically working memory. They were administered according to standard instructions, and the total score was the total number correct for each test.

Animal Fluency

Animal fluency, a measure of semantic verbal fluency, requires participants to generate as many animals as possible for 60 seconds. Higher scores indicate better verbal fluency, with the total number of words, minus intrusions and perseverations analyzed.

Brief Test of Attention (BTA)

The BTA (5) is a measure of divided auditory attention. An examiner administered 10 trials where increasing longer lists of letters and numbers (containing 4-18 items) were read. Participants were instructed to keep track of how many numbers were read during each trial, disregarding the number of letters, and were told to keep their hands in fists to discourage counting on their fingers. Only the numbers portion of the test was administered. The total score was the total number of trials correct out of 10.

Trail Making Tests A and B (Trails A and Trails B)

Trailmaking test A and B(6) are tests of attention and executive functioning, respectively, specifically cognitive control and visuo-motor scanning. Participants were instructed to draw lines between consecutive numbers (Trails A) or alternate between numbers and letter (Trails B) as fast as they could while a stop watch recorded time. When errors were committed the participant corrected the error by returning to his/her last correct response and continued from there. The stop-watch ran while corrections

were made. Scores reflected time to completion (in seconds) separately for Trails A and B. Higher scores indicate poorer performance.

Clock Drawing Test – Clock to Command (CDT)

The Clock Drawing Test (7) is a test of visuo-spatial and visuo-constructional abilities. Participants are asked to draw a clock, put in all of the numbers and set the hands for 10 after 11. Scores are assessed for the clock face (0-2), numbers (0-4) and hands (0-4), with a range from 0 to 10, with higher scores indicating more accurate clock drawing. Participants who did not score a 10 on the command version of the test were asked to copy a clock with the time set to 10 after 11.

Wide Range Achievement Test – 3rd Edition: Word and Letter Reading Subtest (WRAT)

The WRAT Word and Letter Reading Subtest (8) is a test of verbal knowledge, frequently used as a proxy for literacy and educational quality. Participants were asked to pronounce a list of 50 words that increased in difficulty. If a criterion of the first five words correctly pronounced was not reached, letter reading was administered. The tan form was administered according to standard instruction and the score was the total number of words correctly pronounced.

Center for Epidemiological Studies Depression Scale (CES-D)

The CES-D (9) is a 20-item measure of depressive symptoms. Participants are asked to rate the frequency and severity of symptoms over the past week. Scores range from 0 to 60, with scores of 16 and higher indicating significant depressive symptoms, and scores of 20 and higher indicating significant clinically depressive symptoms.

APPENDIX II: Description of mixed-effects regression models

The mixed-effects regression models can be summarized as follows:

	Multi-level models vs. **Composite models**	
Eq.		
1.1-1.4	$Y_{ij} = \pi_{0i} + \pi_{1i} Time_{ij} + \varepsilon_{ij}$	$Y_{ij} = \gamma_{00} + \gamma_{0a} X_{aij} + \sum_{k=1}^{l} \gamma_{0k} Z_{ik}$
	$\pi_{0i} = \gamma_{00} + \gamma_{0a} X_{aij} + \sum_{k=1}^{l} \gamma_{0k} Z_{ik} + \zeta_{0i}$	$+ \gamma_{10} Time_{ij} + \gamma_{1a} X_{aij} Time_{ij}$
	$\pi_{1i} = \gamma_{10} + \gamma_{1a} X_{aij} + \sum_{m=1}^{n} \gamma_{1m} Z_{im} + \zeta_{1i}$	$+ \sum_{m=1}^{n} \gamma_{1m} Z_{im} Time_{ij}$
		$+ (\zeta_{0i} + \zeta_{1i} Time_{ij} + \varepsilon_{ij})$

Where Y_{ij} is the outcome (cognitive test scores) for each individual "i" and visit "j"; π_{0i} is the level-1 intercept for individual i; π_{1i} is the level-1 slope for individual i; γ_{00} is the level-2 intercept of the random intercept π_{0i}; γ_{10} is the level-2 intercept of the slope π_{1i}; Z_{ik} is a vector of fixed covariates for each individual i that are used to predict level-1 intercepts and slopes and included baseline age (Age_{base}) among other covariates. $X_{ija,}$ represents the main predictor variables (thyroid hormone exposures); ζ_{0i} and ζ_{1i} are level-2 disturbances; ε_{ij} is the within-person level-1 disturbance. Of primary interest are the main effects of each exposure X_a (γ_{0a}) and their interaction with *TIME* (γ_{1a}), as described in a previous methodolgical paper.(10)

References:

1. Folstein MF, Folstein SE, McHugh PR. "Mini-mental state". A practical method for grading the cognitive state of patients for the clinician. J Psychiatr Res 1975; 12:189-198
2. Delis DC, Freeland J, Kramer JH, Kaplan E. Integrating clinical assessment with cognitive neuroscience: construct validation of the California Verbal Learning Test. J Consult Clin Psychol 1988; 56:123-130

3. Benton AL, ed. Revised visual retention test (fifth edition). New York: The Psychological Corportation; 1974.

4. Wechsler D. WAIS-R manual. Cleveland: The Psychological Corporation;1981.

5. Schretlen D, Bobholz JH, Brandt J. Development and psychometric properties of the Brief Test of Attention. Clinical Neuropsychologist 1996; 10:80-89

6. Reitan R. Trail Making Test: Manual for Administration and Scoring. Tucson, AZ: Reitan Neuropsychological Laboratory.

7. Rouleau I, Salmon DP, Butters N, Kennedy C, McGuire K. QUANTITATIVE AND QUALITATIVE ANALYSES OF CLOCK DRAWINGS IN ALZHEIMERS AND HUNTINGTONS-DISEASE. Brain and Cognition 1992; 18:70-87

8. Wilkinson GS. Wide Range Achievement Test–Revision 3. Wilmington, DE: Jastak Association.

9. Nguyen HT, Kitner-Triolo M, Evans MK, Zonderman AB. Factorial invariance of the CES-D in low socioeconomic status African Americans compared with a nationally representative sample. Psychiatry research 2004; 126:177-187

10. Blackwell E, de Leon CF, Miller GE. Applying mixed regression models to the analysis of repeated-measures data in psychosomatic medicine. Psychosom Med 2006; 68:870-878