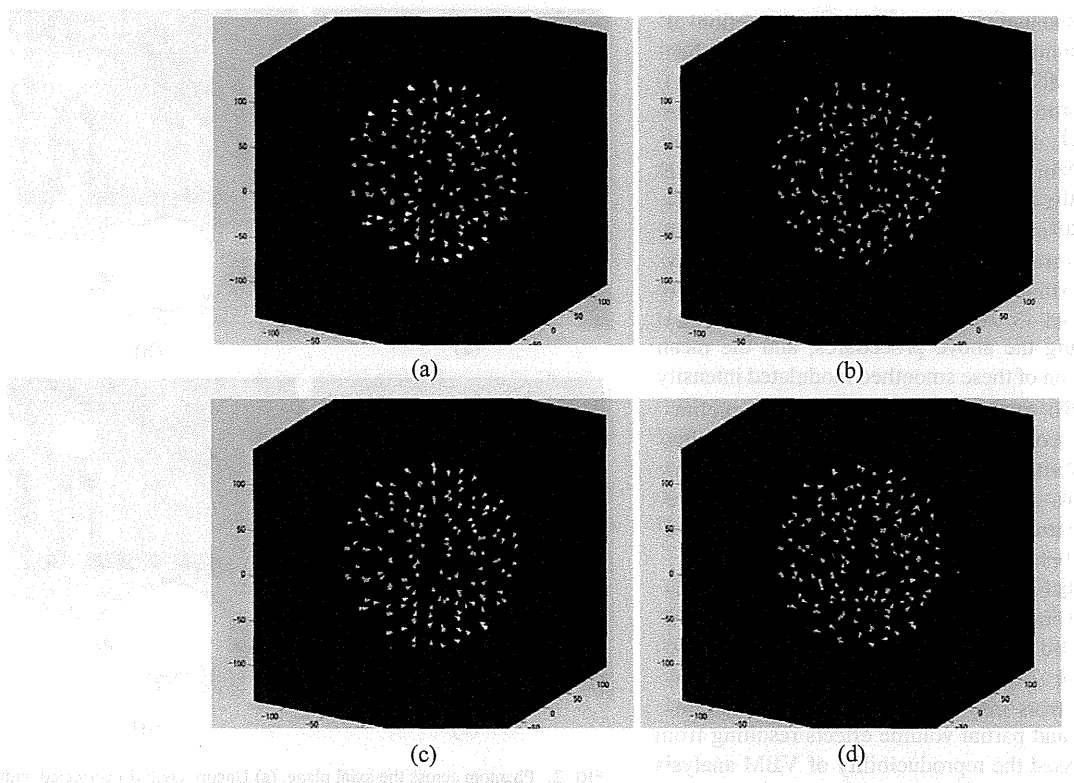

FIG. 3. Spatial distributions of fiducial sphere errors (differences) between obtained and designed coordinate values in phantom images. The arrows indicate the direction and magnitude of error. Colors indicate error levels at corresponding positions of fiducial spheres. The direction of an arrow corresponds to the basis vector with the highest magnitude of error. Red indicates error exceeding 3 mm; yellow indicates error of 2–3 mm; blue indicates error of 1–2 mm; and gray indicates error of 0–1 mm. (a) Uncorrected; (b) corrected with the J-ADNI method; (c) GW correction; and (d) GW and scaling correction.

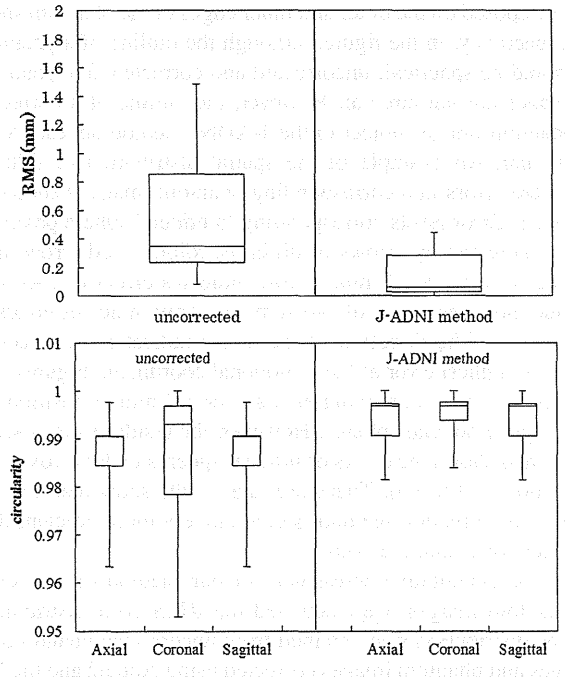

FIG. 4. RMS error and CR plots of uncorrected and corrected (J-ADNI method) phantom images ($N = 42$) obtained from all scanners for evaluation of effectiveness. (a) and (b) RMS errors obtained from analysis of phantom images by Aqual2. (c) and (d) CR plots for axial, coronal, and sagittal sections that were uncorrected (c) and corrected with the J-ADNI method (d).

for uncorrected axial images, 0.993 for uncorrected coronal images, and 0.991 for uncorrected sagittal images. After distortion correction, the median CR was 0.997 for axial images, 0.997 for coronal images, and 0.997 for sagittal images. RMS errors and the CR for all slices significantly improved (Wilcoxon signed-rank test, $p < 0.001$) after J-ADNI distortion correction.

These results show that the CR of a corrected image following application of the J-ADNI method was better than that of an uncorrected image for all slices. It is clear from Fig. 4 that the J-ADNI method improved RMS errors and the CR, and effectively corrected geometrical distortion in all phantom images.

To compare the performance of our method with that of GW correction and that of phantom-based scaling correction in addition to GW correction, we calculated RMS errors and the CR for phantom images as described previously. We obtained this information from GE Healthcare Japan and were therefore only able to apply GW correction to 11 images acquired by a group of GE scanners. The RMS errors in phase-encoding and slice-encoding directions, and CR results in the axial plane, both uncorrected and corrected with each method, are shown in Fig. 5. The top panel indicates that median RMS errors were 0.259 mm for uncorrected phantom images, 0.0976 mm for J-ADNI-corrected images, 0.268 mm for GW-corrected images, and 0.187 mm for GW-plus-scaling-corrected images along the phase-encoding direction.

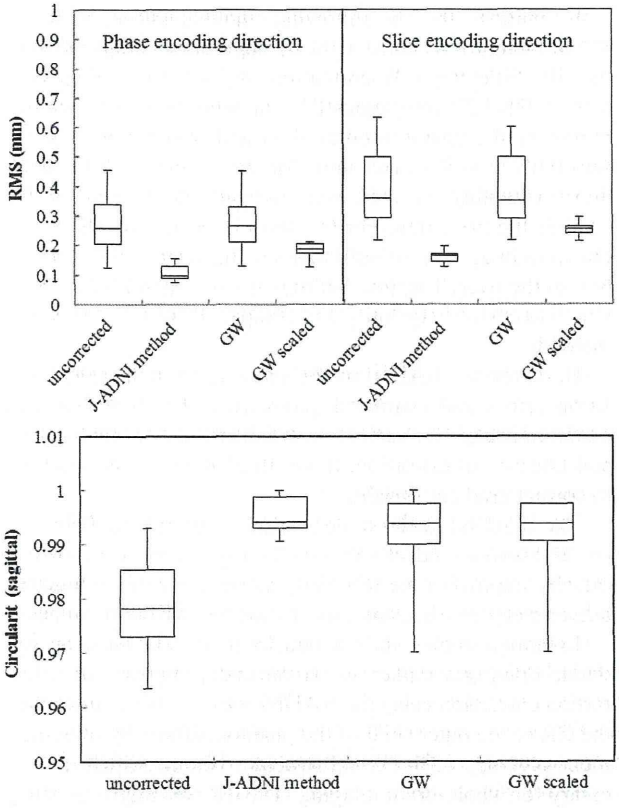

FIG. 5. RMS errors and CR plots ($N = 12$) for comparison of effectiveness of the J-ADNI, GW, and GW-plus-scaling correction methods for a confined group of GE scanners. The top panel shows RMS errors calculated by Aqual2 in the phase-encoding, A/P, direction (a) and slice-encoding, R/L, direction (b). The bottom panel shows CR plots for axial phantom images (c). These directions and (axial) plane suffered only the contribution of the gradient nonlinearity of distortion.

The corresponding values were 0.359 mm for uncorrected phantom images, 0.164 mm for J-ADNI-corrected images, 0.350 mm for GW-corrected images, and 0.255 mm for GW-plus-scaling-corrected images along the slice-encoding direction. The bottom panel shows the median CRs for the axial images: 0.978 for uncorrected images, 0.994 for J-ADNI-corrected images, 0.994 for GW-corrected images, and 0.994 for GW-plus-scaling-corrected images. The J-ADNI method improved RMS errors significantly compared with using uncorrected, GW, and GW-scaling correction methods along the phase-encoding direction and slice-encoding direction ($p < 0.001$).

CRs of the phantom image corrected with the J-ADNI method, GW, and GW-plus-scaling correction showed significant error reduction compared with the uncorrected image (Steel–Dwass test; $p < 0.001$, $p = 0.028$, and $p = 0.014$, respectively). However, there was almost no difference in the CR when comparing the J-ADNI method with the GW and GW-plus-scaling corrections despite there being CR differences for a specific image as shown in Fig. 2.

III.B. Reproducibility of VBM analysis

We performed Z-score analysis using automated VBM software. Ideally, CV for indicators in test–retest studies

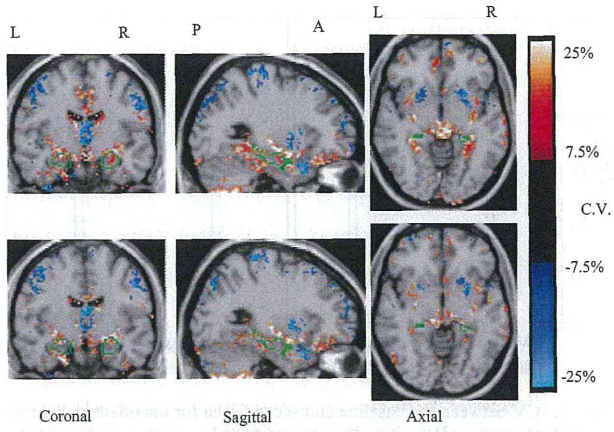

FIG. 6. Map of median CVs obtained by VBM Z-score analysis. Top and bottom rows show CV maps for an uncorrected image and a distortion-corrected image (J-ADNI method), respectively. The color-scaled CV map ranges from −25.0% to −7.5% and from 7.5% to 25.0% are displayed by overlaying orthogonal sections of the anatomically standardized MRI. Areas enclosed by green lines show medial temporal structures. Note the decrease in CV following distortion correction for the whole brain.

should be zero for a subject where there is no expectation of morphological changes associated with brain atrophy. CV maps drawn by comparing the initial and second Z-score analyses of uncorrected and corrected brain images of 38 healthy volunteers are shown in Fig. 6. The top and bottom rows show results for uncorrected and corrected brain images, respectively. In these maps, color voxels have a median value of 38 for the volunteer's CV of the Z-score between the first and second scans and the voxels are overlaid with orthogonal sections of the anatomically standardized MRI. The color scale range was −25.0% to −7.5% and 7.5% to 25.0%. Areas enclosed by green lines indicate medial temporal structures having the most significant decline of gray matter concentration at the very early stages of AD (Matsuda *et al.*[8]).

We obtained CVs of the averaged Z-score in the medial temporal structures. The extent of a region showing significant atrophy between first and second scans obtained using automated VBM software is shown in Fig. 7. The median CV of the averaged Z-score in the medial temporal structures after distortion correction using our method was 2.68%, which is not a significant improvement when comparing with uncorrected images. The median CV of the percentage of coordinates with a Z-score exceeding the threshold in the entire brain improved from 3.46% to 2.70% after distortion correction using the J-ADNI method (Wilcoxon signed-rank test, $p < 0.05$).

IV. DISCUSSION

To accurately measure the atrophy of brain structures such as the hippocampus, and to explore surrogate biomarkers obtained from MRI, geometrical distortion needs to be corrected. The US-ADNI uses two types of distortion correction methods: GW and phantom-based scaling correction. GW is only used to correct nonlinearities of geometrical distortion,

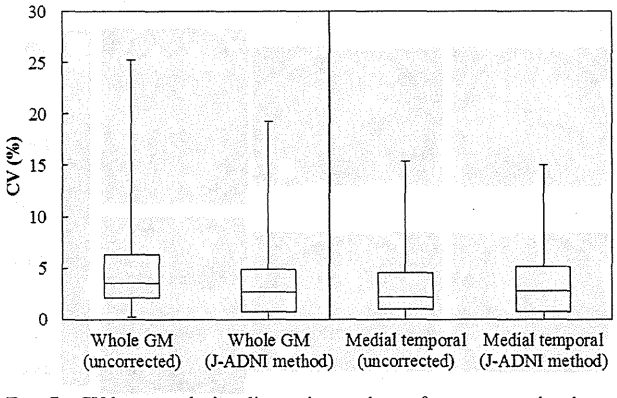

FIG. 7. CV between the baseline and second scan for uncorrected and corrected brain images ($N = 42$). The extent of atrophy variance in the whole-brain gray matter (GM) is shown (a), alongside the average Z-score for the medial temporal structures of the gray matter, which was determined using automated VBM software (b).

whereas scaling correction is used to correct linear scaling distortion.[19] According to Eq. (1), the first order of the polynomial function corresponds to the scaling factor of transformation and can correct linear scaling correction. Higher orders correct residual distortion; i.e., nonlinearities of distortion. Therefore, the J-ADNI method, which is based on a polynomial equation, can correct both nonlinear and linear geometrical distortions concurrently, and the method corresponds to phantom-based scaling correction and GW correction used by the US-ADNI.

Figure 4 shows that RMS errors and CRs in phantom studies significantly improved after J-ADNI distortion correction. In particular, RMS errors reduced by several millimeters to below the voxel resolution of 1.0 mm following application of the J-ADNI method. The structural analysis of brain images, such as voxel-based volume measurements within a region of interest, requires accuracy below the voxel resolution. Figure 4 shows that our method has enhanced reliability for volumetric brain image analysis.

We examined the ability of the J-ADNI, GW, and GW-plus-scaling correction methods to correct distortion arising from gradient nonlinearity. Figure 5 shows that RMS errors after GW correction did not change. This may be because of the efficiency limit of GW, as subsequent scaling correction reduced RMS errors significantly. These results show that the distortion arising from gradient nonlinearity includes a component that cannot be described by generic coil information and demonstrate the necessity of ADNI phantom-based scaling correction in addition to GW correction. In fact, the necessity of scaling correction through the registration of nine degrees of freedom has been reported.[19,27] Furthermore, GW corrects geometrical distortion according to the geometry of the generic gradient coil construction. If the gradient nonlinearity changes for the same scanner model, GW cannot correct this variability. Figure 5 shows that our method can reduce RMS errors more than GW and GW-plus-scaling correction. Hence, our method can describe distortion arising from gradient nonlinearity and correct it.

In contrast, the CR following implementation of the J-ADNI method was not statistically significant compared with the CRs following GW correction and GW-plus-scaling correction. The CR was obtained by drawing two circles on the border of the phantom outer shell and background, and its sensitivity was the pixel unit. Moreover, these circles were drawn manually and there were user-introduced errors at the level of the pixel unit. For the above reasons, the CR of the phantom image can be used to assess the adequacy of correction in the overall region, but may not be a good metric with which to correctly quantify and compare distortion correction methods.

However, the J-ADNI method detected fiducial sphere position errors and estimated geometrical distortion fields in phantom images for each scan, and corrected the nonlinearity and linearity of distortion. It was therefore used successfully to correct gradient variability.

The J-ADNI method demonstrates improved RMS error assessment and can thus correct geometrical distortion, thereby improving the reliability and accuracy of volumetric and/or morphometric analysis of structural MRI brain images.

In human studies, the reproducibility of VBM analysis for whole-brain gray matter was significantly improved after distortion correction using the J-ADNI method. We assumed that the CR of the outer shell of the phantom affects the measurement accuracy of the broad structure volume, which is necessary for whole-brain imaging. The CR was improved using either the J-ADNI or GW method, corresponding to improved accuracy in describing the broad structure. Geometrical distortion resulted in apparent volume changes and partial volume effects at each voxel. Our method corrected these erroneous volume changes and partial volume effects through image warping and mass-preservation resampling, respectively. Repeatability of subsequent analyses was improved after distortion correction. However, the reproducibility of VBM analysis for medial temporal structures was not significantly improved using our method because fiducial position error was more common in the outer region of the phantom image, as shown in Fig. 2. In fact, medial temporal structures originally showed low variation in VBM analysis (Fig. 6). Therefore, effects of distortion correction may be smaller than those observed for the whole brain. Alternatively, the efficacy of distortion correction is seen for the whole-brain gray matter including the region of large distortion. Our future task is to investigate the performance of our method for anatomical structures around the surface of the brain that are expected to suffer from distortion in human studies; e.g., the cortical thickness, lateral temporal lobe, and inferior parietal lobe.

The GW method requires information about the gradient coil, which can only be provided by the manufacturer, and the method can only be applied to MRI scanners produced by GE and Siemens. Our method requires scanning of a phantom image to correct geometric distortion of a subject's brain image and can be adapted to all scanners. The ADNI phantom is commercially available and can be easily purchased. However, the phantom-based correction method requires additional scanning to obtain a phantom image for correction.

A multisite study such as the ADNI can monitor scanner performance by scanning the ADNI phantom. Our method uses this scan, and we can thus correct distortion by making one phantom scan serve a dual purpose.

In future studies, we will investigate the period of validity for the distortion field obtained from a phantom image. This would assist in reducing the cost of conducting a phantom scan. We also intend to propose an optimal phantom scanning design for distortion correction using our developed method associated with scanner performance monitoring.

ACKNOWLEDGMENTS

This study is part of the "Translational Research Promotion Project/Research project for the Development of a Systematic Method for the Assessment of Alzheimer's Disease," sponsored by the New Energy and Industrial Technology Development Organization (NEDO) of Japan. The J-ADNI is also supported by a Grant-in-Aid for Comprehensive Research on Dementia from the Japanese Ministry of Health, Labour, and Welfare, as well as by grants from J-ADNI Pharmaceutical Industry Scientific Advisory Board (ISAB) companies. The authors would like to thank the J-ADNI Imaging ISAB and other organizations for their support of this work.

[a]Author to whom correspondence should be addressed. Electronic mail: maikusa@ncnp.go.jp

[1]C. Fennema-Notestine, D. J. Hagler, Jr., L. K. McEvoy, A. S. Fleisher, E. H. Wu, D. S. Karow, A. M. Dale, and Alzheimer's Disease Neuroimaging Initiative, "Structural MRI biomarkers for preclinical and mild Alzheimer's disease," Hum. Brain. Mapp. **30**(10), 3238–3253 (2009).

[2]B. C. Dickerson, I. Goncharova, M. P. Sullivan, C. Forchetti, R. S. Wilson, D. A. Bennett, L. A. Beckett, and L. deToledo-Morrell, "MRI-derived entorhinal and hippocampal atrophy in incipient and very mild Alzheimer's disease," Neurobiol. Aging **22**, 747–754 (2001).

[3]D. Holland, J. B. Brewer, D. J. Hagler, C. Fennema-Notestine, A. M. Dale, and Alzheimer's Disease Neuroimaging Initiative, "Subregional neuroanatomical change as a biomarker for Alzheimer's disease," Proc. Natl. Acad. Sci. U.S.A. **106**(49), 20954–20959 (2009).

[4]M. Grundman, D. Sencakova, C. R. Jack, Jr., R. C. Petersen, H. T. Kim, A. Schultz, M. F. Weiner, C. DeCarli, S. T. DeKosky, C. Dyck, R. G. Thomas, and L. J. Thal, "Alzheimer's disease cooperative study brain MRI hippocampal volume and prediction of clinical status in a mild cognitive impairment trial," J. Mol. Neurosci. **19**, 23–27 (2002).

[5]C. R. Jack, Jr., R. C. Petersen, Y. C. Xu, S. C. Waring, P. C. O'Brien, E. G. Tangalos, G. E. Smith, R. J. Ivnik, and E. Kokmen, "Medial temporal atrophy on MRI in normal aging and very mild Alzheimer's disease," Neurology. **49**, 786–794 (1997).

[6]S. Klöppel, C. M. Stonnington, C. Chu, B. Draganski, R. I. Scahill, J. D. Rohrer, N. C. Fox, C. R. Jack, Jr., J. Ashburner, and R. S. J. Frackowiak, "Automatic classification of MR scans in Alzheimer's disease," Brain **131**, 681–689 (2008).

[7]B. Magnin, L. Mesrob, S. Kinkingnéhun, M. Pélégrini-Issac, O. Colliot, M. Sarazin, B. Dubois, S. Lehéricy, and H. Benali, "Support vector machine-based classification of Alzheimer's disease from whole-brain anatomical MRI," Neuroradiology **51**, 73–83 (2009).

[8]H. Matsuda, S. Mizumura, K. Nemoto, F. Yamashita, E. Imabayashi, N. Sato, and T. Asada, "Automatic voxel-based morphometry of structural MRI by SPM8 plus diffeomorphic anatomic registration through exponentiated Lie algebra improves the diagnosis of probable Alzheimer disease," AJNR Am. J. Neuroradiol. **33**, 1109–1114 (2012).

[9]P. Vemuri, J. L. Gunter, M. L. Senjem, J. L. Whitwell, K. Kantarci, D. S. Knopman, B. F. Boeve, R. C. Petersen, and C. R. Jack, Jr., "Alzheimer's disease diagnosis in individual subjects using structural MR images: Validation studies," Neuroimage **39**, 1186–1197 (2008).

[10]B. Dubois, H. H. Feldman, C. Jacova, S. T. Dekosky, P. Barberger-Gateau, J. Cummings, A. Delacourte, D. Galasko, S. Gauthier, G. Jicha, K. Meguro, J. O'brien, F. Pasquier, P. Robert, M. Rossor, S. Salloway, Y. Stern, P. J. Visser, and P. Scheltens, "Research criteria for the diagnosis of Alzheimer's disease: Revising the NINCDS-ADRDA criteria," Lancet Neurol. **6**(8), 734–746 (2007).

[11]R. Cuingnet, E. Gerardin, J. Tessieras, G. Auzias, S. Lehéricy, M. O. Habert, M. Chupin, H. Benali, and O. Colliot, "Automatic classification of patients with Alzheimer's disease from structural MRI: A comparison of ten methods using the ADNI database," Neuroimage **56**, 766–781 (2011).

[12]S. K. Madsen, A. J. Ho, X. Hua, P. S. Saharan, A. W. Toga, C. R. Jack, Jr., M. W. Weiner, and P. M. Thompson, "3D maps localize caudate nucleus atrophy in 400 Alzheimer's disease, mild cognitive impairment, and healthy elderly subjects," Neurobiol. Aging **31**, 1312–1325 (2010).

[13]C. R. Jack, Jr., M. A. Bernstein, N. C. Fox, P. Thompson, G. Alexander, D. Harvey, B. Borowski, P. J. Britson, J. L. Whitwell, C. Ward, A. M. Dale, J. P. Felmlee, J. L. Gunter, D. L. Hill, R. Killiany, N. Schuff, S. Fox-Bosetti, C. Lin, C. Studholme, C. S. DeCarli, G. Krueger, H. A. Ward, G. J. Metzger, K. T. Scott, R. Mallozzi, D. Blezek, J. Levy, J. P. Debbins, A. S. Fleisher, M. Albert, R. Green, G. Bartzokis, G. Glover, J. Mugler, and M. W. Weiner, "The Alzheimer's Disease Neuroimaging Initiative (ADNI): MRI methods," J. Magn. Reson. Imaging **27**(4), 685–691 (2008).

[14]T. Iwatsubo, "Japanese Alzheimer's Disease Neuroimaging Initiative: Present status and future," Alzheimers Dement. **6**, 297–299 (2010).

[15]A. Janke, H. Zhao, G. J. Gowin, G. J. Galloway, and D. M. Doddrell, "Use of spherical harmonic deconvolution methods to compensate for nonlinear gradient effects on MRI images," Magnet. Reson. Med. **52**, 115–122 (2004).

[16]Z. Caramanos, V. S. Fonov, S. J. Francis, S. Narayanan, G. B. Pike, D. L. Collins, and D. L. Arnold, "Gradient distortions in MRI: Characterizing and correcting for their effects on SENA-generated measures of brain volume change," Neuroimage **49**, 1601–1611 (2010).

[17]L. N. Baldwin, K. Wachowicz, and S. D. Thomas, "Characterization, prediction and correction of geometric distortion in 3 T MRI images," Med. Phys. **34**, 388–399 (2007).

[18]A. D. Leow, A. D. Klunder, C. R. Jack, Jr., A. W. Toga, A. M. Dale, M. A. Bernstein, P. J. Britson, J. L. Gunter, C. P. Ward, J. L. Whitwell, B. J. Borowski, A. S. Fleisher, N. C. Fox, D. Harvey, J. Kornak, N. Schuff, C. Studholme, G. E. Alexander, M. W. Weiner, P. M. Thompson, and ADNI Preparatory Phase Study, "Longitudinal stability of MRI for mapping brain change using tensor-based morphometry," Neuroimage **31**, 627–640 (2006).

[19]J. L. Gunter, M. A. Bernstein, B. J. Borowski, C. P. Ward, P. J. Britson, and J. P. Felmlee, "Measurement of MRI scanner performance with the ADNI phantom," Med. Phys. **36**, 2193–2205 (2009).

[20]L. Schad, S. Lott, F. Schmitt, V. Sturm, and W. J. Lorenz, "Correction of spatial distortion in MRI image: A prerequisite for accurate stereotaxy," J. Comput. Assist. Tomogr. **11**, 499–505 (1987).

[21]S. Langlois, M. Desvigness, J. M. Constans, and M. Revenu, "MRI geometric distortion: A simple approach to correcting the effects of non-linear gradient fields," J. Comput. Assist. Tomogr. **9**, 821–831 (1999).

[22]C. Menuel, L. Garnero, E. Bardinet, F. Poupon, D. Phalippou, and D. Dormont, "Characterization and correction of distortions in stereotactic magnetic resonance imaging for bilateral subthalamic stimulation in Parkinson disease," J. Neurosurg. **103**, 256–266 (2005).

[23]A. R. Liddle, "Information criteria for astrophysical model selection," Mon. Not. Roy. Astron. Soc. Lett. **377**, L74-L78 (2007).

[24]G. Schwarz, "Estimating the dimension of a model," Ann. Stat. **6**, 461–464 (1978).

[25]J. Jovicich, S. Czanner, D. Greve, E. Haley, A. van der Kouwe, R. Gollub, D. Kennedy, F. Schmitt, G. Brown, J. MacFall, B. Fischl, and A. Dale, "Reliability in multi-site structural MRI studies: Effects of gradient nonlinearity correction on phantom and human data," Neuroimage **30**, 436–443 (2006).

[26]C. J. C. Bakker, M. A. Moerland, R. Bhagwandien, and R. Beersma, "Analysis of machine-dependent and object-induced geometric distortion if 2DFT MR imaging," Magn. Reson. Imaging **10**, 597–608 (1992).

[27]M. J. Clarkson, S. Ourselin, C. Nielsen, K. K. Leung, J. Barnes, J. L. Whitwell, J. L. Gunter, D. L. Hill, M. W. Weiner, C. R. Jack, Jr., N. C. Fox, and Alzheimer's Disease Neuroimaging Initiative, "Comparison of phantom and registration scaling corrections using the ADNI cohort," Neuroimage **47**, 1506–1513 (2009).

ORIGINAL ARTICLE

A Phase 3 Trial of Semagacestat for Treatment of Alzheimer's Disease

Rachelle S. Doody, M.D., Ph.D., Rema Raman, Ph.D., Martin Farlow, M.D., Takeshi Iwatsubo, M.D., Ph.D., Bruno Vellas, M.D., Steven Joffe, M.D., M.P.H., Karl Kieburtz, M.D., M.P.H., Feng He, M.S., Xiaoying Sun, M.S., Ronald G. Thomas, Ph.D., and Paul S. Aisen, M.D., for the Alzheimer's Disease Cooperative Study Steering Committee; and Eric Siemers, M.D., Gopalan Sethuraman, Ph.D., and Richard Mohs, Ph.D., for the Semagacestat Study Group

ABSTRACT

BACKGROUND

Alzheimer's disease is characterized by the presence of cortical amyloid-beta (Aβ) protein plaques, which result from the sequential action of β-secretase and γ-secretase on amyloid precursor protein. Semagacestat is a small-molecule γ-secretase inhibitor that was developed as a potential treatment for Alzheimer's disease.

METHODS

We conducted a double-blind, placebo-controlled trial in which 1537 patients with probable Alzheimer's disease underwent randomization to receive 100 mg of semagacestat, 140 mg of semagacestat, or placebo daily. Changes in cognition from baseline to week 76 were assessed with the use of the cognitive subscale of the Alzheimer's Disease Assessment Scale for cognition (ADAS-cog), on which scores range from 0 to 70 and higher scores indicate greater cognitive impairment, and changes in functioning were assessed with the Alzheimer's Disease Cooperative Study–Activities of Daily Living (ADCS-ADL) scale, on which scores range from 0 to 78 and higher scores indicate better functioning. A mixed-model repeated-measures analysis was used.

RESULTS

The trial was terminated before completion on the basis of a recommendation by the data and safety monitoring board. At termination, there were 189 patients in the group receiving placebo, 153 patients in the group receiving 100 mg of semagacestat, and 121 patients in the group receiving 140 mg of semagacestat. The ADAS-cog scores worsened in all three groups (mean change, 6.4 points in the placebo group, 7.5 points in the group receiving 100 mg of the study drug, and 7.8 points in the group receiving 140 mg; $P=0.15$ and $P=0.07$, respectively, for the comparison with placebo). The ADCS-ADL scores also worsened in all groups (mean change at week 76, –9.0 points in the placebo group, –10.5 points in the 100-mg group, and –12.6 points in the 140-mg group; $P=0.14$ and $P<0.001$, respectively, for the comparison with placebo). Patients treated with semagacestat lost more weight and had more skin cancers and infections, treatment discontinuations due to adverse events, and serious adverse events ($P<0.001$ for all comparisons with placebo). Laboratory abnormalities included reduced levels of lymphocytes, T cells, immunoglobulins, albumin, total protein, and uric acid and elevated levels of eosinophils, monocytes, and cholesterol; the urine pH was also elevated.

CONCLUSIONS

As compared with placebo, semagacestat did not improve cognitive status, and patients receiving the higher dose had significant worsening of functional ability. Semagacestat was associated with more adverse events, including skin cancers and infections. (Funded by Eli Lilly; ClinicalTrials.gov number, NCT00594568.)

From the Alzheimer's Disease and Memory Disorders Center, Department of Neurology, Baylor College of Medicine, Houston (R.S.D.); the Division of Biostatistics and Bioinformatics, Department of Family and Preventive Medicine (R.R., F.H., X.S., R.G.T.), and the Department of Neurosciences (R.R., R.G.T., P.S.A.), University of California at San Diego, San Diego; the Department of Neuropathology and Neuroscience, University of Tokyo, Tokyo (T.I.); the Clinic of Internal Medicine and Gerontology, Hôpital La Grave-Casselardit, Toulouse, France (B.V.); the Department of Hematology–Oncology, Boston Children's Hospital, Boston (S.J.); the Center for Human Experimental Therapeutics, University of Rochester Medical Center, Rochester, NY (K.K.); and Eli Lilly and the Semagacestat Study Group (E.S., G.S., R.M.) and the Indiana Alzheimer Disease Center, Indiana University (M.F.) — all in Indianapolis. Address reprint requests to Dr. Doody at Baylor College of Medicine, Department of Neurology, Alzheimer's Disease and Memory Disorders Center, 1977 Butler Blvd., Suite E 5.101, Houston, TX 77030, or at rdoody@bcm.edu.

N Engl J Med 2013;369:341-50.
DOI: 10.1056/NEJMoa1210951

Alzheimer's disease begins decades before the appearance of clinical symptoms, with the deposition of aggregated amyloid-beta (Aβ) peptide plaques in the cortex and hippocampus. This protein is cleaved from the amyloid precursor protein (APP) by the sequential action of β- and γ-secretases, producing fragments that include Aβ1-40 and Aβ1-42. Since the accumulation of aggregated Aβ is associated with disease progression, both β-secretase and γ-secretase represent potential therapeutic targets. Multiple small molecules can inhibit γ-secretase in vitro,[1-4] but Notch and other transmembrane proteins are also substrates for γ-secretase,[1-4] and studies have raised concern that the inhibition of γ-secretase could interfere with Notch receptor–related nuclear signaling and with the function of cell-surface receptors and proteins involved in embryonic development, hematopoiesis, cell adhesion, and other cell-to-cell contacts. Eli Lilly developed semagacestat as a potential therapy for Alzheimer's disease because studies had shown inhibition of γ-secretase in vitro, reduction of Aβ synthesis by approximately 25% over a period of 24 hours (at a dose of 100, 140, or 280 mg) in humans (as measured with the use of an Aβ synthesis assay), and reduction of Aβ levels in the brains, cerebrospinal fluid (CSF), and plasma of PDAPP transgenic mice carrying human *APP* with a V717F mutation.[5-8] Semagacestat functionally inhibits γ-secretase rather than competitively inhibiting the enzyme active sites.[9]

Phase 1 studies in human volunteers without Alzheimer's disease showed a dose-dependent decrease in plasma levels of Aβ after 14 days of treatment with 5 to 50 mg of semagacestat[10]; patients with mild-to-moderate Alzheimer's disease had a significant reduction in plasma levels of Aβ but not in CSF levels of Aβ after 6 weeks of treatment with a maximum daily dose of 40 mg of semagacestat.[7] A small phase 2 safety study testing doses of 100 or 140 mg of semagacestat daily for up to 14 weeks as compared with placebo in patients with mild-to-moderate Alzheimer's disease showed dose-related reductions in plasma levels of Aβ40 but no significant reductions in CSF levels (measured after a lumbar puncture performed 6 hours after administration of the last dose).[11] Adverse events affecting the skin and subcutaneous tissue were more common among patients receiving semagacestat than among those receiving placebo (P=0.05); semagacestat was also associated with reduced uric acid levels (P<0.001), prolongation of the prothrombin time and partial-thromboplastin time (P<0.05), and reduced percentages of CD19 cells (P<0.05).

On the basis of the possibility that semagacestat could have disease-modifying effects in patients with Alzheimer's disease, Eli Lilly launched a multinational phase 3 trial comparing the two doses used in phase 2 with placebo in a randomized, double-blind trial. This trial was halted by the company after the results of a preplanned interim analysis performed by an independent data and safety monitoring board (DSMB) suggested that semagacestat was associated with worsening of cognition and functioning as well as excess skin cancers and a futility analysis indicated that the trial could not show a benefit with respect to the primary outcomes.

METHODS

SCREENING

Otherwise healthy patients 55 years of age or older who had mild-to-moderate Alzheimer's disease without depression were randomly assigned to receive 100 mg or 140 mg of semagacestat or placebo daily for 76 weeks. Depression at screening was assessed with the Geriatric Depression Scale (score range, 0 to 15, with higher scores indicating more severe depression)[12]; a score of 6 or less was considered to indicate the absence of depression. Mild-to-moderate Alzheimer's disease was documented on the basis of a score of 16 to 26 on the Mini–Mental State Examination (MMSE; score range, 0 to 30, with higher scores indicating better cognitive function)[13] and the criteria of the National Institute of Neurological and Communicative Diseases and Stroke–Alzheimer's Disease and Related Disorders Association.[14] Patients being treated with background cholinesterase inhibitors, memantine, or both were admitted to the study.

OVERSIGHT

The research protocol was approved by the institutional review board at each institution where the trial was conducted, and all participants provided written informed consent. The study protocol and details on informed consent are available with the full text of this article at NEJM.org. All the authors vouch for the completeness and veracity of the data and the data analysis and for

the fidelity of this report to the protocol. Eli Lilly or its contract research organization collected the data, and the Alzheimer's Disease Cooperative Study (ADCS) performed the analysis and wrote the report, which was approved by the voting members of the Data Analysis and Publication Committee (a collaborative body of the ADCS and Eli Lilly, with the latter not having voting rights) and the ADCS steering committee (for a list of members of both committees, see the Supplementary Appendix, available at NEJM.org).

SAFETY ASSESSMENTS

Monitoring for adverse events was performed at each visit. Additional safety assessments included measurement of lymphocytes, examination of the skin, urinalysis, and electrocardiography (ECG). Because semagacestat strongly activates the production of cytochrome P-450 3A (CYP3A), levels of representative drugs likely to be affected by CYP3A were assessed periodically, and donepezil levels were measured at baseline, week 12, and week 52.

OUTCOME MEASURES

Efficacy measures included the cognitive subscale of the Alzheimer's Disease Assessment Scale (ADAS-cog; score range, 0 to 70, with higher scores indicating greater cognitive impairment),[15] the Alzheimer's Disease Cooperative Study–Activities of Daily Living (ADCS-ADL) scale (score range, 0 to 78, with higher scores indicating better function),[16] the Clinical Dementia Rating–Sum of Boxes (CDR-SB),[17,18] the Neuropsychiatric Inventory (NPI),[19] the Resource Utilization in Dementia Lite (RUD-Lite) scale,[20] the European Quality of Life–5 Dimensions (EQ-5D) scale (proxy version),[21] and the MMSE.[13]

Apolipoprotein E genotyping was performed at study entry, and lymphocyte phenotyping was performed periodically. Plasma levels of Aβ were assessed throughout the study, and the following studies were performed for a subset of patients at baseline and at week 76 or at the time of early termination: CSF analysis for levels of Aβ and tau protein, volumetric magnetic resonance imaging (MRI), positron-emission tomographic (PET) imaging with the use of ^{18}F-florbetapir for the visualization of Aβ. The analyses of plasma and CSF and the performance of ^{18}F-fluorodeoxyglucose (FDG) PET and other imaging studies were carried out to obtain evidence of disease modification and pharmacodynamic effect.

STATISTICAL ANALYSIS

The primary analysis compared the change in scores on the ADAS-cog and the ADCS-ADL scale from baseline to week 76 in the three study groups. The DSMB charter called for an interim analysis of the coprimary outcomes after 50% of patients had completed 12 months of treatment or had dropped out of the study. DSMB members were initially unaware of group assignments but were to be informed of them if the P value for the differences in coprimary outcomes between any two groups was less than 0.05. When it was determined that the outcomes for the two treatment groups were worse than they were for the control group, the DSMB decided that the futility analysis, originally scheduled to take place after 50% of participants had been in the study for 18 months, should be performed. The DSMB recommended stopping the trial on the basis of the results of the futility analysis.

We used a mixed-model repeated-measures analysis to compare model-adjusted least-squares means at 76 weeks. The model included the fixed effects of baseline scores on the ADAS-cog, ADCS-ADL scale, CDR-SB, MMSE, RUD-Lite scale, NPI, and EQ-5D scale; group assignment (high-dose semagacestat, low-dose semagacestat, or placebo); MMSE score at screening (continuous variable); visit and treatment-by-visit interaction; concomitant treatment with a cholinesterase inhibitor or memantine at baseline (yes or no); and age at baseline. The random effect (random intercept) was "participant." Study visit was treated as a categorical variable. The primary efficacy analysis was based on a modified intention-to-treat population, which included all randomly assigned participants for whom there was at least one postbaseline observation. A secondary analysis was performed for all participants who completed the initial 76-week period of treatment with the study medication.

The coprimary outcomes were the change in scores on the ADAS-cog and the ADCS-ADL scale from baseline to study end point (week 76). Secondary outcomes were the change from baseline in scores on the CDR–SB, MMSE, NPI, RUD-Lite scale, and EQ-5D scale; the values for plasma and CSF levels of Aβ and for CSF levels of tau; MRI volumetric measures; and imaging studies performed with FDG-PET and ^{18}F-florbetapir–PET.

Safety analyses were based on the full intention-to-treat population, and all biomarker analy-

ses were calculated with the use of data from patients with at least one postbaseline value. The baseline characteristics of the three study groups were compared with the use of a Fisher–Freeman–Halton test for categorical variables and an analysis of variance for continuous variables. An analysis of covariance was used to assess changes from baseline in plasma and CSF biomarkers and in imaging studies. Safety analyses included summary listings of adverse events, with Fisher's exact test used for pairwise comparisons. R Statistical software, version 2.14.1 (R Development Core Team 2011), was used for all statistical analyses. For the coprimary outcomes, the analysis was duplicated with the use of SAS software, version 9 (SAS Institute). The significance level was set at $P<0.05$. All statistical testing was two-sided. Table S1 in the Supplementary Appendix shows the slight differences between the ADCS analysis plan and the original Eli Lilly analysis plan.

RESULTS

COMPLETION RATES AND BASELINE CHARACTERISTICS

A total of 2009 patients were screened for study participation and 1537 underwent randomization (Fig. 1). The most common reason for not proceeding to randomization was failure to meet the entry criteria (accounting for 77% of screened patients who did not meet the entry criteria). Completion rates were 38% for participants in the placebo group, 30% for those receiving 100 mg of semagacestat, and 23% for those receiving 140 mg of semagacestat; most discontinuations were due to the termination of the trial. Discontinuation due to adverse events was higher in the groups receiving semagacestat ($P<0.001$).

There were no significant differences among the groups with regard to baseline characteristics (Table 1). Approximately 82% of participants were being treated with a cholinesterase inhibitor, 33% with memantine, and 29% with both.

COGNITIVE AND CLINICAL OUTCOMES

Coprimary Outcomes

Table 2 shows modeled changes from baseline for the coprimary and secondary outcomes at 76 weeks in the modified intention-to-treat population (also see Fig. S2 through S7 in the Supplementary Appendix). Cognitive status worsened in all three groups (mean change in the ADAS-cog score at week 76, 6.4 points in the placebo group, 7.5 points in the group receiving 100 mg of semagacestat, and 7.8 points in the group receiving 140 mg of semagacestat; $P=0.15$ for the comparison of 100 mg with placebo, and $P=0.07$ for the comparison of 140 mg with placebo). Functional status also worsened in all three groups (mean change in the ADCS-ADL score at week 76, −9.0 points with placebo, −10.5 points with 100 mg of semagacestat, and −12.6 points with 140 mg of semagacestat; $P=0.14$ for the comparison of 100 mg with placebo, and $P<0.001$ for the comparison of 140 mg with placebo). An analysis of the 679 patients who completed 76 weeks of treatment showed no significant differences in scores on either the ADAS-cog or the ADCS-ADL scale among the three groups.

Secondary Outcomes

Scores on the CDR-SB showed worsening severity of dementia in all three study groups, with greater deterioration in the two semagacestat groups than in the placebo group (Fig. S3 in the Supplementary Appendix). MMSE and NPI scores showed significant worsening in the group receiving 140 mg of semagacestat as compared with the placebo group in the modified intention-to-treat population (Fig. S4 and S5 in the Supplementary Appendix). These differences were not observed in an analysis restricted to participants who completed the 76-week period of treatment. There were no significant differences among groups with respect to the RUD-Lite score (Table S7 in the Supplementary Appendix). As compared with the placebo group, the group receiving 140 mg of semagacestat had greater worsening on the EQ-5D scale (U.S. and U.K. index scores) at weeks 52 and 76, and the group receiving 100 mg of semagacestat had greater worsening at weeks 28, 52, and 76 (Table S7 in the Supplementary Appendix).

SAFETY AND SIDE-EFFECT PROFILE

Weight

By week 52, patients in both active-treatment groups had lost weight (mean [±SD] change, −1.6±4.4 kg in the group receiving 100 mg of semagacestat and −1.6±4.7 kg in the group receiving 140 mg), whereas those in the placebo group had gained an average of 0.4±3.9 kg by week 52 ($P<0.001$ for both comparisons). The percentage

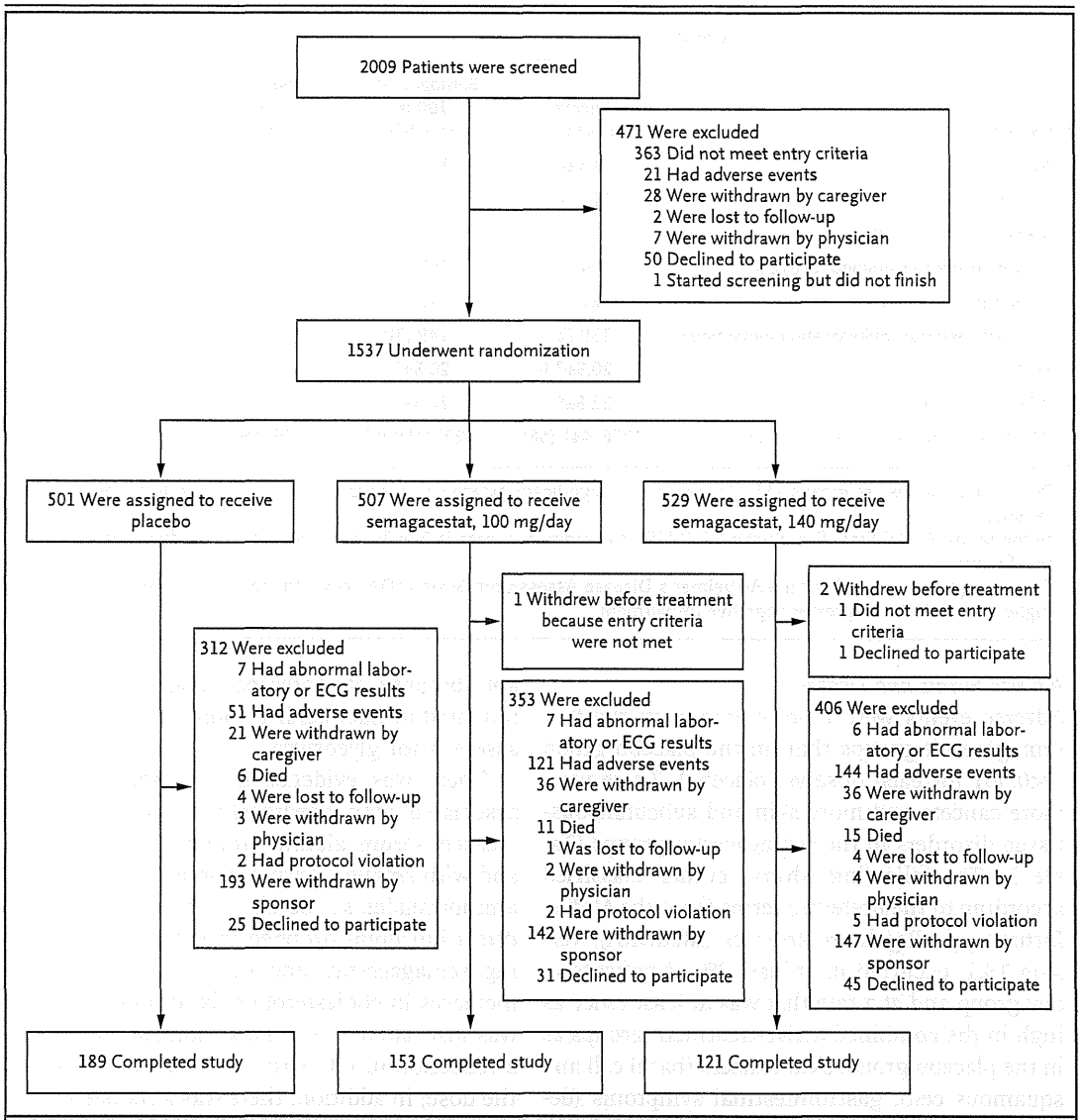

Figure 1. Screening, Randomization, and Follow-up during the Initial 76-Week Treatment Period.
Patients were withdrawn by the study sponsor because of adverse effects of the study drug and futility (according to the futility analysis) reported by the data and safety monitoring board. ECG denotes electrocardiographic.

of patients who lost more than 7% of their baseline weight by week 52 was 7% in the placebo group, as compared with 19% in the lower-dose semagacestat group and 20% in the higher-dose semagacestat group (P<0.001 for both comparisons). The percentage of patients who lost more than 4% of their baseline weight at week 52 was 15% in the placebo group, as compared with 31% in the lower-dose semagacestat group and 33% in the higher-dose semagacestat group (P<0.001 for both comparisons).[22]

ECG Results

ECGs showed a small increase from baseline in the QT interval corrected for heart rate at week 72 in the two treatment groups (mean increase, 3.5±15.3 msec in the lower-dose semagacestat group and 3.7±15.4 msec in the higher-dose semagacestat group, vs. a decrease of 0.44±13.7 msec in the placebo group). The increase was significantly larger in the two active-treatment groups than in the placebo group (P<0.001 for both comparisons).

Table 1. Baseline Characteristics of the Study Participants.*

Variable	Placebo (N=501)	Semagacestat, 100 mg (N=506)	Semagacestat, 140 mg (N=527)	Total (N=1534)
Age — yr	73.3±8.1	73.0±8.0	73.3±8.5	73.2±8.2
Male sex — no. (%)	223 (45)	227 (45)	264 (50)	714 (47)
Medication — no. (%)				
Cholinesterase inhibitor use only	274 (55)	263 (52)	272 (52)	809 (53)
Memantine use only	21 (4)	21 (4)	16 (3)	58 (4)
Cholinesterase inhibitor and memantine use	139 (28)	149 (30)	101 (31)	449 (29)
MMSE score†	20.9±3.6	20.8±3.5	20.7±3.5	20.8±3.5
ADAS-cog score‡	22.8±9.1	22.5±8.9	23.1±8.8	22.9±9.0
APOE ε4 carrier — no./total. no. (%)	256/441 (58)	291/459 (63)	268/463 (58)	815/1363 (60)

* Plus–minus values are means ±SD. There were no significant differences among groups with regard to baseline characteristics.

† In the Mini–Mental State Examination (MMSE), the range of scores is 0 to 30, and higher scores indicate better cognitive function.

‡ On the cognitive subscale of the Alzheimer's Disease Assessment Scale (ADAS-cog), the range of scores is 0 to 70, and higher scores indicate greater cognitive impairment.

Adverse Events and Deaths

Adverse events were more common in the two semagacestat groups than in the placebo group (P<0.001 for each dose vs. placebo). There were more cancers and more skin and subcutaneous-tissue disorders in the semagacestat groups (Table 3). The following adverse events, classified according to the preferred terms from the *Medical Dictionary for Regulatory Activities* (MedDRA), version 13.1, occurred in at least 2% of patients in any group and at a rate that was at least twice as high in the combined active-treatment groups as in the placebo group: skin cancers (basal cell and squamous cell), gastrointestinal symptoms (decreased appetite, nausea, vomiting, and weight loss), hair and eyelash hypopigmentation, epistaxis, rashes and pruritus, alopecia, and syncope (Table 3). The percentage of patients who discontinued the study drug because of adverse events was higher with semagacestat than with placebo (26% with 100 mg and 30% with 140 mg vs. 11% with placebo, P<0.001 for both comparisons). There were more cancers (especially nonmelanoma skin cancer) and more infections in the two groups receiving semagacestat (Table S4 in the Supplementary Appendix).

Several laboratory findings pointed to drug-related and clinically significant effects of semagacestat (Table S5 in the Supplementary Appendix). There was indirect evidence of Fanconi's syndrome with a dose-related reduction in serum uric acid and phosphorus levels (presumably because of increased renal excretion) and impaired urinary acidification. Patients were not assessed for glycosuria.

There was evidence that semagacestat was associated with hepatocellular injury, with increased serum alanine aminotransferase levels and with smaller elevations in levels of aspartate aminotransferase. Levels of bilirubin (primarily direct bilirubin) dropped in both groups receiving semagacestat, and there were dose-related increases in cholesterol levels. Immune function was also altered. The most notable change was a reduction in IgG levels that was not related to the dose; in addition, there was a transient elevation in IgM levels that was unrelated to the dose and a transient dose-related reduction in IgA levels. Lymphocyte counts, particularly CD19 cells, declined (a change that was not related to the dose), and there were small, dose-dependent increases in monocyte and eosinophil counts.

The rate of serious adverse events was higher in the semagacestat groups than in the placebo group: 24% in the group receiving 100 mg and 25% in the group receiving 140 mg, as compared with 14% in the group receiving placebo (P<0.001 for both comparisons) (Table S6 in the Supplementary Appendix). More deaths occurred among patients receiving semagacestat (9 in the group receiving 100 mg and 14 in the group receiving 140 mg vs. 6 in the group receiving placebo, P=0.25 for both comparisons) (Table S8 in the Supplementary Appendix).