

# Predictive Utility of Marketed Volumetric Software Tools in Subjects at Risk for Alzheimer Disease: Do Regions Outside the Hippocampus Matter?

 T.P. Tanpitukpongse,  M.A. Mazurowski,  J. Ikhen, and  J.R. Petrella, for the Alzheimer's Disease Neuroimaging Initiative



## ABSTRACT

**BACKGROUND AND PURPOSE:** Alzheimer disease is a prevalent neurodegenerative disease. Computer assessment of brain atrophy patterns can help predict conversion to Alzheimer disease. Our aim was to assess the prognostic efficacy of individual-versus-combined regional volumetrics in 2 commercially available brain volumetric software packages for predicting conversion of patients with mild cognitive impairment to Alzheimer disease.

**MATERIALS AND METHODS:** Data were obtained through the Alzheimer's Disease Neuroimaging Initiative. One hundred ninety-two subjects (mean age, 74.8 years; 39% female) diagnosed with mild cognitive impairment at baseline were studied. All had T1-weighted MR imaging sequences at baseline and 3-year clinical follow-up. Analysis was performed with NeuroQuant and Neuroreader. Receiver operating characteristic curves assessing the prognostic efficacy of each software package were generated by using a univariable approach using individual regional brain volumes and 2 multivariable approaches (multiple regression and random forest), combining multiple volumes.

**RESULTS:** On univariable analysis of 11 NeuroQuant and 11 Neuroreader regional volumes, hippocampal volume had the highest area under the curve for both software packages (0.69, NeuroQuant; 0.68, Neuroreader) and was not significantly different ( $P > .05$ ) between packages. Multivariable analysis did not increase the area under the curve for either package (0.63, logistic regression; 0.60, random forest NeuroQuant; 0.65, logistic regression; 0.62, random forest Neuroreader).

**CONCLUSIONS:** Of the multiple regional volume measures available in FDA-cleared brain volumetric software packages, hippocampal volume remains the best single predictor of conversion of mild cognitive impairment to Alzheimer disease at 3-year follow-up. Combining volumetrics did not add additional prognostic efficacy. Therefore, future prognostic studies in mild cognitive impairment, combining such tools with demographic and other biomarker measures, are justified in using hippocampal volume as the only volumetric biomarker.

**ABBREVIATIONS:** AD = Alzheimer disease; ADAS-13 = 13-item Alzheimer's Disease Assessment Scale; ADNI = Alzheimer's Disease Neuroimaging Initiative; ADNI-GO = Alzheimer's Disease Neuroimaging Initiative—Grand Opportunities; AUC = area under the curve; HOC = Hippocampal Occupancy Score; MCI = mild cognitive impairment; NQ = NeuroQuant; NR = Neuroreader; ROC = receiver operating characteristic

**A**lzheimer disease (AD) is a progressive neurodegenerative disease leading to synaptic dysfunction, neuronal death, and brain atrophy. Atrophy of specific medial temporal lobe struc-

tures, such as the hippocampus, amygdala, and parahippocampal gyrus, has been associated with the future development of AD in numerous research studies.<sup>1-9</sup> However, use of this information

Received April 21, 2016; accepted after revision October 31.


From the Department of Radiology (T.P.T., M.A.M., J.R.P.), Duke University Medical Center, Durham, North Carolina; and Duke University School of Medicine (J.I.), Durham, North Carolina.

Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative data base (adni.loni.usc.edu). Thus, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at [http://adni.loni.usc.edu/wp-content/uploads/how\\_to\\_apply/ADNI\\_Acknowledgement\\_List.pdf](http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf).

Data collection and sharing for this project were funded by the ADNI (National Institutes of Health Grant U01 AG024904) and DOD-ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, the Alzheimer's Association, Alzheimer's Drug Discovery Foundation, Araclon Biotech, BioClinica, Biogen, Bristol-Myers Squibb, CereSpir Inc, Eisai Inc, Elan Pharmaceuticals, Eli Lilly and

Company, EUROIMMUN, F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Fujirebio Diagnostics, GE Healthcare, IXICO, Janssen Alzheimer Immunotherapy Research & Development, Johnson & Johnson Pharmaceutical Research & Development LLC, Lumosity, Lundbeck, Merck & Co, Meso Scale Diagnostics LLC, NeuroRx Research, NeuroTrack Technologies, Novartis Pharmaceuticals Corporation, Pfizer, Piramal Imaging, Servier, Takeda Pharmaceutical Company, and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health ([www.fnih.org](http://www.fnih.org)). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Disease Cooperative Study at the University of California, San Diego. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

Please address correspondence to Jeffrey R. Petrella, MD, Duke University Hospital, Department of Radiology, Box 3808, Durham, NC 27710; e-mail: [jeffrey.petrella@duke.edu](mailto:jeffrey.petrella@duke.edu)

 Indicates open access to non-subscribers at [www.ajnr.org](http://www.ajnr.org)

<http://dx.doi.org/10.3174/ajnr.A5061>

in the clinical care of patients with memory impairment for prognosis remains challenging.<sup>10,11</sup> Medial temporal lobe volume assessments of MR images with visual ratings or manual or semi-manual volumetric processing have been difficult to implement in a busy clinical environment due to the high interobserver variability of raters and/or the time-consuming nature of obtaining these measurements.<sup>12</sup> These problems have been addressed through the use of fully automated segmentation algorithms available in commercial software programs, providing the user with immediate, detailed volumetric analysis of the hippocampus and other brain regions, which is more sensitive than visual analysis.<sup>13</sup>

Because the atrophy pattern in prodromal AD is spatially distributed, including regions beyond the hippocampus, such as the lateral and inferior temporal lobe, parietal lobe, and cingulate gyrus,<sup>14</sup> incorporation of such information may enhance the prognostic capability of these currently available tools. Indeed, pattern-analysis techniques, incorporating whole-brain morphologic information,<sup>15-19</sup> have been harnessed for this purpose and have shown a high prognostic value in individual patients. Such techniques, however, have not yet been implemented in commercially available products. The purpose of this study was to assess the prognostic efficacy of using the complete set of raw volumetric measures available in 2 fully automated, commercially available brain volumetric software packages. Such tools have been FDA 510(k) cleared for clinical use but have not yet been validated for specific diagnostic or prognostic purposes in AD. We sought to determine whether combining volumetrics by using multivariable approaches, including machine learning,<sup>20-22</sup> would add to the prognostic efficacy of individual measures alone, such as hippocampal volume, in predicting conversion from mild cognitive impairment (MCI) to AD. We hypothesized that the multivariable approach would enhance the prognostic value of already existing individual measures available in commercial volumetric software.

## MATERIALS AND METHODS

### Subjects

All subject data were available through the Alzheimer's Disease Neuroimaging Initiative (ADNI), a multicenter trial with a publicly available data base ([adni.loni.usc.edu](http://adni.loni.usc.edu)).<sup>15</sup> The ADNI was launched in 2003 as a public-private partnership, led by Principal Investigator Michael W. Weiner, MD. The primary goal of the ADNI has been to test whether MR imaging, PET, other biologic markers, and clinical and neuropsychological assessment can be combined to measure the progression of MCI and early AD. For up-to-date information, see [www.adni-info.org](http://www.adni-info.org).

All subjects classified as having MCI under ADNI-1 or late MCI under Alzheimer's Disease Neuroimaging Initiative-Grand Opportunities (ADNI-GO), with baseline MR imaging and a baseline and 3-year clinical assessment available on or before November 11, 2013, were selected.

Conversion of MCI to AD was based on the National Institute of Neurological Disorders and Stroke and the Alzheimer's Disease and Related Disorders Association criteria for probable or possible AD, determined by review from a committee according to protocol guidelines. Those subjects who did not fulfill these crite-

ria for conversion were defined as nonconverters, including subjects with MCI and those who reverted to a normal state.

### Image Acquisition and Image Analysis

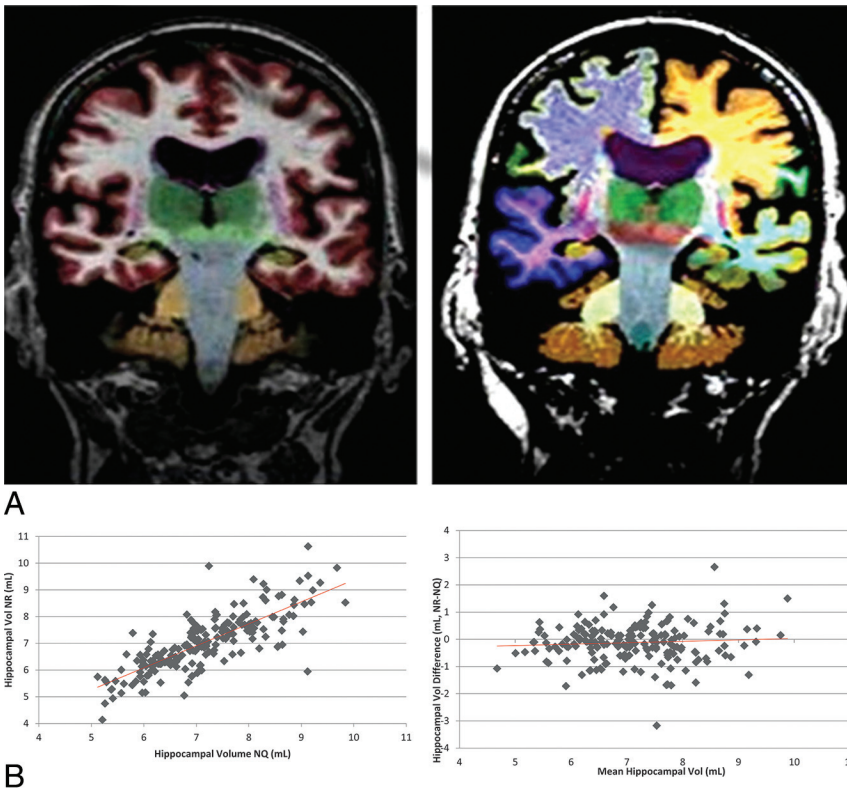
All subjects underwent brain MR imaging at 1.5T or 3T with sagittal 3D T1-weighted (MPRAGE) scans. Further details on the MR imaging scanner protocol and MR image acquisition are available elsewhere (<http://adni.loni.usc.edu/methods/documents/mri-protocols/>). Because we selected only subjects with late MCI from the ADNI-GO, all subjects were therefore previous members of ADNI-1 and had their original scan with the ADNI-1 protocol. The MPRAGE sequence was processed with NeuroQuant (NQ; CorTechs Labs, San Diego, California) (Original Version 1.0), which is a commercially available automated image-analysis software program (<http://www.cortechslabs.com/neuroquant/>). Separate left and right volumes, available in the generated morphometry report, were combined into a total volume for the following 11 brain regions (features): amygdala, caudate, cerebellum, cortical gray matter, forebrain parenchyma, hippocampus, inferior lateral ventricle, lateral ventricle, pallidum, putamen, and thalamus.

In addition, the MPRAGE sequence was also processed with Neuroreader (NR; Brainreader Aps, Horsens, Denmark), another commercially available software program for volumetric segmentation (<http://brainreader.net/>).<sup>23</sup> Both NQ and NR are FDA 510(k) cleared and "intended for automatic labeling, visualization and volumetric quantification of segmentable brain structures from a set of MR images" ([http://www.accessdata.fda.gov/cdrh\\_docs/pdf6/K061855.pdf](http://www.accessdata.fda.gov/cdrh_docs/pdf6/K061855.pdf) and [https://www.accessdata.fda.gov/cdrh\\_docs/pdf14/K140828.pdf](https://www.accessdata.fda.gov/cdrh_docs/pdf14/K140828.pdf)). Because the NR volumetric report yielded many more regions compared with NQ, 11 regions yielding the best individual predictions were chosen from the NR morphometry report to keep the models equally parsimonious. Combined bilateral volumes were obtained for the following 11 brain regions (features): amygdala, cerebellum, frontal lobe, hippocampus, lateral ventricle, occipital lobe, parietal lobe, putamen, temporal lobe, thalamus, and ventral diencephalon. Technical aspects of these commercial software packages have been previously described in detail.<sup>24,25</sup>

Of note, all ADNI studies included 2 echo-spoiled gradient echo/MPRAGE sequences performed back-to-back, the original and a repeat sequence accelerated by parallel imaging. In cases in which the original MPRAGE sequence could not be processed by the NR or NQ programs, the repeat MPRAGE sequence was used instead. If processing also failed on the repeat sequence, with either NR or NQ, the case was excluded from all further analysis. All image processing and analyses were performed by 2 authors (T.P.T. and J.I.). Sample segmentations from NQ and NR are shown in Fig 1A.

### Statistical Analysis

We compared demographics between the MCI converter and nonconverter groups by using a 2-sample Student *t* test, assuming equal variance when appropriate. To assess the prognostic performance of individual features and multivariable models combining the features, we used the receiver operating characteristic (ROC) curve methodology. Area under the curve (AUC) was used to summarize the performance. Comparison of AUCs was performed by using the method of Delong et al.<sup>26</sup>



**FIG 1.** Example of NeuroQuant (A, left) and Neuroreader (A, right) color segmentations in the same subject and hippocampal scatter (B, left) and Bland-Altman plots (B, right). Note the high correlation ( $r = 0.79$ ,  $P < .05$ ) of hippocampal volumetrics between software packages across all subjects and the small underestimation bias ( $-0.12$  mL,  $P < .05$ ) of NR with respect to NQ.

**Table 1: Characteristics of study populations<sup>a</sup>**

Characteristic	All Patients	MCI (Nonconverter)	MCI to AD (Converter)	P Value
No.	192	107	85	
Age (yr)	74.8 ± 7.3	74.7 ± 7.6	75.0 ± 6.9	.7604
Sex				.3262
Female	75 (39)	38 (36)	37 (44)	
Male	117 (61)	69 (64)	48 (56)	
Education (yr)	15.7 ± 2.9	15.7 ± 3.0	15.6 ± 2.8	.7637
ADAS-13 (baseline)	17.7 ± 6.4	15.2 ± 6.1	20.9 ± 5.3	<.0001
MMSE (baseline)	27.1 ± 1.7	27.6 ± 1.7	26.5 ± 1.6	<.0001

**Note:**—MMSE indicates Mini-Mental State Examination.

<sup>a</sup>MCI or AD status was determined at 3-year follow-up by the ADNI criteria. Data are means. Percentages listed in parentheses are rounded to the nearest percentage.

To test whether a combination of the features would outperform models based on a single brain region, we constructed 2 multivariable models: a classic multivariable logistic regression model and a more novel machine-learning method called the Random Forest Classifier.<sup>27</sup> For the Random Forest Classifier, 2000 trees were used. The leave-one-out cross-validation method was used for training and testing of the multivariable models. In addition to the data-driven multivariable models, we also tested one a priori multivariable model, known as the Hippocampal Occupancy (HOC) score, defined as the ratio of hippocampal volume to the sum of hippocampal and inferior lateral ventricle volumes. This measure is thought to differentiate individuals with congenitally small hippocampi from those with degeneration.<sup>28</sup> The HOC score for each separate hemisphere was averaged to provide a single composite measurement. Because inferior lateral

ventricle (temporal horn) volume was not available for NR, this measure was calculated only with NQ.

To assess correlations between assessment of the same volume by the NQ and NR software, we used the Pearson correlation. Statistical analyses were performed with Matlab, Version 8.1.0.604 R2013a (MathWorks, Natick, Massachusetts); R statistical and computing software, Version 3.1.3 (<http://www.r-project.org>); and JMP, Version 11.0 (SAS Institute, Cary, North Carolina) software packages.

## RESULTS

Initially 281 subjects were identified in the ADNI databases who met the inclusion criteria. Eighty-four (29.9%) subjects were excluded due to failure to generate an NQ morphometry report. Five (1.8%) subjects were excluded for other reasons, including failure to generate an NR morphometry report. A total of 192 remaining subjects met the inclusion and exclusion criteria. All subjects had a 3-year clinical follow-up visit recorded. Mean follow-up was  $3.05 \pm 0.14$  years (range, 2.47–3.63 years). All 192 included subjects were diagnosed as having MCI at baseline, and all started under the ADNI-1 protocol. At the end of the 3-year follow-up, 186/192 subjects had their 3-year evaluation under the ADNI-1 protocol and 6/192 subjects had their 3-year evaluation under the ADNI-GO protocol. For the 192 subjects, at the 3-year follow-up, the final diagnosis was AD in 85/192 (44.3%), MCI in 99/192 (51.6%), and healthy in 8/192 (4.2%).

Characteristics of the study populations are listed in Table 1. Two subjects did not have the 13-item Alzheimer's Disease Assessment Scale (ADAS-13) scores at baseline. There were no significant differences ( $P > .05$ ) between the MCI (nonconverter) and AD (converter) groups in terms of age, sex, or education. There was a significant difference ( $P < .0001$ ) in the ADAS-13 score at baseline and the Mini-Mental State Examination score at baseline between the 2 groups. The AUC for predicting conversion at baseline was 0.76 and 0.68 for the ADAS-13 and Mini-Mental State Examination, respectively.

AUC values for NQ and NR are listed in Table 2. With NQ, the most predictive feature for conversion of MCI to AD was hippocampal volume, with an AUC of 0.69. The most predictive feature for NR was also hippocampal volume, with an AUC of 0.68. The multivariable analysis did not improve on either of these.

**Table 2: AUC values for different brain regions separated by software package (NQ, NR) and method of analysis (univariable-multivariable)**

Feature	AUC NQ (95% CI)	AUC NR (95% CI)
Univariable analysis		
Hippocampus	0.69 (0.61–0.76)	0.68 (0.60–0.76)
Amygdala	0.67 (0.59–0.74)	0.65 (0.57–0.73)
Cerebellum	0.58 (0.49–0.66)	0.57 (0.49–0.66)
Putamen	0.58 (0.50–0.66)	0.62 (0.54–0.70)
Thalamus	0.56 (0.47–0.64)	0.56 (0.48–0.64)
Lateral ventricle	0.54 (0.46–0.62)	0.54 (0.46–0.62)
Pallidum	0.52 (0.43–0.60)	NA
Caudate	0.51 (0.42–0.59)	NA
Cortical gray matter	0.64 (0.56–0.72)	NA
Forebrain parenchyma	0.62 (0.54–0.70)	NA
Inferior lateral ventricle	0.60 (0.52–0.68)	NA
Temporal lobe	NA	0.63 (0.55–0.71)
Parietal lobe	NA	0.61 (0.53–0.69)
Frontal lobe	NA	0.60 (0.52–0.68)
Occipital lobe	NA	0.59 (0.51–0.68)
Ventral diencephalon	NA	0.51 (0.43–0.60)
Multivariable analysis		
Logistic regression	0.63 (0.55–0.71)	0.65 (0.58–0.73)
Random forest	0.60 (0.52–0.68)	0.62 (0.54–0.71)

Note:—NA indicates not applicable.

**Table 3: Pearson correlation coefficients for NQ and NR volumetrics in regions with the same name**

Feature	Pearson Coefficient ( <i>r</i> )	95% CI	<i>P</i> Value
Thalamus	0.60	0.51–0.69	<.001
Putamen	0.61	0.51–0.69	<.001
Lateral ventricles	0.99	0.99–0.99	<.001
Hippocampus	0.79	0.72–0.83	<.001
Cerebellum	0.87	0.83–0.90	<.001
Amygdala	0.71	0.64–0.78	<.001

Intracranial volume normalization did not result in improvement of the performance of hippocampal volume in the linear model (AUC = 0.65 for NQ, and 0.61 for NR), and adding age as a covariate to the normalized intracranial hippocampal volumes did not result in improvement (AUC = 0.66 for NQ and 0.62 for NR).

With the NR, the AUC value for the hippocampus was significantly greater ( $P < .05$ ) than the AUC value for the cerebellum, lateral ventricle, and the thalamus. With the NQ, the AUC value for the hippocampus was significantly greater ( $P < .05$ ) than the AUC value for the caudate, cerebellum, lateral ventricle, pallidum, putamen, and thalamus.

There was no statistically significant difference in AUC values by using hippocampal volumes obtained by NR versus NQ ( $P = .657$ ).

Pearson correlation coefficients are listed in Table 3. There was a statistically significant correlation between NQ and NR absolute volumes in all compatible regions tested: thalamus, lateral ventricle, hippocampus, cerebellum, putamen, and amygdala ( $P < .05$ ). The scatterplot and Bland-Altman plot for hippocampal volume are shown in Fig 1B. There was a small bias for NR with respect to NQ of  $-0.12$  mL (95% CI,  $-0.02$  to  $-0.22$  mL).

The AUC for the HOC score was 0.64 (95% CI, 0.56–0.72). There was no statistically significant difference in AUC values by using the NQ HOC score versus NQ hippocampal volume ( $P = .1827$ ).

## DISCUSSION

Our study, with the complete set of raw volumetric measures available in 2 fully automated, commercially available brain volumetric software packages, confirms previous studies documenting the prognostic utility of hippocampal volume in patients with MCI. Indeed, hippocampal volumes provided the highest prognostic value of all individual regions available in both the NQ and NR volumetric reports. Multivariable analysis, including the use of a cross-validated machine-based learning classifier algorithm, which incorporated other brain regions available in these software packages, did not provide additional predictive value compared with a model based on just hippocampal volumes. The results of our study are in agreement with previous studies that have found hippocampal volumes to be most predictive of conversion of MCI to AD, with little added benefit from additional brain regions. For example, a large study of patients in the ADNI using semiautomated methods reported that hippocampal volumes were the most predictive of conversion compared with other regional and whole-brain measures.<sup>29</sup> Automated volumetric measurements of the hippocampus also had high predictive values for conversion.<sup>2</sup> A 2-year clinical follow-up study of patients with MCI with manual methods demonstrated that baseline hippocampal volumes had high predictive value.<sup>5</sup> Despite the differences in segmentation techniques (manual, semiautomated, and automated) among the 3 studies, all these studies reported that hippocampal volumes were the most predictive of MCI conversion.

On the other hand, a number of previous studies have implicated brain regions other than the hippocampus as either adding to the prognostic value of hippocampal volume or having prognostic value in their own right, though direct comparison with our study is limited due to differences in methodologies and primary outcomes. In one study using manual methods, a bivariate model of hippocampal volumes and follow-up changes to either ventricular volume or whole-brain volume was shown to have prognostic value in predicting conversion of MCI to AD.<sup>7</sup> Another study with automated methods found that amygdala and caudate nucleus volumes were predictive of MCI conversion, whereas hippocampal volumes were not.<sup>12</sup> A separate study with automated methods reported that deep gray matter structures, including the amygdala, thalamus, putamen, and nucleus accumbens, were predictive of conversion of MCI to AD, in addition to the hippocampus.<sup>30</sup> Temporal horn volumes were shown to be more predictive than hippocampal volumes in one study using semiautomated methods,<sup>31</sup> whereas the left lateral temporal lobe and left parietal cortex were the most predictive factors in another study using automated methods.<sup>32</sup>

A likely reason for discrepancies with these studies is that volume loss in AD-affected regions is correlated with that of the hippocampus, and these collinear effects are therefore lost in the multivariable model when one accounts for hippocampal volume. For example, in our study, volume of the hippocampus was highly correlated with that of the amygdala ( $r = 0.71$ ) and cortical gray matter ( $r = 0.52$ ) for NQ, and amygdala ( $r = 0.75$ ) and temporal lobe ( $r = 0.57$ ) for NR. Another explanation for the differences in results, compared with our study, could be the use of different segmentation methods. In our study, we used an automated segmentation method with NQ and NR, which may be

less accurate in measuring smaller and deeper structures compared with manual or semiautomated segmentation methods.

Also of note in this study, the HOC score, a composite index of hippocampal and temporal horn volume, did not outperform hippocampal volume alone. The HOC score has been advocated as a more accurate measurement of hippocampal tissue loss, which accounts for individual variations in hippocampal size by accounting for temporal horn volume. In our study population, temporal horn volume was somewhat collinear with hippocampal volume ( $r = -0.33$ ), and we did not find the HOC to be more predictive of conversion than hippocampal volume alone, suggesting that the additional variance added by temporal horn volume might be redundant information and/or noise. Although temporal horn volume was a significant predictor of conversion in the univariable logistic regression model, it lost significance when hippocampal volume was added to the model and negligibly increased the AUC. Nevertheless, our results do not exclude the possibility that combining these 2 measures may be helpful in the preclinical or mild dementia phases of the disease spectrum.

Our study has several limitations. It was theoretically possible that the ROC area for hippocampal volume might have been larger if hippocampal volumes were normalized to intracranial volume and adjusted for age.<sup>33</sup> Such corrections might adjust for bias across subjects since ICV-normalized and age-adjusted values are already available in the volumetric reports. Nevertheless, for our primary analysis, we chose to use the raw values, in keeping with our purpose of comparing single-versus-combined measures within the same subjects, which should not be affected by normalizing by intracranial volume, a practice that could introduce additional noise into the measures. Although age may differentially affect various brain regions, we did not have age-adjusted values available for most regions in the NQ volumetric report and therefore did not include these measures in our primary analysis. Even so, secondary analyses of hippocampal volume with intracranial volume normalization, as well as including age as a covariate in the linear model, did not yield additional prognostic efficacy. Of note, 1 group, the Coalition Against Major Diseases, reported higher AUC values for the prediction of MCI conversion to AD (in 2 years) in their de novo analysis, based on automated hippocampal volume obtained by several methods: 0.7565 (LEAP; Learning Embeddings for Atlas Propagation, <http://www.ixico.com/additional-information/leap-analysis>), 0.7516 (NeuroQuant), 0.7536 (FreeSurfer; <http://surfer.nmr.mgh.harvard.edu>), and 0.7290 (HMAPS; Hippocampus Multi-Atlas Propagation and Segmentation).<sup>34</sup> However, the same committee also reported a range of AUC values between 0.60 and 0.77 based on their literature review, and our results are within this range.

Both the NQ and NR programs provide a large number of data values (55 for the general morphometry report through NQ, 140 for the NR report) that we did not fully include in our machine-based learning classifier program. Increasing the number of features can hurt the performance of a classifier, given a limited sample size. Initially, we found that the addition of the full dataset resulted in a worse predictive value for MCI conversion, which we speculate was due to overfitting of the training samples. This is a common phenomenon when the

number of predictor variables is high compared with the number of subjects in the training set.<sup>35</sup> Hence, we decided to limit our analysis to the most promising features provided in the NQ and NR volumetric reports.

In some cases, we were not able to process the MR images for the baseline sequence, which excluded some patients from the study, all of whom were chosen through the ADNI data base. In total, 89 of 281 patients were excluded due to inadequate NQ or NR morphometry data. This was most often due to patient age and sex missing from the anonymized header information, a requirement of the NQ processing pipeline but a situation unlikely to occur in clinical practice because these are standard DICOM data fields. In addition, some patients had inadequate MPRAGE sequences due to technical factors, including motion, which we were not able to segment by using the NQ or NR programs; in these cases, we substituted a repeat MPRAGE sequence. Despite occasional differences in our ability to process a particular sequence, we found a strong linear correlation ( $r > 0.60$ ) between NQ and NR for all compatible regions tested. Of interest, for hippocampal volume, there was a small underestimation bias for NR with respect to NQ, which may be attributable to different segmentation algorithms. Two outliers were noted, showing differences of approximately  $\pm 3$  mL between software packages, which were attributable to segmentation differences.

Another limitation is that we did not incorporate biomarkers, neuropsychological assessments, or longitudinal imaging measures in our study. Rather, we sought to limit our study to testing the prognostic efficacy of 2 commercially available brain volumetric software packages in their own right, rather than by using additional data, which might confound direct comparisons. Of note, the Mini-Mental State Examination and ADAS-13 AUC scores were as high as, or higher than, those of hippocampal volume alone, in agreement with previous studies looking at the prognostic efficacy of multiple biomarkers in MCI.<sup>36</sup> This result is not surprising given that a dichotomous end point measure, conversion, was used to assess cognitive decline, and cognitive measures at baseline would be expected to strongly predict how close a subject is to the point of conversion. Future effort in evaluating the prognostic efficacy of these software packages should involve combining volumetric data with a suite of biomarkers and neuropsychological assessments by using machine-based learning approaches and continuous, rather than categorical, outcomes of cognitive decline. Subject factors such as financial or socioeconomic ones also remained unadjusted in the analysis. These factors would unlikely affect the within-subject design and were omitted to keep the comparison straightforward between single-versus-multiple volumetric outputs directly available from the imaging software packages.

Despite these limitations, our study had several strengths, including a large sample size of 192 patients through the ADNI data base. Indeed, ADNI uses a wide variety of vendor platforms, field strengths, and harmonized pulse sequences that patients are likely to encounter in future clinical settings. We also selected patients with 3-year clinical follow-up, which provided a more accurate designation of future conversion status, compared with studies with shorter follow-up duration.

## CONCLUSIONS

Of the multiple regional volume measures available in current FDA-cleared brain volumetric software, hippocampal volume remains the best single predictor of MCI conversion to AD at 3-year follow-up. Combining volumetrics, by using multivariable approaches including a machine-learning classifier, does not appear to add additional prognostic efficacy. Therefore, future prognostic studies in mild cognitive impairment, combining such tools with demographic and other biomarker measures, are justified in using hippocampal volume as the only volumetric biomarker.

## ACKNOWLEDGMENTS

The authors thank Dr. James Brewer (for NeuroQuant) and Dr. Jamila Ahdidan (for Neuroreader) for providing comments for this article.

Disclosures: Jeffrey R. Petrella—UNRELATED: Board Membership: Piramal Healthcare, Comments: advisory board; Payment for Lectures including Service on Speakers Bureaus: Eli Lilly, Comments: Speakers Bureau. J. Ikkena was funded by the 2015 RSNA Research Medical Student Grant #RMS1504.

## REFERENCES

1. 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;25:1197–203 CrossRef Medline
2. Fritzsche KH, Stieltjes B, Schlindwein S, et al. **Automated MR morphometry to predict Alzheimer's disease in mild cognitive impairment.** *Int J Comput Assist Radiol Surg* 2010;5:623–32 CrossRef Medline
3. Douaud G, Menke RA, Gass A, et al. **Brain microstructure reveals early abnormalities more than two years prior to clinical progression from mild cognitive impairment to Alzheimer's disease.** *J Neurosci* 2013;33:2147–55 CrossRef Medline
4. Wang PN, Lirng JF, Lin KN, et al. **Prediction of Alzheimer's disease in mild cognitive impairment: a prospective study in Taiwan.** *Neurobiol Aging* 2006;27:1797–806 CrossRef Medline
5. Eckerström C, Olsson E, Borga M, et al. **Small baseline volume of left hippocampus is associated with subsequent conversion of MCI into dementia: the Göteborg MCI study.** *J Neurol Sci* 2008;272:48–59 CrossRef Medline
6. Varon D, Barker W, Loewenstein D, et al; Alzheimer's Disease Neuroimaging Initiative. **Visual rating and volumetric measurement of medial temporal atrophy in the Alzheimer's Disease Neuroimaging Initiative (ADNI) cohort: baseline diagnosis and the prediction of MCI outcome.** *Int J Geriatr Psychiatry* 2015;30:192–200 CrossRef Medline
7. Jack CR Jr, Shiung MM, Weigand SD, et al. **Brain atrophy rates predict subsequent clinical conversion in normal elderly and amnesic MCI.** *Neurology* 2005;65:1227–31 CrossRef Medline
8. Desikan RS, Fischl B, Cabral HJ, et al. **MRI measures of temporoparietal regions show differential rates of atrophy during prodromal AD.** *Neurology* 2008;71:819–25 CrossRef Medline
9. Devanand DP, Pradhaban G, Liu X, et al. **Hippocampal and entorhinal atrophy in mild cognitive impairment: prediction of Alzheimer disease.** *Neurology* 2007;68:828–36 CrossRef Medline
10. Petrella JR, Coleman RE, Doraiswamy PM. **Neuroimaging and early diagnosis of Alzheimer disease: a look to the future.** *Radiology* 2003;226:315–36 CrossRef Medline
11. Petrella JR. **Neuroimaging and the search for a cure for Alzheimer disease.** *Radiology* 2013;269:671–91 CrossRef Medline
12. Liu Y, Paajanen T, Zhang Y, et al; Addneuromed Consortium. **Analysis of regional MRI volumes and thicknesses as predictors of conversion from mild cognitive impairment to Alzheimer's disease.** *Neurobiol Aging* 2010;31:1375–85 CrossRef Medline
13. Ross DE, Ochs AL, Seabaugh JM, et al; Alzheimer's Disease Neuroimaging Initiative. **Man versus machine: comparison of radiologists' interpretations and NeuroQuant® volumetric analyses of brain MRIs in patients with traumatic brain injury.** *J Neuropsychiatry Clin Neurosci* 2013;25:32–39 CrossRef Medline
14. Davatzikos C, Fan Y, Wu X, et al. **Detection of prodromal Alzheimer's disease via pattern classification of magnetic resonance imaging.** *Neurobiol Aging* 2008;29:514–23 CrossRef Medline
15. Misra C, Fan Y, Davatzikos C. **Baseline and longitudinal patterns of brain atrophy in MCI patients, and their use in prediction of short-term conversion to AD: results from ADNI.** *Neuroimage* 2009;44:1415–22 CrossRef Medline
16. Davatzikos C, Bhatt P, Shaw LM, et al. **Prediction of MCI to AD conversion, via MRI, CSF biomarkers, and pattern classification.** *Neurobiol Aging* 2011;32:2322.e19–27 CrossRef Medline
17. Vemuri P, Wiste HJ, Weigand SD, et al; Alzheimer's Disease Neuroimaging Initiative. **MRI and CSF biomarkers in normal, MCI, and AD subjects: predicting future clinical change.** *Neurology* 2009;73:294–301 CrossRef Medline
18. Plant C, Teipel SJ, Oswald A, et al. **Automated detection of brain atrophy patterns based on MRI for the prediction of Alzheimer's disease.** *Neuroimage* 2010;50:162–74 CrossRef Medline
19. Fennema-Notestine C, McEvoy LK, Hagler DJ Jr; The Alzheimer's Disease Neuroimaging Initiative. **Structural neuroimaging in the detection and prognosis of pre-clinical and early AD.** *Behav Neurol* 2009;21:3–12 CrossRef Medline
20. Moradi E, Pepe A, Gaser C, et al; Alzheimer's Disease Neuroimaging Initiative. **Machine learning framework for early MRI-based Alzheimer's conversion prediction in MCI subjects.** *Neuroimage* 2015;104:398–412 CrossRef Medline
21. Zhan Y, Chen K, Wu X, et al; Alzheimer's Disease Neuroimaging Initiative. **Identification of conversion from normal elderly cognition to Alzheimer's disease using multimodal support vector machine.** *J Alzheimers Dis* 2015;47:1057–67 CrossRef Medline
22. Adaszewski S, Dukart J, Kherif F, et al; Alzheimer's Disease Neuroimaging Initiative. **How early can we predict Alzheimer's disease using computational anatomy?** *Neurobiol Aging* 2013;34:2815–26 CrossRef Medline
23. Ahdidan J, Raji CA, DeYoe EA, et al. **Quantitative neuroimaging software for clinical assessment of hippocampal volumes on MRI.** *J Alzheimers Dis* 2015;49:723–32 CrossRef Medline
24. Brewer JB, Magda S, Airriess C, et al. **Fully-automated quantification of regional brain volumes for improved detection of focal atrophy in Alzheimer disease.** *AJNR Am J Neuroradiol* 2009;30:578–80 CrossRef Medline
25. Ahdidan J, Hviid LB, Chakravarty MM, et al. **Longitudinal MR study of brain structure and hippocampus volume in major depressive disorder.** *Acta Psychiatr Scand* 2011;123:211–19 CrossRef Medline
26. DeLong ER, DeLong DM, Clarke-Pearson DL. **Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach.** *Biometrics* 1988;44:837–45 CrossRef Medline
27. Breiman L. **Random forests.** *Machine Learning* 2001;45:5–32 CrossRef
28. Heister D, Brewer JB, Magda S, et al; Alzheimer's Disease Neuroimaging Initiative. **Predicting MCI outcome with clinically available MRI and CSF biomarkers.** *Neurology* 2011;77:1619–28 CrossRef Medline
29. Risacher SL, Saykin AJ, West JD, et al; Alzheimer's Disease Neuroimaging Initiative (ADNI). **Baseline MRI predictors of conversion from MCI to probable AD in the ADNI cohort.** *Curr Alzheimer Res* 2009;6:347–61 CrossRef Medline
30. Yi HA, Moller C, Dieleman N, et al. **Relation between subcortical grey matter atrophy and conversion from mild cognitive impairment to Alzheimer's disease.** *J Neurol Neurosurg Psychiatry* 2016;87:425–32 CrossRef Medline
31. Macdonald KE, Bartlett JW, Leung KK, et al; ADNI investigators. **The**

- value of hippocampal and temporal horn volumes and rates of change in predicting future conversion to AD. *Alzheimer Dis Assoc Disord* 2013;27:168–73 CrossRef Medline
32. Karas G, Sluimer J, Goekoop R, et al. **Amnesic mild cognitive impairment: structural MR imaging findings predictive of conversion to Alzheimer disease.** *AJNR Am J Neuroradiol* 2008;29:944–49 CrossRef Medline
33. Wyman BT, Harvey DJ, Crawford K, et al; Alzheimer's Disease Neuroimaging Initiative. **Standardization of analysis sets for reporting results from ADNI MRI data.** *Alzheimers Dement* 2013;9:332–37 CrossRef Medline
34. Hill DL, Schwarz AJ, Isaac M, et al. **Coalition Against Major Diseases/European Medicines Agency biomarker qualification of hippocampal volume for enrichment of clinical trials in prodromal stages of Alzheimer's disease.** *Alzheimers Dement* 2014;10:21–9.e3 CrossRef Medline
35. Hawkins DM. **The problem of overfitting.** *J Chem Inf Comput Sci* 2004;44:1–12 CrossRef Medline
36. Gomar JJ, Bobes-Bascaran MT, Conejero-Goldberg C, et al; Alzheimer's Disease Neuroimaging Initiative. **Utility of combinations of biomarkers, cognitive markers, and risk factors to predict conversion from mild cognitive impairment to Alzheimer disease in patients in the Alzheimer's Disease Neuroimaging Initiative.** *Arch Gen Psychiatry* 2011;68:961–69 CrossRef Medline