



Published in final edited form as:

Magn Reson Imaging. 2011 May ; 29(4): 546–553. doi:10.1016/j.mri.2010.12.001.

Application of machine learning methods to describe the effects of conjugated equine estrogens therapy on region-specific brain volumes

Ramon Casanova¹, Mark A. Espeland¹, Joseph S. Goveas², Christos Davatzikos³, Sarah A. Gaussoin¹, Joseph A. Maldjian⁴, Robert L. Brunner⁵, Lewis H. Kuller⁶, Karen C. Johnson⁷, W. Jerry Mysiw⁸, Benjamin Wagner⁴, Susan M. Resnick⁹, and Women's Health Initiative Memory Study

¹Department of Biostatistical Sciences, Wake Forest University School of Medicine, Winston-Salem, NC

²Department of Psychiatry, Medical College of Wisconsin, Milwaukee, WI

³Department of Radiology, University of Pennsylvania, Philadelphia, PA

⁴Department of Radiology, Wake Forest University School of Medicine, Winston-Salem, NC

⁵Family and Community Medicine, University of Nevada School of Medicine, Reno, NV

⁶Department of Epidemiology, University of Pittsburgh, Pittsburgh, PA

⁷University of Tennessee at Memphis, Memphis, TN

⁸Department of Physical Medicine and Rehabilitation, Ohio State University School of Medicine, Columbus, OH

⁹Laboratory of Personality and Cognition, Intramural Research Program, National Institute on Aging, NIH, 5600 Nathan Shock Drive, Baltimore, MD

Abstract

Use of conjugated equine estrogens (CEE) has been linked to smaller regional brain volumes in women aged ≥ 65 years, however it is unknown whether this results in a broad-based characteristic pattern of effects. Structural MRI was used to assess regional volumes of normal tissue and ischemic lesions among 513 women who had been enrolled in a randomized clinical trial of CEE therapy for an average of 6.6 years, beginning at ages 65–80 years. A multivariate pattern analysis, based on a machine learning technique that combined Random Forest and logistic regression with L_1 penalty, was applied to identify patterns among regional volumes associated with therapy and whether patterns discriminate between treatment groups. The multivariate pattern analysis detected smaller regional volumes of normal tissue within the limbic and temporal lobes among women that had been assigned to CEE therapy. Mean decrements ranged as high as 7% in the left entorhinal cortex and 5% in the left perirhinal cortex, which exceeded the effect sizes reported previously in frontal lobe and hippocampus. Overall accuracy of classification based on these patterns, however, was projected to be only 54.5%. Prescription of CEE therapy for an average of 6.6 years is associated with lower regional brain volumes, but it does not induce a characteristic spatial pattern of changes in brain volumes of sufficient magnitude to discriminate users and non-users.

Keywords

Hormone therapy; MRI; Random Forest; WHIMS

Introduction

Conjugated equine estrogen (CEE) increases the risks of dementia and mild cognitive impairment in women aged 65 years or older and is associated with deficits in cognitive function (1,2). Additionally, relative to placebo, CEE-based therapy is associated with smaller mean hippocampal and frontal lobe volumes, but not ischemic lesion volumes (3,4). It is not known whether CEE therapy differentially affects other regions to produce distinctive spatial patterns of reduced volumes and/or lesions..

This work has two main objectives: first, to test the primary hypothesis about the existence of spatial patterns of tissue atrophy and/or lesions that are associated to the CEE-based therapy that could have been missed in previous work due the coarse spatial resolution that was employed; the second goal is largely methodological. We introduce a new sophisticated multivariate technique based on Random Forest (RF) and penalized logistic regression to analyze MRI data, which we have developed to overcome some of the challenges in identifying patterns across large numbers of potential predictors (5,6).

RF belongs to the category of the so called ensemble methods for classification because a committee of learners (e.g. trees) is generated and each one casts a “vote” to classify cases. Trees are built using classification and regression trees methodology (CART). RF has several properties that explain its increasing popularity in bioinformatics: 1) can be used when there are more variables than observations; 2) can deal with both two class and multi-class problems; 3) does not overfit; 4) can handle mixtures of categorical and continuous variables; 5) contains a built-in cross-validation method; 6) generates measures to evaluate variable importance (7). Precisely this last RF property is used in this work by making use of the RF’s permutation importance index to obtain ranks of the MRI volumetric measures. On the other hand, the L_1 penalized logistic regression is a machine learning technique that allows dealing with many variables and forces to zero non-relevant variables via an embedded mechanism .

Differently from our previous work (3,4) we conducted our analyzes at a much finer spatial scale by using 157 measurements of regional volumes of normal tissue and ischemic lesions from magnetic resonance images acquired from women 1-4 years after they had participated in a large randomized placebo-controlled clinical trial of CEE therapy to identify treatment-related differences.

SUBJECTS and METHODS

The Women’s Health Initiative Memory Study (WHIMS), an ancillary study to the Women’s Health Initiative (WHI), consisted of parallel placebo-controlled randomized clinical trials of 0.625 mg/day CEE with and without 2.5 mg/day medroxyprogesterone acetate (MPA) in women with a uterus or post-hysterectomy (8). Participants were recruited from 39 clinical centers of the WHI CEE-Alone and CEE+MPA clinical trials (9,10). They were 65 to 79 years of age and free of dementia. Written informed consent was obtained. The National Institutes of Health and Institutional Review Boards of participating institutions approved the protocols and consent forms.

Modified MiniMental State (3MS) exams (11) were administered at baseline by trained technicians. Information on other pre-trial risk factors for cognitive decline was obtained by standardized interviews and self-reports (12).

This report is limited to the CEE-Alone trial to avoid possible influences of MPA. This trial was terminated early (February, 2004) due to an increased risk of stroke and lack of evidence for prevention of cardiovascular disease (10).

MRI acquisition and image processing

The WHIMS Magnetic Resonance Imaging (WHIMS-MRI) study was designed to contrast MRI outcomes between women who had been assigned to active versus placebo therapy during the WHIMS trials in 14 of its clinical sites (13). Exclusion criteria included presence of pacemakers, other implants, foreign bodies, and other contraindications for MRI. None of those scanned had been classified as having cognitive impairment at WHIMS enrollment. A total of 1,403 women completed MRI scans that met central reading criteria for analysis, of which 520 were from the CEE-Alone trial (3,4). Complete data on dementia risk factors were available on 513 (98.6%) of these women, who are described in this report. These women had MRIs an average (range) of 8.0 (6.5, 9.3) years and 1.4 (0.8, 2.2) years after trial enrollment and termination.

Standardized protocols were used for acquisition and processing of MRI scans and for measuring volumes (3,4). Images were acquired with field of view=22 cm and matrix size=256×256. The sequences included oblique axial spin density/T2-weighted spin echo (TR = 3200 ms, TE = 30/120 ms, slice thickness = 3 mm), FLAIR T2-weighted spin echo (TR = 8000 ms, TI = 2000 ms, TE = 100 ms, slice thickness = 3 mm), and oblique axial 3D T1-weighted gradient echo (TR = 21 ms, TE = 8 ms, flip angle = 30 degrees, slice thickness = 1.5 mm) acquired from the vertex to the skull base parallel to the anterior commissure-posterior commissure (AC-PC) plane.

To quantify regional brain volumes, the T1-weighted volumetric MRI scans were first preprocessed according to a standardized protocol (14): 1) alignment to the AC-PC orientation; 2) removal of extracranial material; and 3) segmentation of brain parenchyma into gray matter (GM), white matter (WM), and CSF. Regional volumetric measurements of GM, WM, and CSF were subsequently obtained via a validated, automated computer-based template warping method. This technique is based on a digital atlas labeled for brain lobes and individual structures, including the hippocampus. Atlas definitions were transferred to MRI scans via an image-warping algorithm called HAMMER that performs pattern matching of anatomically corresponding brain regions (15). The volumes of GM, WM, and CSF of each labeled brain region were obtained by summing the number of respective voxels within each region. Volumes of brain lesions and periventricular abnormal WM were also measured separately via the same procedure, using the three sets of images; total lesion volume was measured, as described in (4). Intracranial volume was estimated as the total cerebral hemispheric volumes including ventricular cerebrospinal fluid and the cerebral spinal fluid within the sulcal spaces.

Analytical methods

We used a multivariate pattern analysis method that combined Random Forest (RF) and logistic regression with L_1 penalty (LR- L_1) to identify regional volumes that were useful to discriminate women's treatment assignment. Our approach consists of a two-step backward elimination scheme. First, RF was used to rank the regional brain volumes according to their importance for classifying women based on the RF permutation importance index, which is one of the measures generated by RF for predictor relevance identification. Then, LR- L_1

was used to assess the relevance of each subset of predictors via its classification accuracy. These two steps were embedded into a 10-fold cross-validation procedure to reduce selection bias (16). One-by-one, MRI measures were eliminated based on the RF ranking to reach the smallest subset that maintained the maximum overall accuracy. We characterized the performance of our method in terms of classification accuracy, power, and false positive rate using simulations. These simulations explored how differences in the number of relevant predictors, their signal-to-noise ratio, and the heterogeneity of relationships between predictors and treatment influenced performance. These details appear in an appendix.

Although prior analyses of WHIMS-MRI data found little evidence that CEE-based therapies influenced ischemic lesion volumes, these findings were based on a limited number of regions. We thus chose to include both volumes of ischemic lesions and normal tissues in our analyses: 157 total volumes (see Table 1). Each individual volume was divided by the intracranial volume to adjust for difference in skull sizes and adjusted for age, education level, body mass index, and baseline 3MS scores using linear regression.

To portray mean differences in volumes from the regions identified by our approach, we used analyses of covariance to gauge the relative magnitudes of observed effects, with adjustment for the above factors.

RESULTS

Women had been enrolled in the CEE-Alone trial for an average (range) of 6.6 (5.5, 7.7) years. Table 2 describes their risk factors for cognitive decline and dementia, which were balanced by original trial randomization. Fifty percent reported prior use of hormone therapy; 21% had used it for more than 10 years. Among those reporting prior use, 52% had begun therapy at or before the time of their last menstrual period. During follow-up, these women met the study criterion for adherence to their assigned treatment ($\geq 80\%$ of expected, based on pill counts) an average of 70% of the time.

Our analytical approach identified volumes that optimally discriminated between treatment groups, with a projected overall accuracy of 53.1% for women assigned to CEE therapy and 56.4% for women assigned to placebo: 54.5% overall. The selected regional volumes appear in Table 3. Assignment to CEE therapy was associated with smaller regional volumes of normal brain tissue in the limbic cortex (left cingulate and left perirhinal cortex) and temporal lobe (left entorhinal cortex and right inferior temporal gyrus); unadjusted analyses of covariance yielded $p < 0.05$. Our approach also identified parietal lobe regions (left angular gyrus and right superior parietal lobe) for which CEE-based therapy was associated with slightly larger volumes of normal tissue, i.e. regions for which no adverse treatment effects were evident. Only one regional ischemic lesion volume (total occipital lobe) was selected as being potentially useful for classification of treatment assignment, however this difference was not pronounced (unadjusted $p = 0.4383$).

We grouped women according to whether they were classified correctly $\geq 50\%$ of the time across 10 runs of the multivariate pattern analysis and examined whether rates varied among subgroups defined by risk factors for cognitive impairment (Table 4). Classification rates were not markedly influenced by age, hypertension, or education. Women who had been assigned to CEE therapy were more likely to be classified correctly if they had higher baseline 3MS scores. Women who had been assigned to placebo were more likely to be classified correctly if they had lower baseline 3MS or prior cardiovascular disease. Classification accuracy was not related to pre-trial use of hormone therapy.

Because our approach did not yield classification rules that performed much better than chance, it was important to confirm its validity. We were concerned whether heterogeneity in associations between CEE therapy and regional volumes, such as might occur if adverse effects were limited to subgroups of women, might diminish performance and whether our sample size was adequate to identify predictors. As described in the appendix, we found that the approach was fairly robust even when heterogeneity was marked or the number of predictors varied. We projected that as few as 300 participants were sufficient to provide powers ranging from 69% to 98% for signal-to-noise ratios ≥ 1 and false positive rates from 3% to 16%.

DISCUSSION

We previously reported that assignment to 4-6 years of CEE-based therapy was associated with sustained smaller mean regional volumes of approximately 1% in the frontal lobe overall and 2% in the hippocampus (3). The current analyses sought more pronounced imprints of CEE therapy on normal and ischemic lesion volumes across an extensive atlas of brain regions that could be used to identify users.

We confirmed that our analytical approach had reasonable levels of statistical power to identify regions useful for classifying women with respect to treatment. We report three findings of potential interest. First, the adverse effects of CEE therapy appeared to be greatest for brain volumes of normal tissue within several limbic and temporal lobes regions. Second, we found little evidence of associations between ischemic lesion volumes and CEE therapies. Third, classification rates were highest among subgroups of participants defined by pre-treatment cognitive function and prior cardiovascular disease.

We confirmed that CEE therapy was associated with smaller volumes of normal tissue in some regions, with mean decrements ranging as high as 7% in the left entorhinal cortex (ERC) and 5% in the left perirhinal cortex. These exceed the effect sizes originally reported for the regions targeted in the original WHIMS-MRI report, i.e. frontal lobe, hippocampus, and overall (3). This suggests that the adverse impact of treatment may be more diffuse than originally reported and may differentially affect particular sub-regions of the brain. The magnitudes of the effects we detected, however, were not sufficiently large to classify women accurately according to treatment assignment, which is a stronger requirement. As an example, for Gaussian distributions, a classification accuracy of 54.5% is associated with a mean shift of approximately 0.25 standard deviations: for relatively large sample sizes, such as ours, this mean shift would be expected to reach statistical significance, as in our original report, despite the relatively modest discrimination it provides. Thus, the effects of CEE therapy on brain structure are generally not so pronounced as to be easily distinguished from other influences. For women who convert to cognitive impairment following CEE-based therapy, however, the affects may be more strongly expressed as relative decrements in regional brain volumes, at least in the hippocampus (28).

Like the hippocampal formation, the ERC plays a key role in episodic memory function and is among the earliest brain region affected by AD neuropathology (17,18). ERC volume loss may be a better neuroimaging marker than hippocampal or frontal lobe atrophy in identifying middle-aged and older adults who are at-risk of memory decline (19) and disease progression to dementia (20-22). A higher degree of AD-type neuropathological lesions in the ERC and marked loss of layer II ERC neurons are found very early in persons at-risk to develop AD, relative to the normal elderly (23,24). ERC volume decline precedes hippocampal atrophy in patients with MCI who eventually progress to AD when compared with MCI non-converters and healthy normal subjects (20). Thus, the greater ERC volume

declines relative to hippocampal and frontal lobe atrophy we saw may suggest a causative mechanism that links CEE use and incident dementia among postmenopausal women.

We are intrigued by the laterality in our findings, which highlight atrophy in the left ERC and perirhinal cortex. Although in general, patients with AD have symmetric amyloid deposition in the cortical and subcortical structures, a higher degree of amyloid deposition in the left brain structures has previously been demonstrated in individuals at-risk for AD (25) and greater left parietal and temporal lobe atrophy is seen in MCI converters when compared to non-converters (26,27). Moreover, greater left parietal lobe hypometabolism is seen in APOE epsilon 4 carriers who are cognitively healthy relatives of AD patients (28).

CEE-based therapies increase the risk of stroke, regardless of age (29). Our analyses confirm the finding of Coker et al., however, that these therapies did not result in markedly increased volumes of subclinical lesion volumes in women in the WHI (4).

The multivariate pattern analysis appeared to perform best for identifying CEE therapy among women who had relatively high pre-treatment global cognitive function. This may seem, at first, counterintuitive in that CEE therapy has been found to have smaller average effects on cognitive function and brain volumes in such women (2,3). However, even within this subgroup of women, CEE therapy is associated with increased risks for cognitive impairment and atrophy (1,30). It may be that the expression of treatment effects are variable in this subgroup of women and that when present to a sufficient degree, they are more detectable against the backdrop of the relatively health brains of women who are not exposed to CEE therapy. Likewise, our classification algorithms appeared to be more successful in identifying placebo therapy among women with lower pre-treatment cognitive function and prior cardiovascular disease. It may be that it identifies, among these women, those whose lower cognitive abilities are associated with vascular disease and not atrophy.

Our findings may reassure women over the age of 65 years who have used CEE therapy. While CEE therapy is associated with increased risk of cognitive impairment in this age range and is not recommended for use other than treatment of menopausal symptoms, its typical affect on brain structure, while statistically significant as in the primary WHIMS report, appears not to be large. It may be that the effects of CEE therapy on regional brain volumes are highly variable among women so that no characteristic patterns emerge. While we cannot discount this, we feel it is more likely that the effect sizes, although reaching nominal levels of statistical significance, are just too small relative to other sources of variability among women to provide effective discrimination.

The machine learning technique introduced in this work based on Random Forests and penalized logistic regression produced good power and control of false positives in the situations tested by our simulations, even when signal to noise ratios were low due to the presence of heterogeneity in the data. Heterogeneity is to be expected in settings such as WHIMS where the mechanisms that define the effects of the CEE drugs are very complex and very likely different for different groups of women (31).

Acknowledgments

WHIMS-MRI Clinical Centers: Albert Einstein College of Medicine: Sylvia Wassertheil-Smoller; Medical College of Wisconsin, Milwaukee: Jane Morley Kotchen; Stanford Center for Research in Disease Prevention: Marcia L. Stefanick; The Ohio State University: Rebecca Jackson; University of California at Davis: John Robbins; University of California at Los Angeles: Lauren Nathan; University of Florida: Marian Limacher; University of Iowa: Jennifer Robinson; University of Massachusetts: Judith Ockene; University of Minnesota: Karen Margolis; University of Nevada: Robert Brunner; University of North Carolina, Chapel Hill: Carol Murphy; University of Pittsburgh: Lewis Kuller.

WHIMS-MRI Clinical Coordinating Center: Wake Forest University Health Sciences: Sally Shumaker.

WHIMS-MRI Quality Control Center: University of Pennsylvania: Nick Bryan.

U.S. National Institutes of Health: National Institute on Aging: Neil Buckholtz, Susan Molchan, Susan Resnick; National Heart, Lung, and Blood Institute: Jacques Rossouw, Linda Pottern.

Funding: The Women's Health Initiative is funded by the National Heart, Lung, and Blood Institute of the National Institutes of Health, U.S. Department of Health and Human Services. WHIMS was funded in part by Wyeth Pharmaceuticals, Inc, St. Davids, PA. SMR is supported by the Intramural Research Program, NIA, NIH.

APPENDIX

Analytical Details and Performance Assessment

Detailed descriptions of RF and LR- L_1 methods are in (32-33). We used the R package randomForest (34) considered one of the best RF implementations available (35) and the LIBLINEAR toolbox (36) implementation of LR- L_1 .

Multivariate Analysis Algorithm

Our method can be summarized, as follows.

1. Loop K-fold cross-validation

For each fold, compute a ranking of predictors based on RF using the training data set. N_{rf} RFs were computed, the corresponding permutation importance scores were averaged and the predictors were sorted according to these average scores. The predictors with average scores ≤ 0 were discarded. We used the default values for RF parameters in the randomForest package: number of trees (*ntree*) 500 and the number of variables sampled as candidates at each split (*mtry*) equal \sqrt{p} , $p = 157$.

Loop across # of predictors [$p, p - 1, \dots, 3, 2, 1$], the top features in the RF ranking.

- a. Train the LR1 classifier using grid search with K-fold cross-validation to determine the optimal regularization parameter C based on the maximum overall accuracy.
- b. Apply the classifier obtained in step a) to the test data and obtain estimates of the overall and intra-class accuracies.

end

end

2. Average the overall and intra-class accuracies across the K folds for each subset of predictors.
3. Estimate the optimal number features (N_{opt}) as the size of the smallest subset of predictors that produced the maximum overall accuracy.
4. Compute N_{rf} RFs using the whole data set, averaging the permutation importance scores. Sort the predictors to these average scores.
5. Selected top N_{opt} predictors in this ranking.

Simulations

We created simulated datasets with 300 women, evenly divided into two groups, and $p = 100$ potential predictors. We studied our method's performance under different signal-to-noise ratios and fractions of relevant predictors and the influence of heterogeneity in relationships between predictors and group assignment among women, calculating the mean and standard deviation (SD) of its overall accuracy, power, and false positive rate from 30 simulated datasets as:

$$Acc = \frac{CCC}{TC} \times 100$$

$$P = \frac{TP}{\#Target \ Features} \times 100$$

$$FPR = \frac{FP}{FP+TN} \times 100$$

where CCC is the number of correctly classified cases, TC is the total number of cases, TP is the number of correctly identified relevant features, $\#Target \ Features$ is the number of relevant predictors present in the data, FP is the number of predictors incorrectly identified as relevant and TN is the number of features correctly identified as non-relevant. All simulations used the randomForest package default values for the main RF parameters ($n_{tree} = 500$, $m_{try} = \sqrt{p}$) and the number of CV folds was fixed to 5.

Simulation 1

This simulation tested our method under different signal to noise ratios. The predictors are independently and identically distributed (i.i.d) generated from $N(0,1)$. Ten predictors (10%) were fixed to be relevant by adding different constant values ($A = 2, 1, 0.5, 0$) to all the subjects in the target class ($Y = -1$). We fixed N_{rf} equal 20.

Simulation 2

This simulation assessed performance when different fractions of features were relevant. The predictors were i.i.d generated from $N(0,1)$. We fixed the signal-to-noise ratio to the lower level we analyzed in simulation 1 ($A = 0.5$) excluding the no signal case. We varied the fraction of relevant features to be 0.05, 0.1, 0.2 and 0.4. We fixed N_{rf} equal 20.

Heterogeneity

All the above simulations were repeated without and with heterogeneity. Predictors were i.i.d generated from $N(0,1)$. We simulated 50% heterogeneity by adding the signal to one-half of the relevant predictors in one-half of the subjects of the target class ($Y = -1$) and for the second half of the relevant predictors the signal is added to the second half of the subjects in the same class.

Results

Table A.1 presents results from the first simulations. In both situations (heterogeneity present or not) the mean overall accuracy decreased as the signal-to-noise ratio decreased. Heterogeneity, as expected, also decreased the overall accuracy. False positive rates (FPR) tended to increase as signal-to-noise ratio decreased, which was also expected. The power generally followed a similar trend, with the exception when the signal-to-noise ratio was 2. There are two reasons for this, both related to the performance of the regularized logistic regression as a wrapper. In general, regularization makes the classifier more robust to the presence of noisy predictors. Also, when the signal-to-noise ratio is high, fewer relevant predictors will be sufficient to produce high classification accuracy rates. This is a good characteristic from the point of view of classification accuracy, but decreases the wrapper's sensitivity to the correct number of relevant predictors.

Table A.2 shows the results of the second simulation that evaluates performance for different fractions of relevant predictors. An increased fraction of relevant predictors leads to an improvement of the overall classification accuracy in both situations. This is to be expected, since the amount of information that allows the classifier to discriminate between the two classes increases with the fraction of relevant predictors. The power and false positive rates of the method decrease with the increase of the fraction of relevant predictors when no heterogeneity is present. A different pattern for both power and false positive rates is observed when heterogeneity is present. A maximum value is obtained when 20% of the predictors are relevant.

Table A.1

Mean (standard deviation) projected overall accuracy, power and false positive rate across 30 realizations of data for different signal-to-noise ratios

SNR levels	No Heterogeneity			Heterogeneity		
	Accuracy (%)	Power (%)	FPR (%)	Accuracy (%)	Power (%)	FPR (%)
2	99.9 (0.2)	69.3 (11.2)	0 (0)	94.3 (1.1)	96.7 (6.6)	1.7 (5.1)
1	93.7 (1.7)	98.0 (5.0)	5.2 (10.0)	76.3 (2.0)	89.0 (10.3)	2.7 (4.8)
0.5	77.2 (2.4)	86.7 (12.5)	3.2(5.0)	59.6 (4.2)	47.3 (20.0)	8.7 (6.9)
No signal	53.1 (2.7)	-	14.0 (10.0)	52.6 (2.7)	-	8.7 (7.3)

Table A.2

Mean (standard deviation) projected overall accuracy, power and false positive rate across 30 realizations of data for different fractions of relevant features is presented. The SNR level was selected to be 0.5

Proportion of Relevant Predictors	No Heterogeneity			Heterogeneity		
	Accuracy (%)	Power (%)	FPR (%)	Accuracy (%)	Power (%)	FPR (%)
0.05	69.5 (3.0)	90.0 (14.7)	6.8 (9.4)	57.0 (3.2)	43.3 (17.5)	9.5 (8.0)
0.1	77.5 (2.2)	91.0 (11.4)	5.2 (6.3)	62.0 (3.4)	53.7 (20.3)	13.7 (9.0)
0.2	85.0 (2.5)	86.2 (11.1)	3.5 (4.2)	65.6 (3.1)	57.8 (19.8)	14.5 (9.4)
0.4	90.5 (1.9)	84.8 (11.0)	3.2 (4.9)	70.2 (2.9)	51.4 (16.0)	8.9 (7.6)

REFERENCES

1. Shumaker SA, Legault C, Kuller L, et al. Conjugated equine estrogens and incidence of probable dementia and mild cognitive impairment in postmenopausal women: Women's Health Initiative Memory Study. *JAMA*. 2004; 291:2947–2958. [PubMed: 15213206]
2. Espeland MA, Rapp SR, Shumaker SA, et al. Conjugated equine estrogens and global cognitive function in postmenopausal women. *JAMA*. 2004; 291:2959–2968. [PubMed: 15213207]
3. Resnick SR, Espeland MA, Jaramillo SA, et al. Effects of postmenopausal hormone therapy on regional brain volumes in older women: The Women's Health Initiative Magnetic Resonance Imaging Study (WHIMS-MRI). *Neurology*. 2009; 72:135–142. [PubMed: 19139364]
4. Coker LH, Hogan PE, Bryan NR, et al. The effects of postmenopausal hormone therapy on volumetric sub-clinical cerebrovascular disease: The Women's Health Initiative Memory Study - Magnetic Resonance Imaging Study (WHIMS-MRI). *Neurology*. 2009; 72:125–134. [PubMed: 19139363]
5. Donoho, D. High-Dimensional Data Analysis: The Curses and Blessings of Dimensionality; Lecture on August 8,2000, To the American Mathematical Society ' Math Challenges of the 21st Century;
6. Clarke R, Ressom HW, Wang A, Xuan J, Liu MC, Gehan EA, Wang Y. The properties of high-dimensional data spaces: implications for exploring gene and protein expression data.
7. Diaz-Uriarte R, Alvarez de Andres S. Gene selection and classification of microarray data using random forest. *BMC Bioinformatics*. 2006
8. Shumaker SA, Reboussin BA, Espeland MA, et al. The Women's Health Initiative Memory Study: a trial of the effect of estrogen therapy in preventing and slowing the progression of dementia. *Control Clin Trials*. 1998; 19:604–621. [PubMed: 9875839]
9. Writing Group for the Women's Health Initiative Investigators. Risks and benefits of estrogen plus progestin in healthy postmenopausal women: principal results from the Women's Health Initiative randomized controlled trial. *JAMA*. 2002; 288:321–333. [PubMed: 12117397]
10. Women's Health Initiative Steering Committee. Effects of conjugated equine estrogen in postmenopausal women with hysterectomy. *JAMA*. 2004; 291:1701–1712. [PubMed: 15082697]
11. Teng EL, Chui HC. The Modified Mini-Mental State (3MS) examination. *J Clin Psychiatry*. 1987; 48:314–318. [PubMed: 3611032]
12. The Women's Health Initiative Study Group. Design of the Women's Health Initiative clinical trial and observational study. *Control Clin Trials*. 1998; 19:61–109. [PubMed: 9492970]
13. Jaramillo SA, Felton D, Andrews LA, et al. Enrollment in a brain magnetic resonance study: results from the Women's Health Initiative Memory Study Magnetic Resonance Imaging Study (WHIMS-MRI). *Acad Radiol*. 2007; 14:603–612. [PubMed: 17434074]
14. Goldszal AF, Davatzikos C, Pham DL, Yan MX, Bryan RN, Resnick SM. An image-processing system for qualitative and quantitative volumetric analysis of brain images. *J Comput Assist Tomogr*. 1998; 22:827–837. [PubMed: 9754125]
15. Shen D, Davatzikos C. HAMMER: hierarchical attribute matching mechanism for elastic registration. *IEEE Trans Med Imaging*. Nov; 2002 21(11):1421–39. [PubMed: 12575879]
16. Ambrose C, McLachlan GJ. Selection bias in gene extraction on the basis of microarray gene-expression data. *Proc Natl Acad Sci*. 2002; 99:6562–6566. [PubMed: 11983868]
17. Braak H, Braak E. Staging of Alzheimer's disease-related neurofibrillary tangles. *Neurobiol Aging*. 1995; 16:271–84. [PubMed: 7566337]
18. Lee BCP, Mintun M, Buckner RL, Morris JC. Imaging of Alzheimer's disease. *J Neuroimaging*. 2003; 13:199–214. [PubMed: 12889165]
19. Rodrigue KM, Raz N. Shrinkage of the entorhinal cortex over five years predicts memory performance in healthy adults. *J Neurosci*. 2004; 24:956–963. [PubMed: 14749440]
20. Dickerson BC, Feczko E, Augustinack JC, et al. Differential effects of aging and Alzheimer's disease on medial temporal lobe cortical thickness and surface area. *Neurobiol Aging*. 2009; 30:432–440. [PubMed: 17869384]
21. deToledo-Morrell L, Stoub TR, Bulgakova M, et al. MRI-derived entorhinal volume is a good predictor of conversion from MCI to AD. *Neurobiol Aging*. 2004; 25:1197–1203. [PubMed: 15312965]

22. Devanand DP, Pradhaban G, Liu X, Khandji A, et al. Hippocampal and entorhinal atrophy in mild cognitive impairment: prediction of Alzheimer disease. *Neurology*. 2007; 68:828–836. [PubMed: 17353470]
23. Gómez-Isla T, Price JL, McKeel DW Jr, Morris JC, Growdon JH, Hyman BT. Profound loss of layer II entorhinal cortex neurons occurs in very mild Alzheimer's disease. *J Neurosci*. 1996; 16:4491–4500. [PubMed: 8699259]
24. Kordower JH, Chu Y, Stebbins GT, et al. Loss and atrophy of layer II entorhinal cortex neurons in elderly people with mild cognitive impairment. *Ann Neurol*. 2001; 49:202–213. [PubMed: 11220740]
25. Raji CA, Becker JT, Tsopelas ND, et al. Characterizing regional correlation, laterality and symmetry of amyloid deposition in mild cognitive impairment and Alzheimer's disease with Pittsburgh Compound B. *J Neurosci Meth*. 2008; 172:277–282.
26. Karas G, Sluimer J, Goekoop R, et al. Amnesic mild cognitive impairment: structural MR imaging findings predictive of conversion to Alzheimer disease. *Am J Neuroradiol*. 2008; 29:944–949. [PubMed: 18296551]
27. Risacher SL, Saykin AJ, West JD, Shen L, Firpi HA, McDonald BC. Baseline MRI predictors of conversion from MCI to probable AD in the ADNI cohort. *Curr Alzheimer Res*. 2009; 6:347–361. [PubMed: 19689234]
28. Small GW, Mazziotta JC, Collins MT, et al. Apolipoprotein E type 4 allele and cerebral glucose metabolism in relatives at risk for familial Alzheimer disease. *JAMA*. 1995; 273:942–947. [PubMed: 7884953]
29. Rossouw JE, Prentice RL, Manson KE, et al. Postmenopausal hormone therapy and risk of cardiovascular disease by age and years since menopause. *JAMA*. 2007; 297:1465–1477. [PubMed: 17405972]
30. Espeland MA, Tindle HA, Bushnell CA, et al. Regional brain and ischemic lesion volumes in women with cognitive impairment following exposure to conjugated equine estrogen therapies: The Women's Health Magnetic Resonance Imaging Study (WHIMS-MRI). *J Gerontol Med Sci*. 2009; 64:243–1250.
31. Turgeon JL, Carr MC, Maki PM, Wendelsohn NE, Wise PM. Complex actions of sex steroids in adipose tissue, the cardiovascular system, and brain: insights from basic science and clinical studies. *Endocr Rev*. 2006; 27:575–605. [PubMed: 16763155]
32. Breiman L. Random forests. *Machine Learning*. 2001; 45:5–32.
33. Hastie, T.; Tibshirani, R.; Friedman, J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer; 2009.
34. Liaw A, Wiener M. Classification and regression by randomForest. *Rnews*. 2002; 2:18–22.
35. Siroky DS. Navigating random forests and related advances in algorithmic modeling. *Statist Surv*. 2009; 3:147–163.
36. Fan RE, Chang KW, Hsieh CJ, Wang XR, Lin CJ. LIBLINEAR: A library for large linear classification. *J Mach Learning Res*. 2008; 9:1871–1874.

Table 1

Analyzed brain regions (N=157): each contributed four volumes (right and left sides; abnormal and normal tissue) except the corpus callosum which contributed a single. Note that total volumes are also included for lobes containing at least five regions (frontal, occipital, parietal, and temporal lobes)

Lobe	Region
Basal Ganglia	Caudate nucleus, Globus palladus, Putamen, Thalamus
Frontal	Inferior frontal gyrus, Middle frontal gyrus, Superior frontal gyrus, Lateral orbital frontal gyrus, Medial orbital frontal gyrus, Medial frontal gyrus, Precentral gyrus Total frontal lobe
Occipital	Inferior occipital gyrus, Middle occipital gyrus, Superior occipital gyrus, Lateral occipital temporal gyrus, Cuneus, Medial occipital gyrus, Occipital pole, Total occipital lobe
Parietal	Angular gyrus, Postcentral gyrus, Precuneus, Superior parietal lobule, Supramarginal gyrus, Total parietal lobe
Limbic	Cingulate region, Insula, Perirhinal cortex
Temporal	Inferior temporal gyrus, Middle temporal gyrus, Superior temporal gyrus, Entorhinal cortex, Hippocampal formation, Parahippocampal gyrus, Uncus, Lingual gyrus Temporal pole, Total temporal lobe
Corpus Callosum	Corpus callosum

Table 2

Characteristics of the cohort of women from the CEE-Alone trial

Risk Factor	CEE-Alone Trial Cohort		
	Treatment Assignment Mean (SD) or Percent		p-value
	CEE N=257	Placebo N=263	
Age at WHI Enrollment	70.7 (3.7)	70.7 (3.8)	0.9582
Age at MRI Scan	78.6 (3.7)	78.6 (3.8)	0.9951
Duration of On-trial Follow-up	6.6 (0.6)	6.6 (0.5)	0.1842
Hypertension	47.5	52.9	0.2542
No Yes	52.5	47.2	
Prior CVD	91.8	90.9	0.7562
No Yes	8.2	9.1	
Baseline 3MS	95.6 (3.7)	95.7 (3.6)	0.1842
Education	70.4	74.5	0.3263
Not college grad College grad	29.6	25.5	
Prior hormone therapy use	51.0	48.3	0.5987
No Yes	49.0	51.7	

Table 3

Mean volumes (mm³) of the regions selected by the multivariate pattern analysis and intra-lobe sums, based on analyses of covariance with adjustment for intracranial volume, age, education, body mass index, and pre-treatment 3MS score

Regional Volume	Volumes Selected To Classify The CEE-Alone Cohort		
	Treatment Assignment Mean (SE) or Percent		Unadjusted p-value
	CEE	Placebo	
Normal tissue			
Frontal lobe regions			
Right superior frontal gyrus	6.10 (0.09)	6.24 (0.09)	0.2700
Limbic lobe subregions			
Left cingulate	8.52 (0.08)	8.80 (0.08)	0.0177
Left perirhinal cortex	1.174 (0.02)	1.23 (0.02)	0.0227
Sum of two regions above	9.695 (0.09)	10.04 (0.09)	0.0068
Occipital lobe total	21.81 (0.21)	22.25 (0.21)	0.1494
Parietal lobe			
Left angular gyrus	4.93 (0.08)	4.81 (0.07)	0.2989
Right superior parietal lobe	10.34 (0.13)	10.07 (0.13)	0.1468
Sum of two regions above	15.27 (0.19)	14.88(0.19)	0.1413
Temporal lobe			
Left entorhinal cortex	0.65 (0.01)	0.70 (0.01)	0.0004
Right inferior temporal gyrus	5.16 (0.05)	5.31 (0.05)	0.0340
Sum of two temporal lobe regions	5.81 (0.05)	6.01 (0.05)	0.0086
Ischemic lesion volume			
Occipital lobe total	0.23 (0.02)	0.21 (0.02)	0.4383

Table 4

Characteristics of women that were classified correctly or incorrectly at least 50% of the time across the ten repetitions of the multivariate pattern analysis

Risk Factor	CEE			Placebo			p-value
	Number of Women Often		p-value	Number of Women Often		p-value	
	Correctly Classified <50% N=125 (49%)	Correctly Classified ≥50% N=128 (51%)		Correctly Classified <50% N=116 (45%)	Correctly Classified ≥50% N=144 (55%)		
Age of WHI Enroll							
65-69	63/123 (53%)	58/123 (47%)	0.65	54/130 (42%)	76/130 (58%)		
70-74	42/93 (45%)	51/93 (55%)		41/90 (46%)	49/90 (54%)		0.46
75+	20/39 (51%)	19/39 (49%)		21/40 (52%)	19/40 (48%)		
Age at MRI Scan							
70-74	25/49 (51%)	24/49 (49%)	0.83	14/47 (30%)	33/47 (70%)		0.07
75-79	53/113 (47%)	58/113 (53%)		59/121 (49%)	62/121 (51%)		
80+	47/93 (51%)	46/93 (49%)		43/92 (47%)	49/92 (53%)		
Hypertension							
No	63/119 (53%)	56/119 (47%)	0.29	62/136 (46%)	74/136 (54%)		0.74
Yes	62/134 (46%)	72/134 (54%)		54/124 (44%)	70/124 (56%)		
Prior CVD							
No	114/233 (49%)	119/233 (51%)	0.60	110/236 (47%)	126/236 (53%)		0.04
Yes	11/20 (55%)	9/20 (45%)		6/24 (25%)	18/24 (75%)		
Baseline 3MS							
<95	43/70 (61%)	27/70 (39%)	0.02	26/77 (34%)	51/77 (66%)		0.02
95-100	82/183 (45%)	101/183 (55%)		90/183 (49%)	93/183 (51%)		
Education							
Not college grad	94/180 (52%)	86/180 (48%)	0.16	81/194 (42%)	113/194 (58%)		0.11
College grad	31/73 (42%)	42/73 (58%)		35/66 (53%)	31/66 (47%)		
Prior HT use							
No	69/130 (53%)	61/131 (47%)	0.32	57/126 (45%)	69/126 (45%)		0.84
Yes	59/126 (47%)	67/126 (53%)		59/134 (44%)	75/134 (56%)		