Title

Effects of Hormone Therapy on List and Story Recall in Post-Menopausal Women

Journal

Experimental Aging Research, 45(3)

ISSN

0361-073X

Authors

Ober, Beth A
Shenaut, Gregory K
Taylor, Sandra L

Publication Date

2019-05-27

DOI

10.1080/0361073x.2019.1609169

Peer reviewed

HHS Public Access

Author manuscript

Exp Aging Res. Author manuscript; available in PMC 2020 May 01.

Published in final edited form as:

Exp Aging Res. 2019 ; 45(3): 199–222. doi:10.1080/0361073X.2019.1609169.

Effects of hormone therapy on list and story recall in post-menopausal women

Beth A. Ober, **Gregory K. Shenaut**, and **Sandra L. Taylor**

University of California, Davis

Abstract

Background/Study Context.—A number of longitudinal randomized controlled trials (LRCT) have used free verbal recall tests to study the effects of post-menopausal estrogen hormone therapy (HT) on episodic memory, but none have explicitly explored contrasts between list and story recall, in spite of cognitive differences between the tasks. For example, list recall provides little support for the use of gist, while story recall emphasizes it, and there is evidence that estrogen produces gist bias. Moreover, we present a literature tabulation that also suggests a task-specific HT effect.

Methods.—In a LRCT with up to eight yearly test sessions, post-menopausal women were randomly assigned either to placebo (N=56) or to an estrogen formulation (N=44); subgroups received either estrogen alone (hysterectomy; E-Alone; *N*=16) or with progestin (intact uterus; E+P; *N*=28). Participants were tested on immediate and delayed list and story recall at each session.

Results.—Linear mixed effects analyses of longitudinal trajectories showed that relative to placebo, the HT group declined significantly faster on immediate list recall and slower on immediate story recall. Separate analyses produced a sharpened version of this pattern for the E-Alone subgroup, but found no significant effects for the E+P subgroup. No significant effects were found in delayed testing.

Conclusion.—The dissociation we found for immediate list and story recall is similar to the pattern of results in our literature tabulation. Fuzzy-Trace Theory posits parallel verbatim and gist traces plus a meta-cognitive review which becomes more gist-biased with age. Our results suggest that: (1) estrogen increases gist bias, hastening the normal age-related decline of list recall but slowing the decline of story recall relative to placebo; (2) decay of the verbatim trace over time generally causes a shift to gist, thereby accounting for the absence of a delayed recall difference; and (3) progestin weakens the effects of estrogen, thereby accounting for why the dissociation found in E-Alone was absent in the E+P subgroup.

Keywords

Episodic Memory; Women; Aging; Estrogen; Longitudinal Randomized Clinical Trial; Fuzzy-Trace Theory

Correspondence concerning this article should be addressed to: baober@ucdavis.edu.

Beth A. Ober and Gregory K. Shenaut, Human Development and Family Studies, Department of Human Ecology, University of California, Davis; Sandra L. Taylor, Department of Public Health Sciences, School of Medicine, University of California, Davis.

This paper reports the results of the first longitudinal randomized controlled trial (LRCT) to focus on the direct comparison of the rate of change over time of list versus story recall, using both immediate and delayed testing in postmenopausal women receiving either estrogen alone, estrogen plus progestin, or placebo. The results of this study as well as those of a systematic review of the subset of LRCT studies on HT that report change over time in list or story recall are analyzed in terms of Fuzzy-Trace Theory (FTT).

There is evidence that postmenopausal estrogen hormone therapy (HT) affects the rate of age-related decline in verbal free recall performance, and that the effect is modulated by hormone formulation (i.e., addition of progestin for women with intact uteruses). There is also theoretical support for the notion that the effect may be modulated by parameters such as type of study material (i.e., lists versus stories) and time of test (immediate or delayed), but no previous HT study has evaluated the impact of such parameters on verbal free recall.

In the rest of the introduction, we first provide general background information on the relation of HT to: episodic memory and the brain; use of verbatim (literal, sensory) and gist (meaning-related, associative) information during recall; list versus story recall; and immediate versus delayed recall. Next, we introduce the Fuzzy-Trace Theory framework, which we believe is well suited to represent the relevant issues. We then report a systematic tabulation of all HT LRCTs that report specific outcomes for list or story recall. Finally, we present the specific aims and hypotheses of the current study.

HT, Brain, and Verbal Free Recall

In many HT studies, both estrogen and progestin must be considered, because HT for women with an intact uterus is formulated with progestin to reduce the risk of uterine cancer. Differences have been obtained in the neurocognitive effects of estrogen alone (no uterus; E-Alone) versus estrogen + progestin (intact uterus; E+P). In the Women's Health Initiative (WHI) Memory Study (WHIMS), there was doubling of the risk for dementia from all causes for E+P relative to placebo (Shumaker et al., 2003), whereas no such effect was found for E-Alone (Shumaker et al., 2004). However, no differences in rate of change in brain or lesion volumes were found between E+P and E-Alone in a later WHIMS magnetic resonance imaging (MRI) study by Coker et al. (2014), who concluded that the "overall effects of [E-Alone] vs [E+P] therapy on the brain are much more similar than different… suggesting that the primary agent in the effects we observe is [estrogen]".

Hippocampus and prefrontal cortex are particularly affected by declining tissue volume in normal aging; these areas are also critical to episodic memory function (see review by Park & Reuter-Lorenz, 2009). It is thus of great interest that Resnick et al. (2009) reported that both E-Alone and E+P resulted in significantly decreased frontal and hippocampal volumes. Functional MRI (fMRI) has shown estrogen-enhanced activation with E-Alone in prefrontal cortex in a verbal recall task (Joffe et al., 2006), as well as increased prefrontal, cingulate, and parietal (but not hippocampal) activations with E+P as a function of semantic versus phonemic encoding of words (Persad et al., 2009). These fMRI results are consistent with many non-HT studies showing increased frontal activation for older versus younger adults,

likely compensatory, in conjunction with decreased hippocampal activity in episodic memory tasks (Park & McDonough, 2013; Park & Reuter-Lorenz, 2009).

A review and descriptive meta-analysis concluded that HT (either E-Alone or E+P) positively affects brain activation during memory processing in postmenopausal women, especially in prefrontal cortex (Comasco, Frokjaer, & Sundstrom-Poromaa, 2014). This leads us to consider whether HT could in some cases counteract the negative effects of aging on prefrontal functioning and, in turn, aspects of episodic memory functioning that depend on prefrontal cortex. Episodic encoding and retrieval strategies that depend on the frontal cortex "working with" the hippocampus (Moscovitch & Winocur, 2002) may include those based on semantic, associative, and related types of gist-based processing. LRCTs that have contrasted E-Alone with E+P for verbal episodic memory without regard to availability or use of semantic/gist processing have yielded mixed results such that E-Alone has been somewhat more likely to show positive or mixed effects, and E+P has been more likely to show negative effects, especially for tests of verbal episodic memory (see Hogervorst & Bandelow, 2010; Maki, 2012).

Estrogen and Gist Bias.

In a series of studies on hormone contraception-using versus normally cycling women, Nielsen and colleagues have shown that estrogen can influence women's preference for gist versus verbatim memory with relatively high emotional content (Nielsen, Ahmed, & Cahill, 2013; Nielsen, Ertman, Lakhani, & Cahill, 2011; Petersen, Patihis, & Nielsen, 2014). For example, in a study involving false memory for misinformation in emotional scenes (Petersen et al., 2014), hormone contraception users—who generally have lower levels of endogenous estradiol than normally cycling women—produced fewer false memories than normally cycling women. The authors interpreted their results as signifying that hormonal contraceptive users show a bias toward gist memory and away from verbatim memory. While the studies by Nielsen and colleagues did not involve post-menopausal HT, the results suggest that estrogen may differentially affect story and list recall by modulating the extent to which participants employ gist versus verbatim information.

List Versus Story Recall.

List and story recall tests are effective measures of verbal episodic memory performance, are widely used in neuropsychological testing of memory, and are strongly correlated with each other (Delis, Cullum, Butters, Cairns, & Prifitera, 1988; Mulligan, 2005), although the two tasks are thought to employ somewhat different mental processes, including greater dependence on executive function in list compared to story recall (Lezak, Howieson, Bigler, & Tranel, 2012). There have been several longitudinal studies of normal aging that compared or permitted comparison of the rate of change in list versus story recall. For example, Small, Dixon, Hultsch, and Hertzog (1999) found, over six-years of assessment, a greater absolute rate of decline in list recall for participants who were 71–86 years of age, versus 55–70, at the initial testing session; for story recall, participants who began the study at age 55–70 improved slightly, while those who began at age 71–86 declined. On the other hand, Zimprich and Kurtz (2015), in a five-year trial with a single group age 65–80 at onset, found a net decline for both list and story recall. In clinical testing, significantly impaired list

versus spared story recall has been reported in patients with head injury (Cooper, Numan, Crosson, & Velozo, 1989), epilepsy (Helmstaedter, Wietzke, & Lutz, 2009), mild cognitive impairment (Tremont, Miele, Smith, & Westervelt, 2010), and late-life depression (Lamar, Charlton, Zhang, & Kumar, 2012). No previous study has raised the question of differences in the effects of HT on list versus story recall.

Immediate Versus Delayed Recall.

In a free verbal recall task, there may be one or more immediate recall trials consisting of a presentation of the stimuli followed immediately by recall; a delayed recall trial may occur 15–30 minutes after the last immediate trial. It is unknown whether HT may differentially affect immediate versus delayed recall. For example, Maki, Gast, Vieweg, Burriss, and Yaffe (2007) showed no effect of HT on immediate recall, but a negative effect on short- and long-delay recall measures. A contrasting finding was reported by Tierney et al. (2009), who found a positive effect of HT on delayed recall at follow-up, but no effect of HT on immediate recall; this difference was observed only for the higher-performing participants on baseline short-delay free recall performance. A number of LRCTs have applied both immediate and delayed verbal free recall tests to HT studies (Hogervorst & Bandelow, 2010), but none have considered whether HT might affect them differently.

Fuzzy-Trace Theory

The theoretical cognitive model known as Fuzzy-Trace Theory (FTT; e.g., Reyna & Brainerd, 2011) can usefully be applied to the differences between list and story recall tests. In the FTT model, two types of memory trace are formed in parallel during encoding. One is a verbatim trace, encoding perceptual details and specifics; the other is a gist trace, encoding context and higher-level associations. The two types of trace are also retrieved in parallel: verbatim retrieval occurs through direct access to the original perceptual experiences, while gist retrieval involves strategic reconstruction of the original episode. Since retrieved verbatim and reconstructed gist traces can conflict, FTT postulates a third element, a metacognitive review that inhibits elements of verbatim or gist traces according to a set of variable criteria. The metacognitive review is generally biased toward gist-based reconstruction, and this bias has been found to increase throughout the lifespan, from childhood when use of gist is undeveloped and ineffective, into later adulthood when verbatim memory becomes increasingly limited and short-lived (Koutstaal & Schacter, 1997; Reyna & Mills, 2007). It is worth noting that when verbatim recall is fragmentary and gist recall dominates, the reconstructed memory may not be veridical; in fact, false memory effects such as false alarms in recognition tasks and intrusions in list recall tasks have been considered indicators of reliance on gist recall (Brainerd, Reyna, & Ceci, 2008).

List recall tasks require verbatim retrieval of presented list items. Verbatim retrieval is of course fully correct in story recall tasks as well, but credit is also given for retrieval of paraphrased narrative elements (i.e., gist). However, gist information can be helpful in both list and story recall. Some word lists support use of gist by inclusion of words belonging to one or more semantic categories, but story recall tasks inherently evoke more meaningful associations because of their use of complete sentences organized into a coherent narrative.

In either task, participants may also make emotional, situational, autobiographical, or strategically mnemonic associations to the study material. FTT differentiates between immediate versus delayed testing in that from the FTT perspective, gist memory "usurps" verbatim memory after a delay (e.g., Reyna & Mills, 2007). That is, the more rapid decay of the verbatim trace after a delay increases reliance on reconstruction based on the slower-decaying gist trace.

A Systematic Tabulation of Relevant Literature

Although there have been systematic literature reviews on HT LRCTs (Hogervorst & Bandelow, 2010; Maki, 2012), the issue of differences due to use of story/list or immediate/delayed recall has not yet been considered explicitly. In order to determine whether there are meaningful patterns relevant to this issue in the existing literature, we systematically tabulated every available result from HT LRCTs using list or story recall.

Approach.

We surveyed every HT LRCT that reported the significance of immediate or delayed word list or story recall for an HT group compared to placebo (no study has explicitly compared the two recall measures). We selected studies from those included in Hogervorst and Bandelow (2010), Maki and Sundermann (2009), and Maki (2012) that fit our criteria, and found several more recent articles via conventional literature keyword searches. Some studies were excluded due to such issues as combining list and story recall into the same factor such that neither could be assessed individually, combining E-Alone and E+P into a single treatment group without reporting either separately, or not fulfilling the definition of LRCT in some way. Seventeen studies met criteria for inclusion. Due to the exploratory nature of this survey, we utilized a liberal criterion of $p < .1$ for classification as a non-null result. We classified differences between HT and placebo reaching $p < .05$ as "significant" findings, and differences reaching $p < .1$ or mixed results (i.e., where some results in the specified table cell were and some were not significant) as "marginally significant" trends. We classified groups receiving E+P separately from those receiving E-Alone, but did not further distinguish among hormone preparations.

Findings.

Table 1 presents the current state of the literature in a Treatment (E-Alone, E+P) × Delay (immediate, delayed) × Measure (list, story) layout. There were 42 results found, 25 E-Alone and 17 E+P. There were 21 list recall and 21 story recall results, and 22 immediate- and 20 delayed-testing results. There were relatively few non-null results (5 with $p < .05$, and 6 with $p < .1$ or mixed; 26% of all results) compared to the number of null results (31; 74% of all results); non-null results were distributed equally between E-Alone and E+P. (The table also notes the use of alternate forms and counterbalancing; we will return to this in the Discussion.) This pattern of findings might be represented by a simple null model: all results are insignificant except some random deviations. However, further inspection reveals that the distribution of non-null results contains regularities inconsistent with the null model. There are several broad regularities: smaller and shorter studies with younger participants

appear to produce relatively more positive and fewer negative results in comparison to larger and longer studies with older participants.

There are several specific regularities relevant to the focus of this paper. First, note that all six negative outliers are due to list recall (there are also two positive list recall results). Second, note that all three story recall outliers are in the immediate E-Alone condition, and are positive (also note that all three are short, small studies with younger participants). Third, most of the outliers are singletons and therefore less likely to be of interest (i.e., worthy of further study) in comparison to clusters of outliers (Aggarwal, 2013), but of the four micro-clusters of two or more outliers, one occurred with E-Alone in story immediate, while three occurred with E+P in list recall, both immediate and delayed. The distribution suggests that if an HT LRCT produces a non-null result for verbal free recall, it is likely to be negative if list recall, and positive (and immediate and with E-Alone) if story recall; this parallels the gist-related difference between story and list recall as modeled by FTT. In the Discussion, we will further consider the distribution of non-null results from the literature, along with our study's findings.

The Current Study

The FTT framework suggests that there are several possible patterns of HT effects in terms of the individual constructs of the theory in the performance of list and recall tasks (verbatim, gist, and metacognitive review). Keeping in mind that both list and story recall require both verbatim and gist processing, an improvement or impairment in either one could improve or worsen scores, respectively, on both tests (although to different degrees). However, bias of metacognitive review in favor of verbatim or gist could improve performance on one test and impair performance on the other. Table 2 lays out the various possibilities.

This paper is based on previously unanalyzed data collected from 1996 to 2004 as part of a WHI ancillary study. Our LRCT involved 100 post-menopausal women, directly comparing the effects of HT versus placebo on the age-related decline of immediate and delayed list and story recall. The overarching hypothesis is that the rate of age-related decline in verbal free recall will be altered by HT. More specifically, we hypothesize that the alteration due to HT (if any) in the rate of decline will be modulated by the following factors: recall type (list vs. story); time of recall (immediate vs. delayed) and HT regimen (E-Alone vs. E+P). Based on our literature tabulation, we also hypothesize that HT effects on story recall, if any, will be positive, with immediate testing, and in E-Alone; effects on list recall, if any, will be in E+P and will be negative. Note that while dividing the HT group into regimen-specific subgroups results in less power for those comparisons, we will include subgroup analyses because of differences (summarized above) that have been found with presence versus absence of progestin.

Methods

All participants came to us via referral from the UC Davis WHIMS office. One hundred participants had been randomly assigned to treatment or placebo before referral. Women

with an intact uterus were given either estrogen + progestin (E+P) or placebo, and women with no uterus were given estrogen (E-Alone) or placebo. Participants were eventually unblinded as 56 placebo (regardless of uterine status) and 44 HT participants (16 E-Alone, 28 E+P). All data were collected in compliance with UC Davis Institutional Review Board guidelines. Our initial session occurred within eight weeks after HT began (median 21.5 days), with follow-up sessions at intervals of about a year (median 395.9 days). Appendix A contains details regarding the WHI and WHIMS projects and the recruitment of our subjects.

Participant Characteristics

Participant demographics for the HT and placebo groups are displayed in Table 3. First-session age ranged from 65.4 to 79.4 years. Participants referred to us by WHIMS were negative for suspected cognitive impairment according to the Modified Mini-Mental State exam (3MS; Teng & Chui, 1987), using cutoffs of 72/100 for women with 0 to 8 years of education and 76/100 with 9 or more years (Shumaker et al., 1998). As a demographic, we administered the American version of the Nelson Adult Reading Test (Grober, Sliwinski, & Korey, 1991) in the first session in order to estimate participants' verbal intelligence (VIQ). There were no significant differences among the groups for the 3MS, VIQ, or any of the demographic variables. Because working memory ability could be related to episodic memory ability, the placebo and both HT subgroups were compared with ANOVA in the first session for digit span backwards (Wechsler, 1981) and computation span (Salthouse & Babcock, 1991); no differences were found, $F(2, 81) = 1.726$, $p = .183$ and $F(2, 81) = 1.420$, $p = 0.247$, respectively.

Hormone Treatment

Participants were randomly assigned to receive, in tablet form, either a placebo, or a preparation of 0.625 mg conjugated equine estrogens (CEE) with 2.5 mg medroxyprogesterone acetate (E+P group), or 0.625 mg CEE alone (E-Alone group). Neither participants nor investigators were informed about the random assignments until the end of the study.

Test Materials

All sessions included a list recall test and a story recall test, each with multiple alternate forms that were used in a counterbalanced fashion (no form repeated until all forms had been used). For list recall we used the University of Southern California Repeatable Episodic Memory Test (USC-REMT), which has seven alternate 15-item lists of semantically unrelated, high-frequency words that do not differ significantly in performance based on validation studies (Parker, Eaton, Whipple, Heseltine, & Bridge, 1995; Parker, Landau, Whipple, & Schwartz, 2004). As per the standard protocol for the USC-REMT, we presented the word lists three times orally, in different orders, and with an immediate recall trial after each presentation. After a 20-min interval filled with unrelated session activity, the delayed recall trial was administered. Our list recall measures were the mean correct (excluding repetitions) over the three immediate trials, and the total correct from the delayed trial (see Parker et al., 1995, for the administration and scoring protocol).

Our story recall measures were the immediate and delayed story recall subtests of the widely used Rivermead Behavioral Memory Test (Wilson, Cockburn, & Baddeley, 1991); there were four alternate forms. Based on validation studies (Wilson, Cockburn, Baddeley, & Hiorns, 1989), performance on Form A of the full 13-subtest battery is significantly correlated with the other three forms of the battery; there is no published comparison of the alternate forms of the story recall subtests per se. After listening to a story being read aloud, participants were asked to recall as much of it as they could (immediate recall). Then, after a 15-min delay, filled with unrelated session activity, they were again asked to recall as much as they could (delayed recall). The scores were the number of standardized “idea units” or discrete story segments recalled for the immediate and delayed recall trials, out of a maximum of 21. Participants received one point for a fully correct idea unit or synonym, and one-half point for a partially correct idea unit or synonym (see Wilson et al., 1991, for the administration and scoring protocol).

Results

Since timing precision is critically important in a longitudinal study such as this one, we first will present information regarding potentially confounding variability from session to session in: time of test, participant age at time of test, and quantity of medication taken at time of test. Next, we specify a statistical model designed to control for this variability. Finally, we conduct per-measure analyses of the rate of change in performance for HT relative to placebo over time in the four memory measures; these analyses are repeated treating E-Alone and E+P separately. A significance level of .05 was adopted in all analyses of our data.

Compliance, Attrition, and Timing Issues

Although affiliated with the WHI clinical trial, our study was not concerned with clinical aspects of HT. Therefore, in order to focus on actual effects of HT on our behavioral measures, we performed per-protocol rather than intent-to-treat analyses. We deemed a session to be in compliance when the participant had taken the assigned study medication for at least 90% of the days of the study up to the time of the session. (Note that in some cases, earlier sessions for a given participant might be excluded but later ones allowed.) Because only four women were in compliance in Session 8, that session was dropped from the analysis (see Appendix B). Note that women could withdraw or pause their participation at any point, but the WHI staff discontinued women either temporarily or permanently who developed certain serious medical conditions (Writing Group for the Women’s Health Initiative Investigators, 2002).

Figure 1 displays the age to the day of each participant at the time of data collection. Careful inspection of this figure suggests that while the three groups initially did not differ in age, attrition (either unavailability for testing or not meeting the 90% criterion for compliance) occurred in slightly different age ranges from group to group. The first notable shift is a “flat” interval for E+P such that their mean age did not change from year 2 to 3. This shifted E+P from the middle position in mean age to the youngest position. There is a similar flat interval between year 4 and 5 for E-Alone. A flat interval arrives when enough older

participants disproportionally are lost to cancel out a year of increase in mean age. The timing and sequence of these flat intervals suggest that they could be related to the premature termination of the E+P branch of WHI followed two years later by the premature termination of the E-Alone branch (see Appendix A). In contrast, during years 1 to 4, the E-Alone group's mean age appears to increase faster than one year per calendar year, indicating that dropping out occurred disproportionately among younger group members during that interval. The E-Alone group started slightly older than the others, and uneven attrition increased this difference during most of the study.

In order to determine whether the pattern of attrition may have been systematically related to the study results, we used ANOVA to compare the average number of in-compliance sessions between groups (E-Alone, E+P, and placebo) and found no significant differences among the groups, $F(2,81) < 1$. Next, we correlated the number of in-compliance sessions with the following variables: starting age, years of education, AMNART performance, and 3MS. For all subjects together, none of these correlations were significant. Taking each of the three groups separately, only initial age was significantly correlated with number of in-compliance sessions, and only for the E+P group, $r = .55$, $p = .006$. Next, we correlated the number of in-compliance sessions with the baseline scores for the four performance variables (immediate and delayed list and story recall). None of the correlations were significant for all groups combined or for the groups individually. Finally, we computed the per-subject slopes for each of the four performance variables and correlated them with the number of in-compliance sessions. For all subjects together, two of those correlations were significant: immediate story recall slope, $r = -.35$, $p = .001$, and delayed story recall slope, $r = -.45$, $p < .001$; neither list recall slope correlation was significant. This pattern was found for each treatment group, except that immediate story recall slope was not significant for the placebo group and immediate list recall slope was correlated significantly with the number of in-compliance sessions for the E-Alone group, $r = -.65$, $p = .006$. See Appendix B for a table presenting the number of participants per year who were tested and who met the 90% criterion for compliance.

The Statistical Model

Due to the substantial variation in time of test relative to the nominal one-year protocol as apparent in Figure 1, we used duration of participation in fractional years to the day as the time variable in all analyses. In order to control for potential age bias due to attrition, we used per-session age in fractional years to the day as a covariate in all analyses. An additional compliance-related potential confound has to do with the quantity of study medication taken as of each session. As noted, participants began medication a median of 21.5 days before the first session; some paused and later resumed medication during the study; after each branch of WHI was terminated, some continued to participate in our testing, but did not take the medication. In order to control for this factor, we also included a per-session quantity of medication covariate, being the number of days study medication was taken as of the session, converted to fractional years.

In the following linear mixed effect (LME) analyses, models were fit for each memory measure separately, linear change in performance being modeled over continuous time with

continuous age and quantity of medication taken as covariates, with random intercept and slope for each subject. Because residual analysis and visual inspection of trajectories did not indicate a need for non-linear terms, we did not include higher-order models in the analyses. Models were fit using R's *lme* function (Pinheiro, Bates, DebRoy, Sarkar, & R Core Team, 2017). See Appendix C for a table of raw means and standard deviations over seven years for the four verbal memory variables in the placebo group and two HT subgroups.

HT Group Relative to Placebo

Turning to our specific hypotheses, we considered changes in performance by the HT group relative to that of the placebo group via a Time × Group interaction term with the placebo group as baseline. Table 4 contains point estimates, 95% confidence intervals, and *p*-values for this analysis. Per session differences between HT and placebo are plotted for each dependent variable in the leftmost column of Figure 2, along with lines representing the LME estimates of slope. Note that for the HT group, a positive slope indicates a slower rate of decline relative to placebo, while a negative slope indicates a more rapid relative rate of decline.

The HT group's rate of decline in immediate list recall performance was significantly more rapid than that of the placebo group. In contrast, their rate of decline in immediate story recall performance was significantly slower than that of the placebo group. The HT group's rate of decline in delayed story performance was slower than that of the placebo group, while the estimated rate of decline for delayed list recall was close to flat for both groups; neither of these group differences approached significance. These findings support our hypotheses that the rate of age-related decline in verbal free recall is altered by HT and that the alteration in the rate of decline due to HT is modulated by recall task and by time of recall; they are also similar to the pattern of non-null results in the literature tabulation.

E-Alone and E+P Subgroups Relative to Placebo

In order to determine whether the HT subgroups show different responses to HT, we performed a second set of LME analyses (see Table 5 for point estimates, 95% confidence intervals, and *p*-values) that was identical to the first except that the Time × Group interaction term, still with placebo as baseline, now contained E-Alone and E+P levels. The LME slope estimates and per session differences from placebo for E-Alone and E+P are plotted in the center and rightmost columns, respectively, of Figure 2.

For the E-Alone subgroup, the rate of decline in immediate list performance was significantly more rapid than for the placebo group, while the rate of decline in immediate story performance was significantly slower. Both delayed list and delayed story recall showed slightly slower rates of decline for the E-Alone subgroup relative to placebo; neither approached significance. For the E+P subgroup, the relative rate of decline in immediate list recall was faster, in immediate story recall was slower, in delayed list recall was faster, and in delayed story recall was slower; none approached significance. These findings support our hypothesis that HT regimen modulates response to treatment. Generally, the overall HT effects appear to be augmented in the E-Alone subgroup and attenuated in the E+P subgroup.

Discussion

Although many LRCTs have used either list or story recall or both, this research is the first LRCT to explicitly evaluate possible differences in the effects over time of HT on list versus story recall. To provide context for our study, we presented a thorough tabulation of all prior LRCTs that published a comparison of HT versus placebo for either list or story recall. Our study involved the collection of performance data on immediate and delayed list and story recall over up to eight annual test sessions from participants who received either E-Alone, E+P, or placebo. To minimize the impact of attrition over the eight years of the study, we used statistical methods that maximized included data, and controlled statistically for potential bias due to variability in participant age, quantity of medication, and timing of annual test sessions.

Summary of Principal Findings

Our primary hypothesis was that there would be different effects over time of postmenopausal HT on the age-related decline of list recall versus story recall. LME analyses revealed significant differences in the linear rate of change in performance for the HT group relative to placebo in two cases: significantly faster decline in immediate list recall, and significantly slower decline in immediate story recall. As indicated in Table 2 this effect is consistent with an increase in gist bias. Since list recall emphasizes verbatim information, increasing gist bias could impair performance. However, since story recall emphasizes gist, an increase in gist bias could improve performance.

A secondary hypothesis was that there would be differential effects related to the time between study and test, i.e., immediate versus delayed testing. Given that we found significant effects only in immediate recall, delayed recall may reduce or eliminate the effect of estrogen found in immediate recall. This is concordant with the observation that delay increases gist bias, as compensation for decay of the verbatim trace.

Another secondary hypothesis was that there would be differences in outcome due to HT formulation, notably estrogen alone versus estrogen plus progestin. When we repeated the LME analysis on separate E-Alone and E+P subgroups, the difference between the tasks remained for the E-Alone group, but not for E+P. This finding could be indicative of progestin muting the effect of estrogen on verbal recall. Further research is required to replicate and extend these findings.

A Conjecture Regarding HT Effects on List and Story Recall

From an FTT perspective, the most likely account of our results is an increase in gist bias resulting from estrogen. As shown in Table 2, although increased verbatim bias would affect list recall more and story less whereas increased gist bias would affect story recall more and list less, both changes would be in the same direction in both tasks. However, we observed a dissociation between list and story, in line with a list/story dissociation visible in the literature for list and story recall in isolation (Table 1). This can be accounted for in FTT terms by positing an HT-related increase in gist bias during metacognitive review, since that would affect the two tasks in the directions found, i.e., relatively improved story recall and

relatively impaired list recall. An HT-related increase in gist bias could also be associated with the HT-related increase in prefrontal activation during verbal memory encoding and/or retrieval found in numerous fMRI studies (reviewed by Comasco et al., 2014), since this could provide support for gist-based processing.

Why did we find significant effects only for immediate list and immediate story? In the FTT framework, it is not surprising for delay to produce gist bias on its own, because as the verbatim trace is effaced with time, participants naturally rely more on gist information (e.g., Reyna & Mills, 2007); this would tend to reduce the differential estrogen-related effect. The story recall data from our literature review shows a tendency similar to our data in that the positive effect relative to placebo is limited to immediate recall. However, the list recall data from our literature review shows a similar negative tendency in both immediate and delayed recall.

A Methodological Issue Regarding List Recall

There are some interesting regularities in the non-null results for list recall, in the literature survey (Table 1), that could help clarify both differences between our study and the literature, as well as certain aspects of the distribution of results in the literature. First, note that no list recall study counterbalanced alternate forms. This is an important qualitative difference from our study. Next, note that there are only two patterns for the use (or not) of alternate forms: aa* (only one form used two or more times) or a*b (Form A used one or more times followed by Form B). Also note that the two aa* non-null results are positive, while the six a*b results are negative. In every case, there is a possibility that systematic inter-session effects due to repetition of the same form may have affected the outcome.

In spite of delays between tests ranging from two months to a year for the studies in Table 1, it is likely that some sort of memory trace remained between sessions. For example, young to middle-aged healthy adults recalled 53% of list and 48% of story items, after a four-week delay, in Mameniskiene, Jatuzis, Kaubrys, and Budrys (2006), and 52% of story items, after an eight-week delay, in Blake, Wroe, Breen, and McCarthy (2000). Therefore, it is likely that some information from the previous session(s) was available to participants in the Table 1 studies, mostly as gist, since gist tends to be retained better than verbatim information. If estrogen increases gist bias, then HT participants might show improved performance relative to placebo when the lingering gist information was relevant (i.e., when the same form was used) and impaired performance when it was not relevant (different form). In cases where alternate forms were used, in the Table 1 list recall studies, they were only used in the last session, which would tend to produce a negative result for HT participants relative to placebo. In contrast, if the same form was used in every session, this might produce a positive HT outcome. Since the current study counterbalanced forms across participants, no systematic effect of gist lingering from the previous session could have affected our results.

Implications of This Study

Given the variety of conflicting results in the literature, it is difficult to know how generalizable the current results will be. Published studies differ along myriad dimensions, and, as we have described in detail, the present study differs from all of them in a number of

ways. Future research will determine which combination of parameters produced the dissociation we found between list and story recall.

Although no other LRCT has explicitly contrasted the effects of HT on list versus story recall, our results share certain aspects with the list and story results from the literature presented in Table 1. The most frequent result is null; in effect, the non-null results are outliers. The field of data-mining has produced methods for analyzing outliers (Aggarwal, 2013), including the strategy of identifying clusters and other patterns, with the goal of identifying and understanding data that is so different from the rest that it may have arisen through a different mechanism. By definition, the appearance of such patterns does not constitute statistically significant tests of specific hypotheses; instead, their utility lies more in the area of hypothesis generation.

Our intended audience for this paper is researchers interested in the effects of estrogen on episodic memory and/or episodic memory processes in adulthood and aging. This focus provides the impetus for taking a careful look at the pattern of outlier findings in the literature. There was, indeed, a good match between our results and the literature's non-null results for story recall in that our positive effect for immediate recall only for E-Alone fell into a small cluster of positive results. The correspondence was less strong for list recall: while our only significant result was negative and E-Alone, there are a number of reports of both positive and negative results for E+P participants. We hope that future research will clarify what processes underlie both the null effects reported in the majority of studies on verbal free recall and HT, and also the tantalizing regularities of the effects found in the minority of studies, including this one.

Limitations

Although this study, with 100 total participants, was of moderate size, it is a limitation of our LRCT that there were only 16 participants in the E-Alone subgroup, which turned out to be the locus of the most interesting effects. This resulted from how the local WHIMS administration referred participants to us (see Appendix A for details), and reflects the proportion of WHI participants who had had or had not had an hysterectomy: nationally, 4,532 women were enrolled in the E+P branch of WHIMS, while only 2,947 enrolled in the E-Alone branch (Coker et al., 2010). Moreover, this study had slightly more placebo participants than active HT participants, an unpredictable outcome of the study randomization. That said, even our E-Alone subgroup is at least as large as the treatment group Ns in a number of published LRCTs on HT in our literature survey (16, Alhola et al., 2010; 14, LeBlanc, Neiss, Carello, Samuels, & Janowsky, 2007; 16, Linzmayer et al., 2001; 17, Maki et al., 2009; 10/9, Phillips & Sherwin, 1992; 19, Schiff, Bulpitt, Wesnes, & Rajkumar, 2005; 10, Sherwin, 1988; 6, Sherwin & Phillips, 1990; 10–13, Wolf, Heinrich, Hanstein, & Kirschbaum, 2005).

The rate of attrition is also a limitation of the LRCT: by Year 7, only 20% of the original 100 participants were compliant with their assigned protocol (see Figure 1 and Appendix B). While this limits the generalizability of the study, our LME analyses of comparative (treatment vs. placebo) rate of change in performance over time, maximized utilization of all available data. Finally, as a result of uneven attrition, we found a shift over time in the age

distribution among groups; however, this was compensated for by the use of continuous age as a covariate in all LME analyses.

Future Directions

Since this LRCT is the first to specifically explore a different effect of HT on list versus story recall, its greatest utility may be in the area of hypothesis generation. For example, it would be helpful to include recognition tests designed to elicit false memories, which are particularly useful as a measure of gist bias (cf. Dennis, Bowman, & Peterson, 2014; Paige, Cassidy, Schacter, & Gutchess, 2016). Also, free list recall tests vary in terms of how much gist information is available. While some studies have compared age-related changes in performance on structured or categorized lists versus unstructured or random lists (e.g., Dixon et al., 2004), it would be interesting to determine the extent to which estrogen modulates those changes. Functional brain imaging studies have considered the effects of estrogen on verbal memory (e.g., Persad et al., 2009), but so far have not manipulated such variables as list versus story recall. Moreover, given that experiments showing an impact of estrogen on gist bias (e.g., Petersen et al., 2014) involved memory for emotion-laden material, it would be interesting to include positive, negative, and neutral words and/or stories in future research on estrogen and verbal memory. Finally, while our study only showed a difference between the specific E-Alone and E+P formulations chosen by the WHI, it would also be of interest to determine whether other related variations such as uterine or ovarian status and/or differences in estrogen formulation (including type of estrogen, size of dose, and whether other sex hormones are included) are associated with gist-related (or other) differences in the effects of HT on verbal memory.

Acknowledgments

The project described was supported by the National Center for Advancing Translational Sciences (NCATS), National Institutes of Health (NIH), through grant #UL1 TR001860. We thank John Robbins, M.D., and the staff of the UC Davis Women's Health Initiative (WHI) Clinical Center for their assistance, and we thank the WHI Memory Study (WHIMS) for referring participants to us. The findings reported here are distinct from and do not involve data from the WHI and WHIMS datasets. Portions of this research were presented as a poster at the Cognitive Neuroscience Society Annual Meeting, April 16, 2013, in San Francisco.

Appendix A: Details on Participant Recruitment

Study Background

The Women's Health Initiative (WHI) estrogen study was a double-blind placebo-based longitudinal study of the effects of HT in women 65+ years old (Women's Health Initiative Steering Committee, 2004; Writing Group for the Women's Health Initiative Investigators, 2002). In the WHI study, women with an intact uterus were given either estrogen + progestin (E+P) or placebo, and women with no uterus were given estrogen (E-Alone) or placebo, with annual follow-ups. A subset of 7,340 WHI participants was enrolled in an ancillary study focused on dementia, the WHI Memory Study (WHIMS). These participants were tested annually for changes in global cognition with the Modified Mini-Mental State Examination (3MS; Teng & Chui, 1987). The WHI E+P branch was halted prematurely in 2002 after an average of 5.2 years of follow-up due to increased risk of coronary heart disease and invasive breast cancer, and the E-Alone branch was stopped in 2004, in part

because of increased risk of stroke (Women's Health Initiative Steering Committee, 2004; Writing Group for the Women's Health Initiative Investigators, 2002).

Soon after the WHIMS study began, the first author received WHI authorization to conduct an ancillary study, "Longitudinal assessment of memory functioning in an elderly normal subsample of the Women's Health Initiative clinical trial enrollees". This ancillary study, which included both list and story verbal recall tasks, assessed participants from September of 1996 through July of 2004.

Recruitment into Current Study

One hundred WHIMS participants enrolled in our WHI ancillary study and came to at least one session, having been assigned per hysterectomy status to E-Alone or E+P, and assigned randomly to placebo versus treatment. Note that at the time this was done, WHIMS researchers expected to find no cognitive differences due to E-Alone versus E+P (e.g., Shumaker et al., 1998), so it was expected that all of our study participants receiving HT would be combined into a single treatment group. (Note that due to progestin-related differences reported since then, the current paper analyses the combined HT group as well as separate E-Alone and E+P subgroups). As a result, there was no attempt to balance the size of the two subgroups. Criteria for acceptance into our study were that participants be enrolled in WHIMS and be willing to undergo a total of seven additional annual assessments. The WHIMS criteria (Shumaker et al., 1998) required active enrollment in WHI and also included a specific age range (65 to 79 years) plus readiness to participate in cognitive assessments for four to six years. The WHI criteria were quite open: they accepted any postmenopausal woman age 50 to 79 who resided near a study center, except for a set of specific exclusion criteria (Appendix A of Shumaker et al., 1998) such as dementia, mental illness including depression, dependence on alcohol or drugs, and a number of serious medical problems. All participants came to us via referral from the UC Davis WHI Clinical Center, where they were asked whether they were willing to volunteer for an additional cognitive study. Before the WHI Study of Cognitive Aging (WHISCA; Resnick et al., 2004) began, all volunteers were referred to us; after WHISCA began, referrals were made to both studies with priority given to WHISCA. We did not receive information about placebo participants' hysterectomy history.

Appendix B: Attrition and Compliance: number of participants per year of study

Year	Placebo		Estrogen		Estrogen + Progestin	
	Compliant	Total	Compliant	Total	Compliant	Total
1	56	56	16	16	28	28
2	46	48	13	15	21	22
3	42	49	11	15	22	24
4	32	40	10	14	19	23
5	29	42	8	12	18	22

Year	Placebo		Estrogen		Estrogen + Progestin	
	Compliant	Total	Compliant	Total	Compliant	Total
6	24	37	7	12	14	20
7	10	32	6	9	4	14
8	1	6	3	3	0	4

Note: Total was defined as the number of participants assessed in the given year. Compliant was defined as the number of participants having taken at least 90 % of the assigned study medication (placebo, estrogen, or estrogen + progestin) from Year 1 to the day of the assessment. Year 8 was excluded from all analyses.

Appendix C: Mean raw verbal recall testing results for Sessions 1–7

List (Immediate)							
Group	**1**	**2**	**3**	**4**	**5**	**6**	**7**
Placebo	7.7 (1.7)	7.2 (1.6)	7.5 (1.7)	7.4 (1.8)	7.0 (1.8)	8.1 (1.8)	7.5 (1.6)
E-Alone	7.1 (1.9)	7.2 (1.3)	6.7 (1.8)	6.3 (1.9)	6.6 (1.7)	6.6 (1.0)	7.3 (0.9)
E+P	7.4 (1.6)	7.4 (1.5)	7.4 (1.5)	7.3 (1.3)	7.2 (1.4)	7.1 (1.5)	9.8 (0.8)
List (Delayed)							
Group	**1**	**2**	**3**	**4**	**5**	**6**	**7**
Placebo	7.3 (2.9)	7.0 (2.8)	6.6 (2.5)	6.4 (2.5)	6.1 (2.7)	7.2 (3.0)	6.8 (2.4)
E-Alone	6.6 (2.4)	6.5 (2.3)	6.4 (3.0)	6.6 (3.2)	6.6 (2.6)	6.1 (2.8)	8.0 (1.7)
E+P	7.1 (2.5)	6.8 (2.6)	7.0 (2.1)	6.7 (2.3)	6.9 (2.5)	6.5 (2.3)	10.0 (3.2)
Story (Immediate)							
Group	**1**	**2**	**3**	**4**	**5**	**6**	**7**
Placebo	8.4 (2.7)	8.5 (2.5)	8.3 (3.2)	7.8 (2.9)	6.5 (2.9)	7.6 (2.7)	6.4 (2.9)
E-Alone	7.2 (2.5)	8.7 (2.6)	8.1 (2.7)	5.8 (2.7)	7.2 (2.2)	8.2 (1.8)	8.8 (2.7)
E+P	8.0 (3.0)	7.5 (2.8)	8.2 (3.1)	6.8 (3.1)	7.2 (3.3)	8.1 (2.8)	9.5 (2.6)
Story (Delayed)							
Group	**1**	**2**	**3**	**4**	**5**	**6**	**7**
Placebo	7.2 (2.8)	7.4 (2.7)	7.5 (3.2)	7.2 (3.1)	6.1 (3.1)	6.9 (2.8)	5.7 (2.7)
E-Alone	5.8 (2.2)	7.2 (3.1)	7.1 (2.7)	5.1 (2.4)	6.6 (2.3)	6.7 (1.6)	7.3 (2.3)
E+P	6.9 (3.1)	6.4 (3.1)	7.4 (3.3)	6.2 (3.0)	6.6 (3.2)	7.5 (2.4)	8.0 (3.1)

Note: Numbers in each cell are raw means, followed by the standard deviation in parentheses, for all in-compliance participants at each session. Note that t-tests found no significant differences in Session 1 among the groups for any measure, although in Story (Delayed), E-Alone trended less than placebo, $t(70) = 1.84$, $p = .066$.

References

Aggarwal CC (2013). Outlier Analysis. New York: Springer.

Alhola P, Tuomisto H, Saarinen R, Portin R, Kalleinen N, & Polo-Kantola P (2010). Estrogen + progestin therapy and cognition: a randomized placebo-controlled double-blind study. Journal of Obstetrics and Gynaecology Research, 36, 796–802. doi: 10.1111/j.1447-0756.2010.01214.x [PubMed: 20666948]

Almeida OP, Lautenschlager NT, Vasikaran S, Leedman P, Gelavis A, & Flicker L (2006). A 20-week randomized controlled trial of estradiol replacement therapy for women aged 70 years and older:

effect on mood, cognition and quality of life. Neurobiology of Aging, 27, 141–149. doi: 10.1016/j.neurobiolaging.2004.12.012 [PubMed: 16298249]

Blake RV, Wroe SJ, Breen EK, & McCarthy RA (2000). Accelerated forgetting in patients with epilepsy: Evidence for impairment in memory consolidation. Brain, 123, 472–483. doi: 10.1093/brain/123.3.472 [PubMed: 10686171]

Brainerd CJ, Reyna VF, & Ceci SJ (2008). Developmental reversals in false memory: a review of data and theory. Psychological Bulletin, 134, 343–382. doi: 10.1037/0033-2909.134.3.343 [PubMed: 18444700]

Coker LH, Espeland MA, Hogan PE, Resnick SM, Bryan RN, Robinson JG, … Shumaker SA (2014). Change in brain and lesion volumes after CEE therapies: The WHIMS-MRI studies. Neurology, 82, 427–434. doi: 10.1212/WNL.0000000000000079 [PubMed: 24384646]

Coker LH, Espeland MA, Rapp SR, Legault C, Resnick SM, Hogan P, … Shumaker SA (2010). Postmenopausal hormone therapy and cognitive outcomes: the Women's Health Initiative Memory Study (WHIMS). Journal of Steroid Biochemistry and Molecular Biology, 118, 304–310. doi: 10.1016/j.jsbmb.2009.11.007 [PubMed: 19932751]

Comasco E, Frokjaer VG, & Sundstrom-Poromaa I (2014). Functional and molecular neuroimaging of menopause and hormone replacement therapy. Frontiers in Neuroscience, 8, 388. doi: 10.3389/fnins.2014.00388 [PubMed: 25538545]

Cooper PV, Numan BK, Crosson B, & Velozo CA (1989). Story and list recall tests as measures of verbal memory in a head-injured sample. Neuropsychology, 3, 1–8. doi: 10.1037/0894-4105.3.1.1

Delis DC, Cullum CM, Butters N, Cairns P, & Prifitera A (1988). Wechsler Memory Scale-Revised and California Verbal Learning Test: Convergence and divergence. The Clinical Neuropsychologist, 2, 188–196. doi: 10.1080/13854048808520100

Dennis NA, Bowman CR, & Peterson KM (2014). Age-related differences in the neural correlates mediating false recollection. Neurobiology of Aging, 35, 395–407. doi: 10.1016/j.neurobiolaging.2013.08.019 [PubMed: 24094578]

Dixon RA, Wahlin Å, Maitland SB, Hultsch DF, Hertzog C, & Bäckman L (2004). Episodic memory change in late adulthood: Generalizability across samples and performance indices. Memory and Cognition, 32, 768–778. doi: 10.3758/BF03195867 [PubMed: 15552354]

Gorenstein C, Renno J Jr., Vieira Filho AH, Gianfaldoni A, Goncalves MA, Halbe HW, … Demetrio FN (2011). Estrogen replacement therapy and cognitive functions in healthy postmenopausal women: a randomized trial. Archives of Women's Mental Health, 14, 367–373. doi: 10.1007/s00737-011-0230-6

Grober E, Sliwinski M, & Korey SR (1991). Development and validation of a model for estimating premorbid verbal intelligence in the elderly. Journal of Clinical and Experimental Neuropsychology, 13, 933–949. doi: 10.1080/01688639108405109 [PubMed: 1779032]

Helmstaedter C, Wietzke J, & Lutz MT (2009). Unique and shared validity of the "Wechsler logical memory test", the "California verbal learning test", and the "verbal learning and memory test" in patients with epilepsy. Epilepsy Research, 87, 203–212. doi: 10.1016/j.eplepsyres.2009.09.002 [PubMed: 19782537]

Hogervorst E, & Bandelow S (2010). Sex steroids to maintain cognitive function in women after the menopause: a meta-analyses of treatment trials. Maturitas, 66, 56–71. doi: 10.1016/j.maturitas.2010.02.005 [PubMed: 20202765]

Joffe H, Hall JE, Gruber S, Sarmiento IA, Cohen LS, Yurgelun-Todd D, & Martin KA (2006). Estrogen therapy selectively enhances prefrontal cognitive processes: a randomized, double-blind, placebo-controlled study with functional magnetic resonance imaging in perimenopausal and recently postmenopausal women. Menopause, 13, 411–422. doi: 10.1097/01.gme.0000189618.48774.7b [PubMed: 16735938]

Koutstaal W, & Schacter DL (1997). Gist-based false recognition of pictures in older and younger adults. Journal of Memory and Language, 37, 555–583. doi: 10.1006/jmla.1997.2529

Lamar M, Charlton R, Zhang A, & Kumar A (2012). Differential associations between types of verbal memory and prefrontal brain structure in healthy aging and late life depression. Neuropsychologia, 50, 1823–1829. doi: 10.1016/j.neuropsychologia.2012.04.007 [PubMed: 22564447]

LeBlanc ES, Neiss MB, Carello PE, Samuels MH, & Janowsky JS (2007). Hot flashes and estrogen therapy do not influence cognition in early menopausal women. Menopause, 14, 191–202. doi: 10.1097/01.gme.0000230347.28616.1c [PubMed: 17194963]

Lezak MD, Howieson DB, Bigler ED, & Tranel D (2012). Neuropsychological Assessment (5th ed.). New York: Oxford University Press.

Linzmayer L, Semlitsch HV, Saletu B, Böck G, Saletu-Zyhlarz G, Zoghlami A, … Grünberger J (2001). Double-blind, placebo-controlled psychometric studies on the effects of a combined estrogen-progestin regimen versus estrogen alone on performance, mood and personality of menopausal syndrome patients. Arzneimittel-Forschung, 51, 238–245. doi: 10.1055/s-0031-1300030 [PubMed: 11304940]

Maki PM (2012). Minireview: effects of different HT formulations on cognition. Endocrinology, 153, 3564–3570. doi: 10.1210/en.2012-1175 [PubMed: 22673228]

Maki PM, Gast MJ, Vieweg AJ, Burriss SW, & Yaffe K (2007). Hormone therapy in menopausal women with cognitive complaints. A randomized, double-blind trial. Neurology, 69, 1322–1330. doi: 10.1212/01.wnl.0000277275.42504.93 [PubMed: 17893293]

Maki PM, Rubin LH, Fornelli D, Drogos L, Banuvar S, Shulman LP, & Geller SE (2009). Effects of botanicals and combined hormone therapy on cognition in postmenopausal women. Menopause, 16, 1167–1177. doi: 10.1097/gme.0b013e3181ace484 [PubMed: 19590458]

Maki PM, & Sundermann E (2009). Hormone therapy and cognitive function. Human Reproduction Update, 15, 667–681. doi: 10.1093/humupd/dmp022 [PubMed: 19468050]

Mameniskiene R, Jatuzis D, Kaubrys G, & Budrys V (2006). The decay of memory between delayed and long-term recall in patients with temporal lobe epilepsy. Epilepsy Behav, 8, 278–288. doi: 10.1016/j.yebeh.2005.11.003 [PubMed: 16359927]

Moscovitch M, & Winocur G (2002). The frontal cortex and working with memory In Stuss DT & Knight RT (Eds.), The Frontal Lobes (pp. 188–209). Oxford: Oxford University Press.

Mulligan NW (2005). Total retrieval time and hypermnesia: investigating the benefits of multiple recall tests. Psychological Research, 69, 272–284. doi: 10.1007/s00426-004-0178-5 [PubMed: 15168120]

Nielsen SE, Ahmed I, & Cahill L (2013). Sex and menstrual cycle phase at encoding influence emotional memory for gist and detail. Neurobiology of Learning and Memory, 106, 56–65. doi: 10.1016/j.nlm.2013.07.015 [PubMed: 23891713]

Nielsen SE, Ertman N, Lakhani YS, & Cahill L (2011). Hormonal contraception usage is associated with altered memory for an emotional story. Neurobiology of Learning and Memory, 96, 378–384. doi: 10.1016/j.nlm.2011.06.013 [PubMed: 21740976]

Paige LE, Cassidy BS, Schacter DL, & Gutchess AH (2016). Age differences in hippocampal activation during gist-based false recognition. Neurobiology of Aging, 46, 76–83. doi: 10.1016/j.neurobiolaging.2016.06.014 [PubMed: 27460152]

Park DC, & McDonough IM (2013). The dynamic aging mind: Revelations from functional neuroimaging research. Perspectives on Psychological Science, 8, 62–67. doi: 10.1177/1745691612469034 [PubMed: 26172252]

Park DC, & Reuter-Lorenz P (2009). The adaptive brain: aging and neurocognitive scaffolding. Annual Review of Psychology, 60, 173–196. doi: 10.1146/annurev.psych.59.103006.093656

Parker ES, Eaton EM, Whipple SC, Heseltine PNR, & Bridge TP (1995). University of Southern California Repeatable Episodic Memory Test. Journal of Clinical and Experimental Neuropsychology, 17, 926–936. doi: 10.1080/01688639508402441 [PubMed: 8847398]

Parker ES, Landau SM, Whipple SC, & Schwartz BL (2004). Aging, recall and recognition: A study on the sensitivity of the University of Southern California Repeatable Episodic Memory Test (USC-REMT). Journal of Clinical and Experimental Neuropsychology, 26, 428–440. doi: 10.1080/13803390490510130 [PubMed: 15512931]

Persad CC, Zubieta JK, Love T, Wang H, Tkaczyk A, & Smith YR (2009). Enhanced neuroactivation during verbal memory processing in postmenopausal women receiving short-term hormone therapy. Fertility and Sterility, 92, 197–204. doi: 10.1016/j.fertnstert.2008.04.040 [PubMed: 18692790]

Petersen N, Patihis L, & Nielsen SE (2014). Decreased susceptibility to false memories from misinformation in hormonal contraception users. Memory, 1–10. doi: 10.1080/09658211.2014.949777 [PubMed: 23731185]

Phillips SM, & Sherwin BB (1992). Effects of estrogen on memory function in surgically menopausal women. Psychoneuroendocrinology, 17, 485–495. doi: 10.1016/0306-4530(92)90007-T [PubMed: 1484915]

Pinheiro J, Bates D, DebRoy S, Sarkar D, & R Core Team. (2017). nlme: Linear and Nonlinear Mixed Effects Models [R package version 3.1–131]. Retrieved from https://CRAN.R-project.org/package=nlme

Resnick SM, Coker LH, Maki PM, Rapp SR, Espeland MA, & Shumaker SA (2004). The Women's Health Initiative Study of Cognitive Aging (WHISCA): a randomized clinical trial of the effects of hormone therapy on age-associated cognitive decline. Clinical Trials, 1, 440–450. doi: 10.1191/1740774504cn040oa [PubMed: 16279282]

Resnick SM, Espeland MA, An Y, Maki PM, Coker LH, Jackson R, … for the WHISCA Investigators. (2009). Effects of conjugated equine estrogens on cognition and affect in postmenopausal women with prior hysterectomy. Journal of Clinical Endocrinology and Metabolism, 94, 4152–4161. doi: 10.1210/jc.2009-1340 [PubMed: 19850684]

Resnick SM, Maki PM, Rapp SR, Espeland MA, Brunner R, Coker LH, … Shumaker SA (2006). Effects of combination estrogen plus progestin hormone treatment on cognition and affect. Journal of Clinical Endocrinology and Metabolism, 91, 1802–1810. doi: 10.1210/jc.2005-2097 [PubMed: 16522699]

Reyna VF, & Brainerd CJ (2011). Dual Processes in Decision Making and Developmental Neuroscience: A Fuzzy-Trace Model. Developmental Review, 31, 180–206. doi: 10.1016/j.dr.2011.07.004 [PubMed: 22096268]

Reyna VF, & Mills BA (2007). Interference processes in fuzzy-trace theory: Aging, Alzheimer's disease, and development In Gorfein DS & MacLeod CM (Eds.), Inhibition in Cognition. Washington, DC, USA: American Psychological Association.

Salthouse TA, & Babcock RL (1991). Decomposing adult age differences in working memory. Developmental Psychology, 27, 763–776.

Schiff R, Bulpitt CJ, Wesnes KA, & Rajkumar C (2005). Short-term transdermal estradiol therapy, cognition and depressive symptoms in healthy older women. A randomised placebo controlled pilot cross-over study. Psychoneuroendocrinology, 30, 309–315. doi: 10.1016/j.psyneuen.2004.08.007 [PubMed: 15694110]

Sherwin BB (1988). Estrogen and/or androgen replacement therapy and cognitive functioning in surgically menopausal women. Psychoneuroendocrinology, 13, 345–357. doi: 10.1016/0306-4530(88)90060-1 [PubMed: 3067252]

Sherwin BB, & Phillips S (1990). Estrogen and cognitive functioning in surgicallly menopausal women. Annals of the New York Academy of Sciences, 592, 474–475. doi: 10.1111/j.1749-6632.1990.tb30379.x

Shumaker SA, Legault C, Kuller L, Rapp SR, Thal L, Lane DS, … Coker LH (2004). Conjugated equine estrogens and incidence of probable dementia and mild cognitive impairment in postmenopausal Women: Women's Health Initiative Memory Study. Journal of the American Medical Association, 291, 2947–2958. doi: 10.1001/jama.291.24.2947 [PubMed: 15213206]

Shumaker SA, Legault C, Rapp SR, Thal L, Wallace RB, Ockene JK, … for the WHIMS Investigators. (2003). Estrogen plus progestin and the incidence of dementia and mild cognitive impairment in postmenopausal women: The Women's Health Initiative Memory Study: A randomized controlled trial. Journal of the American Medical Association, 289, 2651–2662. doi: 10.1001/jama.289.20.2651 [PubMed: 12771112]

Shumaker SA, Reboussin BA, Espeland MA, Rapp SR, McBee WL, Dailey M, … for the WHIMS Investigators. (1998). The Women's Health Initiative Memory Study (WHIMS): A trial of the effect of estrogen therapy in preventing and slowing the progession of dementia. Controlled Clinical Trials, 19, 604–621. doi: 10.1016/S0197-2456(98)00038-5 [PubMed: 9875839]

Small BJ, Dixon RA, Hultsch DF, & Hertzog C (1999). Longitudinal changes in quantitative and qualitative indicators of word and story recall in Young-Old and Old-Old adults. Journal of Gerontology, 54B, P107–P115. doi: 10.1093/geronb/54B.2.P107

Teng EL, & Chui HC (1987). The Modified Mini-Mental State (3MS) Examination. Journal of Clinical Psychiatry, 48, 314–318. [PubMed: 3611032]

Tierney MC, Oh P, Moineddin R, Greenblatt EM, Snow WG, Fisher RH, … MacLusky NJ (2009). A randomized double-blind trial of the effects of hormone therapy on delayed verbal recall in older women. Psychoneuroendocrinology, 34, 1065–1074. doi: 10.1016/j.psyneuen.2009.02.009 [PubMed: 19297102]

Tremont G, Miele A, Smith MM, & Westervelt HJ (2010). Comparison of verbal memory impairment rates in mild cognitive impairment. Journal of Clinical and Experimental Neuropsychology, 32, 630–636. doi: 10.1080/13803390903401328 [PubMed: 20603742]

Wechsler D (1981). WAIS-R manual: Wechsler adult intelligence scale-revised. Psychological Corporation.

Wilson BA, Cockburn J, & Baddeley AD (1991). The Rivermead Behavioural Memory Test (2nd ed.). Bury St Edmunds: Thames Valley Test Co.

Wilson BA, Cockburn J, Baddeley AD, & Hiorns R (1989). The development and validation of a test battery for detecting and monitoring everyday memory problems. Journal of Clinical and Experimental Neuropsychology, 11, 855–870. [PubMed: 2592527]

Wolf OT, Heinrich AB, Hanstein B, & Kirschbaum C (2005). Estradiol or estradiol/progesterone treatment in older women: no strong effects on cognition. Neurobiology of Aging, 26, 1029–1033. doi: 10.1016/j.neurobiolaging.2004.09.012 [PubMed: 15748783]

Women's Health Initiative Steering Committee. (2004). Effects of conjugated equine estrogen in postmenopausal women with hysterectomy: The women's health initiative randomized controlled trial. Journal of the American Medical Association, 291, 1701–1712. doi: 10.1001/jama.291.14.1701 [PubMed: 15082697]

Writing Group for the Women's Health Initiative Investigators. (2002). Risks and benefits of estrogen plus progestin in healthy postmenopausal women: Principal results from the women's health initiative randomized controlled trial. Journal of the American Medical Association, 288, 321–333. doi: 10.1001/jama.288.3.321 [PubMed: 12117397]

Yaffe K, Vittinghoff E, Ensurd KE, Johnson KC, Diem S, Hanes V, & Grady D (2006). Effects of Ultra–Low-Dose Transdermal Estradiol on Cognition and Health-Related Quality of Life. Archives of Neurology, 63, 945–950. doi: 10.1001/archneur.63.7.945 [PubMed: 16831962]

Zimprich D, & Kurtz T (2015). Subjective and Objective Memory Changes in Old Age across Five Years. Gerontology, 61, 223–231. doi: 10.1159/000369927 [PubMed: 25791780]

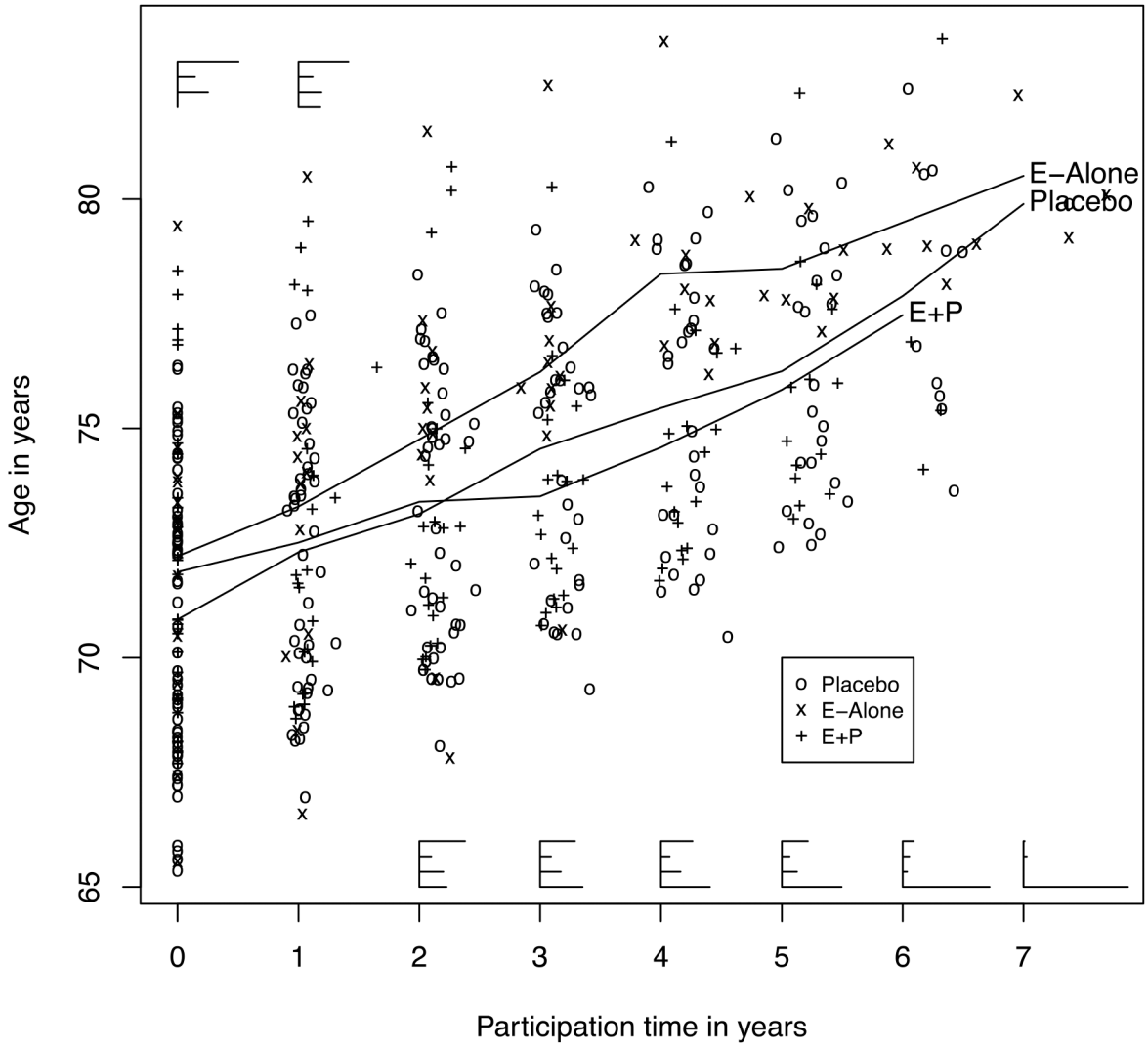

Figure 1.

Plot of the age and time since baseline for each subject included in the analysis for each session. Both axes are measured in days converted to fractional years. Note that Participation Time 0 corresponds to Session 1 and that Session 8 (Participation Time 7) was excluded from analysis due to inadequate sample size. The solid lines give the mean age for each group at 1-year intervals. The horizontal lines of the F-E-L plots represent the distribution of in-compliance subjects at each year, in order from top to bottom: Placebo, E-Alone, E+P, Noncompliant/Withdrawn.

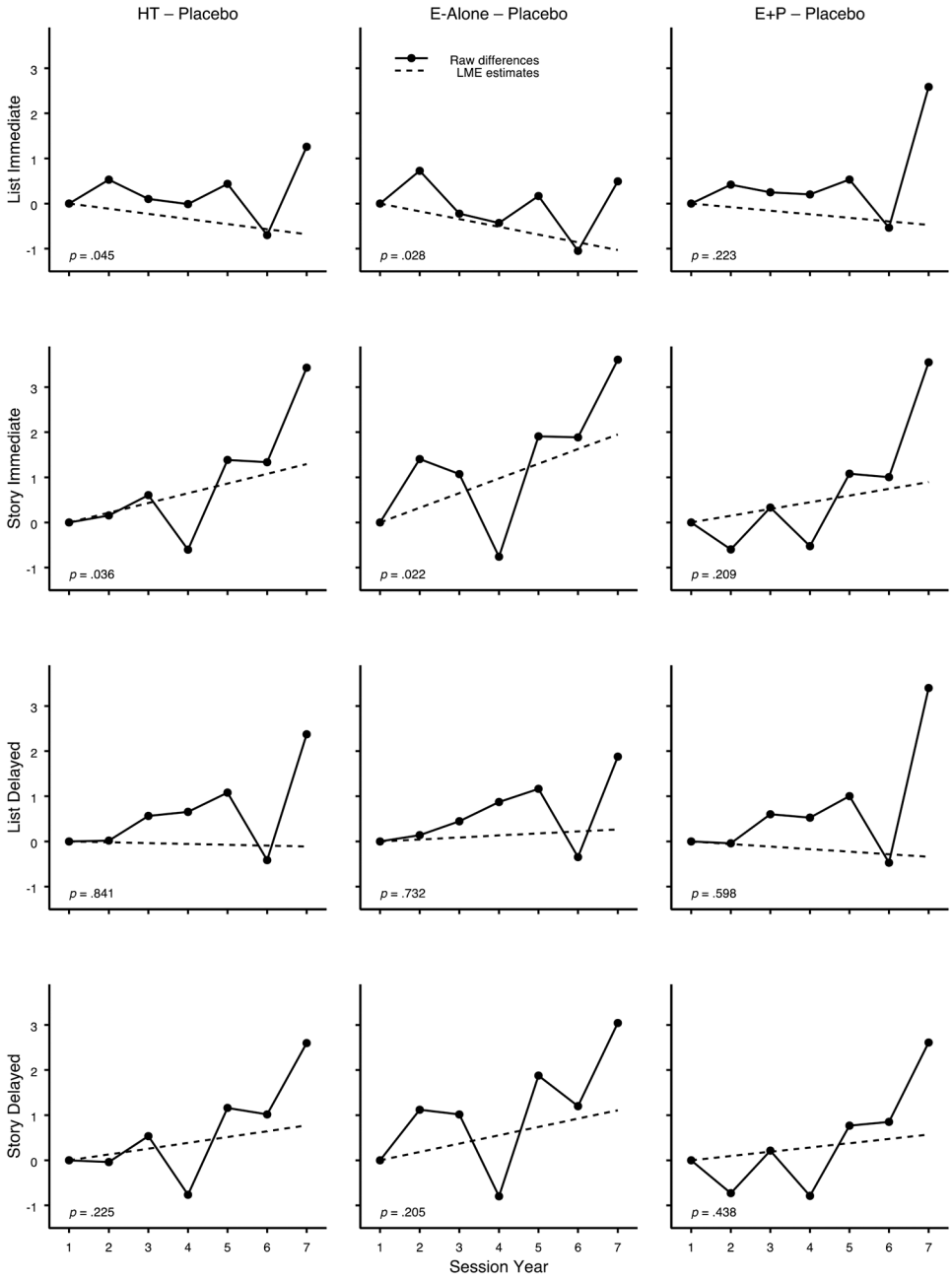

Figure 2.

Zero-adjusted treatment group scores minus placebo group scores over time for combined HT, E-Alone, and E+P treatments, for the immediate and delayed sub-tests of a word List recall task and a Story recall task. Solid lines are raw mean differences plotted at nominal session years. Dashed lines are slope estimates from Linear Mixed Effects (LME) analyses, based on participation time in fractional years, with per-session age in fractional years and quantity of study medication consumed as covariates. The LME-derived *p* values indicate the reliability of the difference between treatment and placebo slopes and are reproduced here for convenience.

Exp Aging Res. Author manuscript; available in PMC 2020 May 01.

Table 1

Review of list and story recall results in HT randomized trials

	E-Alone				E+P			
	Immediate		Delayed		Immediate		Delayed	
	List Recall	Story Recall	List Recall	Story Recall	List Recall	Story Recall	List Recall	Story Recall
HT > Pl		Ss4(ab)[l] Ss4(abcd)[m] Ss4(ab)[n]						
HT ≥ Pl							Sm5aa[a] Ml7aaa[g]	
HT = Pl	Ss7ab[b] Ms7aa[c] Ms5aa[d] Ll6aaa[f]	Ss5abcd[o] Ss5aa[p] Ss6aaa[q] Ll6aaa[f] Ms5aa[k]	Ms7aa[c] Ms5aa[d] Ss7ab[b] Ll6aaa[f] Ms5aa[k]	Ss4(ab)[n] Ss5abcd[o] Ss4(ab)[l] Ss6aaa[q] Ll6aaa[f] Ms5aa[k]	Sm5aa[a] Ms5ab[e] Ml7aaa[g]	Ss5aa[p] Ms5[e] Sl5ab[h] Ss6aaa[q]	Sl5ab[h]	Ms5[e] Sl5ab[h] Ss6aaa[q]
HT ≤ Pl	Ll7aab[i]		Ll7aab[i]		Sl5ab[h]		Ms5ab[e]	
HT < Pl					Ll7aab[j]		Ll7aab[j]	

Note: ‘HT’ hormone therapy, ‘Pl’ placebo, ‘>‘, ‘<‘ $p < .05$, ‘≥’, ‘≤’ $p < .1$, ‘=‘ $p \geq .1$

Entries are of the form NdD where N is the number of participants (Small < 50, Medium < 250, Large ≥ 250), d is the duration of the study (short < 182 days, medium < 365 days, long ≥ 365 days), and D is the decade of mean age of the participants. The remaining letters represent alternate forms used in each session; parentheses indicate that form order was counterbalanced across subjects.

[a]Alhola et al. (2010),

[b]Schiff et al. (2005),

[c]Almeida et al. (2006),

[d]Joffe et al. (2006),

[e]Maki et al. (2007),

[f]Yaffe et al. (2006),

[g]Tierney et al. (2009),

[h] Maki et al. (2009),
[i] Resnick et al. (2009),
[j] Resnick et al. (2006),
[k] Gorenstein et al. (2011),
[l] Phillips and Sherwin (1992),
[m] Sherwin (1988),
[n] Sherwin and Phillips (1990),
[o] LeBlanc et al. (2007),
[p] Linzmayer et al. (2001),
[q] Wolf et al. (2005)

Table 2

Hypothetical outcome patterns in terms of slope of longitudinal trajectories for HT relative to placebo, according to the individual components of Fuzzy-Trace Theory (FTT)

FTT Component	List Recall	Story Recall
Improved verbatim processing	>	≥
Impaired verbatim processing	<	≤
Improved gist processing	≥	>
Impaired gist processing	≤	<
Metacognitive gist bias	<	>
Metacognitive verbatim bias	>	<

Notes: ">" sparing relative to placebo, positive slope; "≥" possible or limited sparing relative to placebo, flat or positive slope; "<" worsening relative to placebo, negative slope; "≤" possible or limited worsening relative to placebo, flat or negative slope.

Table 3

Demographics

Group	*N*	Age	Education Years	3MS	VIQ
Placebo	56	70.8 (3.1)	13.9 (3.0)	95.1 (3.9)	99.5 (5.2)
E-Alone	16	72.2 (3.2)	13.9 (1.8)	93.2 (4.1)	99.3 (4.8)
E+P	28	71.9 (3.2)	14.3 (2.7)	95.3 (4.4)	98.8 (5.4)

Notes. E-Alone = Estrogen only; E+P = Estrogen + Progestin; 3MS = Modified Mini-Mental State Examination (3MS; Teng & Chui, 1987), as scored by WHIMS, percent correct (out of 100), minus number of missing responses; VIQ = verbal intelligence as estimated via the American version of the Nelson Adult Reading Test (Grober et al., 1991).

Table 4

Linear mixed effect analysis of the effects of HT relative to placebo.

	Immediate List		Immediate Story		Delayed List		Delayed Story	
Source	$\widehat{m}$ **± 95% CI**	***p***	$\widehat{m}$ **± 95% CI**	***p***	$\widehat{m}$ **± 95% CI**	***p***	$\widehat{m}$ **± 95% CI**	***p***
Age	−0.100 ± 0.089	.029	−0.171 ± 0.140	.017	−0.133 ± 0.143	.069	−0.182 ± 0.140	.011
Qty	−1.577 ± 1.247	.014	0.451 ± 2.537	.728	−1.008 ± 2.027	.331	0.600 ± 2.663	.659
T	1.695 ± 1.232	.007	−0.575 ± 2.504	.653	1.060 ± 2.003	.300	−0.569 ± 2.630	.672
T * HT	−0.114 ± 0.111	.045	0.215 ± 0.201	.036	−0.018 ± 0.182	.841	0.129 ± 0.208	.225

T (participation time) is an estimate of the slope of change in scores over participation time for the baseline (placebo) group; T * HT is the difference in the slope for the HT group from the placebo group's slope; $\widehat{m}$ is the point estimate of the slope. Covariates: Age is continuous per-session age; Qty is the total number of days in which the study medication was taken up to a given session. Estimates are in unscaled test score units.

Table 5

Linear mixed effect analysis of the effects of HT with and without progestin relative to placebo.

	Immediate List		Immediate Story		Delayed List		Delayed Story	
Source	$\widehat{m}$ **± 95% CI**	***p***	$\widehat{m}$ **± 95% CI**	***p***	$\widehat{m}$ **± 95% CI**	***p***	$\widehat{m}$ **± 95% CI**	***p***
Age	−0.097 ± 0.090	.034	−0.180 ± 0.141	.013	−0.137 ± 0.143	.063	−0.187 ± 0.141	.010
Qty	−1.507 ± 1.252	.019	0.294 ± 2.550	.822	−1.073 ± 2.037	.303	0.519 ± 2.679	.705
T	1.624 ± 1.237	.011	−0.412 ± 2.519	.749	1.127 ± 2.014	.273	−0.484 ± 2.649	.720
T * E	−0.172 ± 0.152	.028	0.325 ± 0.276	.022	0.044 ± 0.253	.732	0.185 ± 0.285	.205
T * EP	−0.078 ± 0.127	.223	0.149 ± 0.232	.209	−0.056 ± 0.210	.598	0.095 ± 0.240	.438

T (participation time) is an estimate of the slope of change in scores over participation time for the baseline (placebo) group; T * E and T * EP are the differences in the slope for the E-Alone and E+P group, respectively, from the placebo group's slope; $\widehat{m}$ is the point estimate of the slope. Covariates: Age is continuous per-session age; Qty is the total number of days in which the study medication was taken up to a given session. Estimates are in unscaled test score units.