Supporting Information

Henderson et al. 10.1073/pnas.1312353110

Protocol Plan

Design. The Early versus Late Intervention Trial with Estradiol (ELITE) was designed as a single-site, randomized, double-blind trial of oral 17β-estradiol (E2; 1 mg daily) or matched placebo. Women with a uterus also receive cyclic micronized progesterone (45 mg) as a 4% vaginal gel or matched placebo (one application daily for 10 d per 30-d cycle). Study drugs were supplied without charge by Teva Pharmaceuticals, Watson Pharmaceuticals, Inc., and Abbott Laboratories. Between July 13, 2005 and September 30, 2008, 643 women were recruited into early and late postmenopause groups. Women in the early postmenopause group were within 6 y of their final menstrual period (natural menopause) or bilateral oophorectomy (surgical menopause). Late postmenopausal women were at least 10 y beyond their final menstrual period or bilateral oophorectomy.

In ELITE, the primary research question concerns differential effects of E2 treatment for women in the early postmenopause group compared with women in the late postmenopause group. For the parent ELITE trial, the primary end point is progression of subclinical atherosclerosis measured as the common carotid artery intima media thickness. The hypothesis is that E2 will reduce atherosclerosis progression if initiated in early postmenopause (when the vascular endothelium is most likely to be relatively healthy) compared with late postmenopause (when the endothelium may have lost much of its responsiveness to estrogen). The cognitive study (ELITE-Cog) is a funded supplement to the parent trial. For ELITE-Cog, the primary end point is the change in a composite measure of verbal episodic memory. This cognitive domain is reported to be sensitive to estrogen effects (1), and verbal memory impairment is associated with increased Alzheimer's disease risk (2–4). We hypothesize that change in the memory composite will differ between postmenopause groups, with better memory performance among early postmenopausal women randomized to hormone therapy compared with placebo but not among women in the late postmenopause group. Primary analyses will be conducted separately within each menopause stratum. Other end points are given below.

Recruitment in the parent trial occurred over a 38-mo period; 2,814 women were screened by telephone for eligibility, of whom 895 women were screened in person, and 643 women were randomized. Common reasons for exclusion were current use of hormone therapy (197 women), not being in a designated postmenopause stratum (not <6 or ≥10 y after menopause; 190 women), indeterminate time from menopause (132 women), and not postmenopausal (90 women). Recruitment was initially based on a 5-y trial with a planned treatment period of 2–5 y depending on when the participant was randomized. The trial was extended with supplemental funding for an additional 2.5 y of treatment. The extension allows a third cognitive assessment as participants complete the trial as well as the determination of coronary artery calcification and assessment of subclinical coronary artery atherosclerosis by coronary computed tomographic angiography.

Standard protocol approval, participant consent, and trial registration. ELITE-Cog procedures were approved by Institutional Review Boards of the University of Southern California and Stanford University. Participants provided written informed consent before study-related procedures were performed. The protocol is registered at ClinicalTrials.gov (NCT00114517).

Participants. Participants were healthy postmenopausal women who had consented to procedures for the parent ELITE trial. Postmenopausal status was based on a serum level of total E2 < 25 pg/mL and absence of vaginal bleeding for at least 6 mo (natural menopause) or bilateral oophorectomy (surgical menopause). ELITE exclusion criteria included use of menopausal hormone therapy within the preceding 1 mo, history or evidence of cardiovascular disease, diabetes mellitus, or fasting serum glucose > 140 mg/dL, uncontrolled hypertension (diastolic blood pressure > 110 mmHg or systolic blood pressure > 160 mmHg), fasting plasma triglyceride level > 500 mg/dL, untreated thyroid disease, renal insufficiency (serum creatinine > 2.0 mg/dL), life-threatening disease, liver disease, history of deep vein thrombosis or pulmonary embolism, history of breast cancer, and inability to participate in outcome assessments.

Cognitive assessment. Cognitive skills were assessed at baseline with a comprehensive neuropsychological battery that emphasized standardized tests sensitive to age-associated change in middle-aged and older adults (Table S2). Other considerations in test selection were our prior experiences with these tests in similar populations (5, 6). We also intended to include tasks in common with the battery that was then being considered for the Kronos Early Estrogen Prevention Study (ClinicalTrials.gov identifier NCT00623311), and for this reason, the Benton Visual Retention Test (7) and the Stroop Color-Interference Test (8) were added to the battery after baseline testing.

Randomization, concealment, and follow-up. Within each postmenopause stratum, assignment to treatment group in a 1:1 ratio used stratified blocked randomization. Other stratification factors were carotid artery intima media thickness (<0.75 mm, ≥0.75 mm) defined by high-resolution B-mode ultrasonography (9) and presence or absence of a uterus at randomization. The computer-generated randomization sequence (SAS statistical software) was prepared before trial initiation by a statistician, and all other investigators, participants, staff, and data monitors were masked to group assignment. Clinic visits occurred every 1 mo for the first 6 mo and every 2 mo thereafter. Cognitive skills were assessed at baseline, at about 33 mo (SD = 2.6, range = 29–50 mo), and as women completed the trial at about 57 mo (SD = 5.8, range = 36–77 mo).

Study power and statistical analysis. ELITE uses a 2 × 2 factorial design, with randomization on one factor (E2 vs. placebo) and menopause group as the second factor (early vs. late postmenopause). Sample size requirements were based on the hypothesis that the exogenous E2 effect on change in carotid artery intima media thickness varies by time since menopause. The projected mean treatment group difference in change rate was 0.0144 mm/y in the early postmenopause group and 0.0021 mm/y in the late group, with the SD of carotid artery intima media change being 0.02 mm/y (9). Sample size estimates to test this interaction at 80% power and a two-sided α-level of 0.05 (allowing for 25% dropout) resulted in a sample size for ELITE of 126 in each of the four strata (a total of 504 subjects). The sample size was subsequently enhanced to increase power to detect treatment interactions and accommodate any higher than anticipated dropout rate among women in the late postmenopause group.

For primary analyses in ELITE-Cog based on verbal episodic memory (described below), we will have 82% power (two-sided α of 0.05) to detect differences in the E2 treatment effect between postmenopausal strata if the early stratum has a treatment effect size of 0.5, which was suggested by some clinical trials in relatively younger women (1, 10–12), and the late stratum has a treatment effect size of 0, which was suggested by clinical trials in relatively older women (13–20). We will have 80% power to detect treat-

ment effect size of 0.34 in the early postmenopause group. For an overall comparison of treatment effects within combined postmenopause strata, we will be able to detect treatment differences of one-fourth SD ($d = 0.22$).

Primary and secondary end points. To guard against a type I error in modified intention-to-treat analyses, we will use two composite measures as primary and secondary end points to test our hypothesis. The primary ELITE-Cog end point is based on a composite score of four measures of verbal episodic memory (word list, immediate and delayed recall; paragraph story, immediate and delayed recall). The secondary end point uses a global composite score that includes all cognitive tests. The composites are calculated as the average of component standardized scores weighted by the inverse intertest correlation matrix as previously described (6). Absolute values of change from baseline on each cognitive end point will be calculated at both postrandomization cognitive follow-ups (2.5 y and end of study). General estimating equation models will be used to analyze the repeated measures of change in the cognitive end points. The primary independent variables will be randomized treatment group and postmenopause stratum; covariates will include randomization stratification variables (carotid artery intima media thickness stratum and hysterectomy status) as well as the baseline value of the cognitive end point. Differences in cognitive change will be tested by treatment group and postmenopause stratum. The main hypothesis of interest will be tested with the interaction (treatment group by postmenopause stratum). Additional analyses will test for interactions with test time (2.5 y vs. end of study) to evaluate if treatment effects (treatment by time interaction) and the differential treatment effect by postmenopause stratum (treatment by postmenopause stratum by test time interaction) vary by follow-up time.

Other end points. To explore whether the E2 treatment effect might be more apparent within cognitive domains other than verbal memory, we will examine changes in an executive function composite measure (in the text) because of reported associations between atherosclerosis and executive functions (21) as well as changes in individual neuropsychological test scores, adjusting for multiple comparisons.

We will compare the treatment groups on the proportion of women with large memory decline, which may represent increased risk for Alzheimer's disease. The cut point for large change will be defined within each postmenopause group by memory change falling below 1.5 SD of the mean change. We will compute the mean and SD within a postmenopause stratum by combining treatment groups. The memory score will be constructed as a weighted linear combination of episodic memory tasks. For this analysis, we will include both verbal and nonverbal memory tasks; decline on each type of task has been predictive of Alzheimer's disease (3, 4, 22), and beneficial estrogen effects, although more often reported for verbal memory, are also claimed for nonverbal memory (23).

To be conservative in terms of multiple hypothesis testing, we specified only two subgroup analyses. We will evaluate the treatment group differences on cognitive change scores separately in participants who experienced a natural vs. surgical menopause. A second subgroup analysis will similarly be performed according to the presence of vasomotor symptoms at baseline. In clinical trials of estrogen therapy that included women with menopausal symptoms, cognitive outcomes tended to favor women assigned to active estrogen therapy (24).

Other analyses, which were specified but we view as exploratory, will consider cognitive outcomes (memory, executive functions, and global cognition) in relation to changes in carotid artery intima media thickness, coronary artery calcium, coronary computed tomographic angiography end points, serum hormone concentrations, reproductive factors, and other health factors that may be affected by E2 and in part, mediate or moderate hypothesized E2 actions on cognition.

SI Materials and Methods

Protocol Approval, Informed Consent, and Trial Registration. ELITE-Cog is a funded supplement to the parent ELITE trial focused on cognitive outcomes (*Protocol Plan*). Procedures were approved by Institutional Review Boards of the University of Southern California and Stanford University. Participants provided written informed consent before study procedures were performed. The protocol is registered at ClinicalTrials.gov (NCT00114517).

Participants. Eligible ELITE-Cog participants were healthy postmenopausal women who had consented to procedures for the parent ELITE trial (Table 1). Postmenopausal status was based on a serum level of total E2 < 25 pg/mL and absence of vaginal bleeding for at least 6 mo (natural menopause) or bilateral oophorectomy (surgical menopause). Demographic information, reproductive history, and other health information were obtained from participants through a structured interview during the screening assessment. Women recorded in a diary the number of daily hot flashes and rated their intensity as mild, moderate, or severe (25) [mean duration of recording before randomization of 40 d (SD = 21)]. Physical activity (26) and alcohol intake during the preceding 1 wk were assessed with standardized interviews. We measured blood pressure, height, and weight and calculated the body mass index (kilograms per meter^2). After an overnight fast, blood was obtained for sex steroid and sex hormone binding globulin (SHBG) assays, DNA extraction for apolipoprotein E genotype, and other blood measures. Three isoforms (ε2, ε3, and ε4) of the apolipoprotein E gene were determined according to two nonsynonymous SNPs (rs429358 and rs7412) (27) encoding arginine for cysteine amino acid variants at codon positions 112 and 158, respectively (TaqMan Assay-on-Demand Genotyping Service; Applied Biosystems) (28).

Cognitive Assessment. Cognitive skills were assessed at baseline with a comprehensive neuropsychological battery (5, 6) that emphasized standardized tests sensitive to age-associated change in middle-aged and older adults (Table S2). Neuropsychological tests and corresponding cognitive skills (29) are listed here: Symbol Digit Modalities Test (30), complex scanning and visual tracking, attention, and psychomotor speed; Trail Making Test, Part B (31), visuomotor tracking, planning, cognitive flexibility, and psychomotor speed; Shipley Institute of Living Scale, Abstraction scale (32), concept formation; Letter–Number Sequencing (33), working memory, attention, and concentration; Block Design (34), visuospatial perception, nonverbal concept formation, planning, and visuoconstructive ability; Judgment of Line Orientation (35), visuospatial perception; animal naming (36), verbal fluency and semantic memory; Boston Naming Test (37), naming and semantic memory; a short version of the California Verbal Learning Test (38), verbal episodic memory, word list learning, and concept formation; East Boston Memory Test (39), verbal episodic memory and logical memory; and Faces I and II (33), visual episodic memory, memory for faces, and visuoperceptual processing. The verbal intelligence quotient was estimated with the Wechsler Test of Adult Reading (40).

To guard against the type I error in modified intention-to-treat analyses, end points in the ELITE-Cog trial focus on cognitive composites for verbal episodic memory, executive functions (executive impairments, like memory impairments, are linked to Alzheimer's disease risk) (4, 41), and overall cognitive performance (global cognition). Each composite is calculated as the average of component standardized scores weighted by the inverse intertest correlation matrix as described in ref. 6. The verbal memory composite was defined a priori by California

Verbal Learning Test (38) (word list recall) and East Boston Memory Test (39) (paragraph recall) immediate and delayed recall scores and the global composite by scores from all neuropsychological measures. Tests used for the executive functions composite were determined by a principal components analysis of baseline scores. This composite used scores from the Symbol Digit Modalities Test (30), the Trail Making Test, Part B (31), the Shipley Institute of Living Abstraction scale (32), Letter–Number Sequencing (33), and category fluency (36).

Serum concentrations of free E2 have been positively—and the log ratio of free T to E2 concentrations has been negatively—associated with naming (semantic memory) (42). In ELITE-Cog, we did not propose a priori to examine semantic memory independent of its contribution to global cognition, but we considered these associations in supplemental analyses (*SI Materials and Methods*, *Other Analyses*). We similarly examined the reported negative association between free T and verbal fluency (43).

Mood and Possible Depression. Mood was assessed with the 20-item Center for Epidemiological Studies depression scale and analyzed continuously and dichotomously using a standard cut point (44) to indicate possible depression.

Sex Steroid and SHBG Concentrations. E2, estrone (E1), testosterone (T), and progesterone (P4) levels were measured in serum as described previously (42, 45). Serum was stored immediately at −20 °C and transferred for long-term storage at 80 °C. Steroids were extracted with hexane and ethyl acetate in a 3:2 ratio. The extract was purified before RIA by Celite column partition chromatography using ethylene glycol as the stationary phase. E2 was eluted with 40% ethyl acetate in isooctane. The assay sensitivity is 2 pg/mL, the intraassay coefficient of variation (CV) is 8.9% at 14 pg/mL, and the interassay CV is 8.8% at 11 pg/mL; 15% ethyl acetate in isooctane was used for E1 elution. The assay sensitivity is 4 pg/mL, the intraassay CV is 7.9% at 26 pg/mL, and the interassay CV is 13% at 22 pg/mL. P4 was eluted with isooctane. The assay sensitivity is 10 pg/mL, the intraassay CV is 7.5% at 190 pg/mL, and the interassay CVs are 14% at 135 pg/mL and 8.0% at 259 pg/mL. T was eluted with 40% toluene in isooctane. The assay sensitivity is 1.5 ng/dL, the intraassay CV is 9.0% at 33 ng/dL, and the interassay CVs are 14% at 9.4 ng/dL and 10% at 27.6 ng/dL.

SHBG for the determination of free (nonprotein bound) E2 and T was quantified by a solid-phase, two-site chemiluminescent immunoassay on the Immulite Analyzer (Siemens Healthcare Diagnostics). The assay sensitivity is 1 nmol/L, the intraassay CV is 5.2% at 63 nmol/L, and the interassay CVs are 6.2% at 21 nmol/L and 4.3% at 72 nmol/L. Free E2 and T were calculated with a validated algorithm (46) based on derived equations as described previously (42).

Statistical Analyses. Baseline characteristics were compared between the two menopause strata with independent sample t tests for continuous variables and χ^2 tests for categorical variables. The associations between individual serum hormone levels and cognitive measures and mood—and the relation between years since menopause and cognitive function—were modeled separately using multivariable linear regression analysis adjusting for age and other covariates (Table 1) that affected β-estimates for cognitive or mood end points on hormones by at least 10%. The relation between hormone levels and possible depression (depression scale score ≥16) was similarly examined by logistic regression adjusted for age and other potential confounders. Associations between individual hormones or SHBG and cognitive end points were modeled first without and then adjusting for other hormone measures. Because of collinearity (Table S1) and similar mechanisms of action, free E2 and E1 levels were not adjusted for each other.

With 271 women in the early group and 372 women in the late group for a two-sided α of 0.05, we had 80% power to detect a standardized difference in β-coefficients between postmenopause strata of at least 0.22. In the absence of modification by postmenopause group, for a two-sided α of 0.05, we had 80% power to detect significant associations between hormone measures and cognitive composites for effect sizes of at least 0.11.

Sensitivity Analyses. Analyses were repeated after excluding women who reported premature (age < 40 y; $n = 24$) or implausibly late (47) (age > 59 y; $n = 4$) menopause.

Other Analyses. We undertook supplemental analyses guided by prior findings in other cohorts. In the Melbourne Women's Midlife Health Project, analyses of postmenopausal women (mean age = 60 y) not using hormone therapy showed a significant positive relation between free E2 and naming performance (42). Findings were similar in ELITE-Cog participants. For free E2 and naming, there was a near-significant interaction between postmenopause strata. In the early postmenopause group (mean age = 55 y), the free E2 concentration was associated with better performance on the 30-item version of the Boston Naming Test (37) (Table S3). Women in the fourth quartile of free E2 levels named, on average, 1.2 more words than women in the first quartile, a difference corresponding to about 0.4 SDs. In the Melbourne cohort, serum E1 was unrelated to naming, and in ELITE-Cog, we also found no association between E1 and naming performance (β = 0.01, SE = 0.25, $P = 0.96$, interaction by postmenopause group $P = 0.26$). We were unable to confirm a negative association between the log ratio of free T to E2 concentrations and naming (42) (β = −0.79, SE = 0.18, $P = 0.66$, interaction $P = 0.26$).

In a Swedish cohort of women ages 35–90 y, there was a significant negative association between free T and verbal fluency (43). Using a somewhat different fluency task, we failed to confirm this association for postmenopausal women (β = −0.62, SE = 0.65, $P = 0.34$, interaction $P = 0.30$).

ELITE Research Group Members. ELITE Research Group Members are given below, with primary trial investigators designated by an asterisk. Study Chair: Howard N. Hodis.* Clinical Center Staff: Liny Zurbrugg (clinic coordinator), Esther Bhimani, Martha Charlson, Irma Flores, Martha Huerta, Thelma LaBree, Sonia Lavender, Violetta McElreath, Janie Teran, and Philip Zurbrugg. Ultrasound Image Acquisition and Processing Laboratory: Robert H. Selzer* (director), Yanjie Li (technical director), Mei Feng, Lora Whitfield-Maxwell, and Ming Yan. Data Coordinating Center: Wendy J. Mack* (director), Stanley P. Azen,* Farzana Choudhury, Carlos Carballo, Laurie Dustin, Adrian Herbert, Naoko Kono, George Martinez, and Olga Morales. Atherosclerosis Research Unit Core Lipid/Lipoprotein Laboratory: Juliana Hwang-Levine* (director), Gail Izumi, Arletta Ramirez, and Luci Rodriguez. Gynecology and Mammography: Donna Shoupe,* Juan C. Felix, Pulin Sheth, and Mary Yamashita. University of Southern California Endocrinology Laboratory: Carole Spencer. Cognition and Mood: Victor W. Henderson,* Carol A. McCleary, and Jan A. St. John. Kaiser Permanente Medical Center Recruitment Site: Malcolm G. Munro. Cardiac Computed Tomography Core Center: Matthew J. Budoff (director), Lily Honoris, Chris Dailing, and Sivi Carson. Apolipoprotein E Genotyping: Hooman Allayee. Data Safety Monitoring Board: Leon Speroff (chair), Robert H. Knopp (deceased), Richard H. Karas, Joan Hilton, and Judy Hannah (ex officio; National Institute on Aging).

1. Sherwin BB, Tulandi T (1996) "Add-back" estrogen reverses cognitive deficits induced by a gonadotropin-releasing hormone agonist in women with leiomyomata uteri. *J Clin Endocrinol Metab* 81(7):2545–2549.
2. Bateman RJ, et al. (2012) Clinical and biomarker changes in dominantly inherited Alzheimer's disease. *N Engl J Med* 367(9):795–804.
3. Elias MF, et al. (2000) The preclinical phase of alzheimer disease: A 22-year prospective study of the Framingham Cohort. *Arch Neurol* 57(6):808–813.
4. Albert MS, Moss MB, Tanzi R, Jones K (2001) Preclinical prediction of AD using neuropsychological tests. *J Int Neuropsychol Soc* 7(5):631–639.
5. Gatto NM, et al. (2009) Subclinical atherosclerosis is weakly associated with lower cognitive function in healthy hyperhomocysteinemic adults without clinical cardiovascular disease. *Int J Geriatr Psychiatry* 24(4):390–399.
6. Henderson VW, et al. (2012) Long-term soy isoflavone supplementation and cognition in women: A randomized, controlled trial. *Neurology* 78(23):1841–1848.
7. Sivan AB (1992) *Benton Visual Retention Test* (Psychological Corporation, San Antonio), 5th Ed.
8. Golden CJ, Freshwater SM (2002) *Stroop Color and Word Test. Revised 2002 Adult Manual for Clinical and Experimental Uses* (Western Psychological Services, Los Angeles).
9. Hodis HN, et al. (2001) Estrogen in the prevention of atherosclerosis. A randomized, double-blind, placebo-controlled trial. *Ann Intern Med* 135(11):939–953.
10. Phillips SM, Sherwin BB (1992) Effects of estrogen on memory function in surgically menopausal women. *Psychoneuroendocrinology* 17(5):485–495.
11. Woo J, et al. (2003) Comparison of *Pueraria lobata* with hormone replacement therapy in treating the adverse health consequences of menopause. *Menopause* 10(4):352–361.
12. Shaywitz SE, et al. (2003) Better oral reading and short-term memory in midlife, postmenopausal women taking estrogen. *Menopause* 10(5):420–426.
13. Binder EF, Schechtman KB, Birge SJ, Williams DB, Kohrt WM (2001) Effects of hormone replacement therapy on cognitive performance in elderly women. *Maturitas* 38(2):137–146.
14. Grady D, et al. (2002) Effect of postmenopausal hormone therapy on cognitive function: The Heart and Estrogen/progestin Replacement Study. *Am J Med* 113(7):543–548.
15. Viscoli CM, et al. (2005) Estrogen therapy and risk of cognitive decline: Results from the Women's Estrogen for Stroke Trial (WEST). *Am J Obstet Gynecol* 192(2):387–393.
16. Resnick SM, et al. (2006) Effects of combination estrogen plus progestin hormone treatment on cognition and affect. *J Clin Endocrinol Metab* 91(5):1802–1810.
17. Yaffe K, et al. (2006) Effects of ultra-low-dose transdermal estradiol on cognition and health-related quality of life. *Arch Neurol* 63(7):945–950.
18. Almeida OP, et al. (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. *Neurobiol Aging* 27(1):141–149.
19. Pefanco MA, et al. (2007) The effect of 3-year treatment with 0.25 mg/day of micronized 17beta-estradiol on cognitive function in older postmenopausal women. *J Am Geriatr Soc* 55(3):426–431.
20. Resnick SM, et al. (2009) Effects of conjugated equine estrogens on cognition and affect in postmenopausal women with prior hysterectomy. *J Clin Endocrinol Metab* 94(11):4152–4161.
21. Moorhouse P, Rockwood K (2008) Vascular cognitive impairment: Current concepts and clinical developments. *Lancet Neurol* 7(3):246–255.
22. Kawas CH, et al. (2003) Visual memory predicts Alzheimer's disease more than a decade before diagnosis. *Neurology* 60(7):1089–1093.
23. Resnick SM, Metter EJ, Zonderman AB (1997) Estrogen replacement therapy and longitudinal decline in visual memory. A possible protective effect? *Neurology* 49(6): 1491–1497.
24. LeBlanc ES, Janowsky J, Chan BKS, Nelson HD (2001) Hormone replacement therapy and cognition: Systematic review and meta-analysis. *JAMA* 285(11):1489–1499.
25. Food and Drug Administration (2003) *Guidance for Industry. Estrogen and Estrogen/Progestin Drug Products to Treat Vasomotor Symptoms and Vulva and Vaginal Atrophy Symptoms—Recommendations for Clinical Evaluation* (Food and Drug Administration, Center for Drug Evaluation and Research, Rockville, MD).
26. Blair SN, et al. (1985) Assessment of habitual physical activity by a seven-day recall in a community survey and controlled experiments. *Am J Epidemiol* 122(5):794–804.
27. Nyholt DR, Yu CE, Visscher PM (2009) On Jim Watson's APOE status: Genetic information is hard to hide. *Eur J Hum Genet* 17(2):147–149.
28. Livak KJ (1999) Allelic discrimination using fluorogenic probes and the 5′ nuclease assay. *Genet Anal* 14(5–6):143–149.
29. Lezak MD, Howieson DB, Loring DW (2004) *Neuropsychological Assessment* (Oxford Univ Press, New York), 4th Ed.
30. Smith A (1991) *Symbol Digit Modalities Test Manual* (Western Psychological Services, Los Angeles).
31. Reitan RM, Wolfson D (1993) *The Halstead-Reitan Neuropsychological Test Battery: Theory and Clinical Interpretations* (Neuropsychology Press, Tucson, AZ), 2nd Ed.
32. Zachary RA (1986) *Shipley Institute of Living Scale. Revised Manual* (Western Psychological Services, Los Angeles).
33. Wechsler D (1997) *WMS-III. Wechsler Memory Scale* (Psychological Corporation, San Antonio), 3rd Ed.
34. Wechsler D (1981) *Weschler Adult Intelligence Scale—Revised* (Psychological Corporation, New York).
35. Benton AL, Varney NR, Hamsher KD (1978) Visuospatial judgment. A clinical test. *Arch Neurol* 35(6):364–367.
36. Troyer AK (2000) Normative data for clustering and switching on verbal fluency tasks. *J Clin Exp Neuropsychol* 22(3):370–378.
37. Mack WJ, Freed DM, Williams BW, Henderson VW (1992) Boston Naming Test: Shortened versions for use in Alzheimer's disease. *J Gerontol* 47(3):154–158.
38. Delis DC, Kramer J, Kaplan E, Ober BA (2000) *CVLT-II. California Verbal Learning Test, Adult Version* (Harcourt-Brace, San Antonio), 2nd Ed.
39. Scherr PA, et al. (1988) Correlates of cognitive function in an elderly community population. *Am J Epidemiol* 128(5):1084–1101.
40. Anonymous (2001) *WTAR. Wechsler Test of Adult Reading Manual* (Harcourt Assessment, San Antonio).
41. Elias MF, et al. (2000) The preclinical phase of Alzheimer disease: A 22-year prospective study of the Framingham Cohort. *Arch Neurol* 57(6):808–813.
42. Ryan J, et al. (2012) Hormone levels and cognitive function in postmenopausal midlife women. *Neurobiol Aging* 33(7):1138–1147.
43. Thilers PP, Macdonald SW, Herlitz A (2006) The association between endogenous free testosterone and cognitive performance: A population-based study in 35 to 90 year-old men and women. *Psychoneuroendocrinology* 31(5):565–576.
44. Andresen EM, Malmgren JA, Carter WB, Patrick DL (1994) Screening for depression in well older adults: Evaluation of a short form of the CES-D (Center for Epidemiologic Studies Depression Scale). *Am J Prev Med* 10(2):77–84.
45. Scott JZ, Stanczyk FZ, Goebelsmann U, Mishell DRJ, Jr. (1978) A double-antibody radioimmunoassay for serum progesterone using progesterone-3-(O-carboxymethyl) oximino-[125I]-iodo-histamine as radioligand. *Steroids* 31(3):393–405.
46. Rinaldi S, et al. (2002) Validity of free testosterone and free estradiol determinations in serum samples from postmenopausal women by theoretical calculations. *Cancer Epidemiol Biomarkers Prev* 11(10 Pt 1):1065–1071.
47. DHEW (1966) *Age at Menopause, United States 1960–1962*, DHEW Publication Number 73-1268 (DHEW, Washington, DC).

Table S1. Correlations in postmenopausal women among serum concentrations of sex steroid hormones and SHBG

	Free E2	E1	P4	Total T	Free T	SHBG
Total E2	0.96	0.71	0.22	0.37	0.43	−0.16
Free E2	1.0	0.67	0.21	0.29	0.47	−0.42
E1		1.0	0.23	0.45	0.48	−0.080*
P4			1.0	0.25	0.27	−0.042†
Total T				1.0	0.89	0.19
Free T					1.0	−0.27

Pearson correlations are based on log concentrations. Sample sizes ranged from 635 to 640. All P values are <0.0001 (two-tailed), except where indicated.

*$P = 0.04$.

†$P = 0.29$.

Table S2. Baseline neuropsychological test scores

Neuropsychological test; potential range of test scores	Early postmenopause (n = 271)	Late postmenopause (n = 371)*	P
Symbol Digit Modalities; 0–110	55.4 (8.3)	51.4 (8.1)	<0.0001
Trail Making Test, Part B; 300–0[†]	72.6 (34.9)	82.9 (40.2)	0.0006
Shipley Abstraction Scale; 0–20	15.2 (3.6)	14.0 (3.8)	<0.0001
Letter–Number Sequencing; 0–21	10.3 (2.3)	9.50 (2.4)	<0.0001
Block Design; 0–51	31.0 (9.5)	28.6 (9.4)	0.002
Judgment of Line Orientation, Form H; 0–30	24.1 (4.7)	23.9 (4.3)	0.63
Category fluency (animal naming); ≥0	31.3 (9.6)	29.0 (8.3)	0.001
Boston Naming Test; 0–30	26.9 (2.9)	26.5 (2.8)	0.07
California Verbal Learning Test			
Immediate recall, three trials; 0–48	29.4 (5.9)	27.7 (6.3)	0.0005
Delayed recall; 0–16	10.4 (3.0)	9.4 (3.2)	<0.0001
East Boston Memory Test			
Immediate recall; 0–6	4.8 (1.1)	4.7 (1.0)	0.44
Delayed recall; 0–6	4.7 (1.1)	4.6 (1.1)	0.44
Visual memory (face recognition)			
Faces I, immediate recall; 0–48	36.8 (4.6)	36.1 (4.5)	0.04
Faces II, delayed recall; 0–48	37.9 (4.2)	37.2 (4.6)	0.04
Composite scores[‡]			
Verbal episodic memory	0.19 (1.31)	−0.14 (1.36)	0.003
Executive functions	0.35 (1.34)	−0.26 (1.32)	<0.0001
Global cognition	0.34 (1.74)	−0.04 (1.60)	0.005

Values given as mean (SD).
*One woman randomized to the late postmenopause group did not undergo baseline cognitive testing. Sample sizes ranged from 268 to 271 for the early group and from 366 to 371 for the late group.
[†]Test scores based on time, with lower scores reflecting better performance.
[‡]Composite scores were derived from weighted standardized scores as described (*SI Materials and Methods*).

Table S3. Associations between serum concentrations of E2 and semantic memory (naming)

	Naming			
	β	SE	P	Interaction P
Combined groups	0.22	0.18	0.20	0.09*
Early group	0.50	0.24	0.04[†,‡]	
Late group	−0.20	0.26	0.43	

There were 268 women in the early postmenopause group and 369 women in the late postmenopause group. Values for free E2 were log-transformed, and analyses were adjusted for age and race/ethnicity. Additional adjustment for P4, free T, and SHBG had no substantial effect.
*0.05 ≤ interaction P < 0.1.
[†]P < 0.05.
[‡]Based on quartiles of free E2, the adjusted mean Boston Naming Test (30 items) score for early postmenopause women was 25.1 (SE = 0.39) for women in the first (lowest) quartile, 26.0 (0.38) for women in the second quartile, 26.3 (0.35) for women in the third quartile, and 26.3 (0.33) for women in the highest quartile. Differences between the first quartile and the third (P = 0.02) and fourth (P = 0.01) quartiles were significant.

Table S4. Associations between serum P4 concentrations and cognitive composite scores: sensitivity analyses

	Verbal episodic memory				Executive functions				Global cognition			
	β	SE	P	Interaction P	β	SE	P	Interaction P	β	SE	P	Interaction P
Combined groups	0.09	0.10	0.38	0.04*	−0.10	0.10	0.32	0.17	0.20	0.12	0.10	0.04*
Early group	0.34	0.14	0.02*						0.52	0.19	0.007[†]	
Late group	−0.11	0.14	0.43						−0.03	0.16	0.86	

Sensitivity analyses excluded 24 women who reported menopause before age 40 y and 4 women who reported menopause after 59 y. Values for progesterone concentrations were log-transformed before analyses. There were 264 women in the early postmenopause group and 343 women in the late postmenopause group. Analyses were adjusted for age, smoking history, systolic blood pressure, and apolipoprotein E genotype.
*Probability P < 0.05.
[†]Probability P < 0.01.