Application of Automated Medial Temporal Lobe Atrophy Scale to Alzheimer Disease

Basil H. Ridha, MRCP; Josephine Barnes, MA, PhD; Laura A. van de Pol, MD; Jonathan M. Schott, MD, MRCP; Richard G. Boyes, ME; Musib M. Siddique, PhD; Martin N. Rossor, MD, FRCP; Philip Scheltens, MD, PhD, FRCP; Nick C. Fox, MD, FRCP

Objective: To compare an automated intensity-based measure of medial temporal atrophy in Alzheimer disease (AD) with existing volumetric and visually based methods.

Design: Longitudinal study comparing a medial temporal atrophy measure with 2 criterion standards: (1) total hippocampal (HC) volume adjusted for total intracranial volume and (2) standard visual rating scale of medial temporal atrophy.

Setting: Cognitive disorders specialist clinic.

Participants: Forty-seven patients with AD and 26 age- and sex-matched controls.

Intervention: Subjects were scanned using volumetric T1-weighted magnetic resonance imaging at baseline and 1 year later.

Main Outcome Measure: Automated Medial Temporal Lobe Atrophy Scale (ATLAS) score, derived from dividing mean intensity of a standardized perihippocampal volume by that of a standardized pontine volume.

Results: Patients with AD had significantly reduced ATLAS scores and HC volumes and increased visual rating scores at baseline and repeat scanning. Rates of HC atrophy and decline in the ATLAS score were significantly higher in patients with AD compared with controls. The ATLAS scores were significantly correlated with HC volumes and visual rating scores. With specificity set at 85%, the sensitivities of HC volume and visual rating scale score were similar (84% and 86%, respectively), whereas ATLAS score had a lower sensitivity (73%). At repeat scanning, all 3 measures had similar sensitivities (86%-87%). Rate of decline in the ATLAS score required a similar sample size to HC atrophy rate to provide statistical power to clinical trials, but being automated, it is less labor intensive.

Conclusions: Like the visual rating scale, ATLAS is a simple medial temporal atrophy measure, which has the additional advantage of being able to track AD progression on serial imaging.

Arch Neurol. 2007;64:849-854

In Alzheimer disease (AD), the pathological process is initially focused in the medial temporal lobe structures such as the entorhinal cortex and hippocampus (HC) before spreading to involve neocortical regions. There have been numerous efforts to develop magnetic resonance (MR) imaging techniques as diagnostic markers of AD or disease progression, and these have particularly focused on assessing atrophy of medial temporal lobe structures, including the HC. Methods to assess HC atrophy have largely been based on volumetric measurements or visual rating scales. Volumetric measurements typically rely on manual outlining of the medial temporal lobe structures on (serial) MR images, which is time-consuming and prone to interrater and intrarater variability. Visual rating scales, although simple to use, making them practical for clinical application, were not designed to detect atrophy progression on serial imaging; their quantized nature makes them insensitive to change over time. A technique that combines the simplicity of visual rating scales and the ability of volumetric measurement to track disease progression would be desirable.

This study describes a simple, quick, and operator-independent quantitative measure of medial temporal lobe atrophy that can detect disease progression on serial imaging: the Automated Medial Temporal Lobe Atrophy Scale (ATLAS). We compared the ability of this method
to distinguish patients from controls and to track change over time with volumetric measures of HC atrophy using manual outlining and the standard Scheltens method for visually rating medial temporal lobe atrophy.4

METHODS

SUBJECTS

Forty-seven patients with probable AD (20 men, mean ± SD age, 65.7 ± 11.5 years) were recruited from the Cognitive Disorders Clinic at the National Hospital for Neurology and Neurosurgery, London, England. Diagnosis was made according to the National Institute of Neurological and Communicative Disorders and Stroke–Alzheimer's Disease and Related Disorders Association criteria.5 Twenty-six neurologically healthy spouses or friends of patients (12 men, mean ± SD age, 65.9 ± 11.4 years) were recruited as controls. Written informed consent was obtained from all participating subjects (with the agreement of the next of kin of all patients with AD) as approved by the local ethics committee. The Mini-Mental State Examination (MMSE)6 was performed on all subjects at the time of baseline ethics committee. The Mini-Mental State Examination (MMSE)6 was performed on all subjects at the time of baseline

MR ACQUISITION

Magnetic resonance imaging was performed at baseline and 1 year later on the same 1.5-T Signa unit (GE Medical Systems, Milwaukee, Wis). T1-weighted volumetric images were obtained using a spoiled fast gradient recalled acquisition in a steady state sequence technique with a 24-cm field of view and 256 × 256 matrix to provide 124 contiguous 1.5-mm-thick slices in the coronal plane. Scan parameters were as follows: repetition time, 15 milliseconds; echo time, 5.4 milliseconds; inversion time, 650 milliseconds; and flip angle, 15°.

HC VOLUME MEASUREMENT

Image analysis was performed with baseline and repeat scans presented side by side in a random order with the operator blind to subject status and scan order. Volumetric measurements were done using the MIDAS program.7 All scans were first registered to a standard brain template using a 6-df algorithm to reduce variability in neuroanatomical landmarks used in delineating the HC.8 Follow-up scans were then accurately registered to the baseline images using a 9-df algorithm.9 Each HC was manually traced using multiple views to include the cornu ammonis, gyrus dentatus, and subiculum. Total (right + left) HC volumes were calculated and adjusted for total intracranial volumes (TIVs) to correct for differences in head size between individuals.9 The TIVs were calculated according to a previously described protocol.10 Total HC volumes were then standardized to mean TIV of control subjects. The standardization was carried out by using the slope of the relationship between total HC volume and TIV, estimated from a linear regression model relating total HC volume to TIV, with both variables on logarithmic scales. Annualized rates of total HC atrophy were calculated as a percentage of total baseline HC volume.

VISUAL RATING OF MEDIAL TEMPORAL ATROPHY

An experienced rater blind to subject status and scan order reviewed the baseline and repeat scans, 1 at a time, and rated the level of right and left medial temporal lobe atrophy on a scale ranging from zero (no atrophy) to 4 (severe atrophy).4 Means of left and right scores were calculated for each scan. Annualized decline on the visual rating scale was calculated by subtracting baseline from repeat mean scores and then dividing by the scan interval in years.

ATLAS SCORE

For the generation of ATLAS scores, subjects were divided into 2 groups: a “method development” subset consisting of 10 patients and 8 controls and a “test” subset consisting of the remaining 37 patients and 18 controls. Scans from the method development subset were used to generate standardized ATLAS reference volumes in the perihippocampal and pontine regions (Figure 1). The standardized perihippocampal volumes (left=9.78 mL and right=10.95 mL) were generated from uniting manually segmented HC volumes and a standardized pontine volume (1.78 mL) from intersecting roughly outlined pontine volumes. The pons was defined laterally by planes containing the entry points of the trigeminal nerves, superiorly by the junction with the midbrain, and inferiorly by the junction with the medulla. The standardized perihippocampal and pontine volumes were then copied onto all baseline and repeat scans of subjects in the test subset. Mean intensity of each standardized region was calculated using MIDAS software. The intensity of each standardized perihippocampal volume was divided by the intensity of the corresponding standardized pontine volume to obtain the ATLAS score. The mean of left and right ATLAS scores was then calculated for each scan. Annualized rates of decline in mean ATLAS scores were expressed as a percentage of baseline mean ATLAS score. To assess whether there was an intrinsic loss of T1-weighted MR signal within the HC region, we also measured the mean signal intensity within the HC formation, normalized to that of the standardized pontine region (HC formation–pons intensity ratio).
Medial temporal lobe atrophy measures and MMSE scores of the test subset of subjects with AD and controls were compared using 2-sample \( t \) tests. The receiver operating characteristic curves were plotted for the 3 medial atrophy measures at baseline and repeat scanning and their annualized rates of decline. For comparison purposes, the sensitivity of each medial temporal lobe atrophy measure was calculated with specificity set at 85%. The relationships between medial temporal atrophy measures were assessed within patients with AD and controls separately, using the Pearson correlation coefficient (\( r \)). For illustrative purposes, a comparison was made between the sample sizes needed to power a clinical trial using HC atrophy rate and decline in the ATLAS score, based on the assumption of a treatment effect corresponding to a 20% reduction in the annualized rate of change of the outcome measure, 90% power, and a 2-sided \( \alpha = 0.05 \) level of significance.

### RESULTS

**SUBJECTS**

Patients with AD and control groups (method development and test subgroups combined) were not significantly different in terms of age (2-sample \( t \) test, \( P = .92 \)) or sex (\( \chi^2 \) test, \( P = .78 \)). Patients with AD had significantly lower mean±SD MMSE scores than controls at baseline (19.7±4.0 vs 29.5±0.7; \( P < .001 \)). Within the test subset, MMSE score declined by mean±SD 1.9±3.3 points per year among patients with AD compared with 0.6±1.0 point per year among controls (\( P = .12 \)) (Table).

**HC VOLUME MEASUREMENT**

Patients with AD had significantly reduced total HC volumes corrected for TIV at baseline (mean±SD, 4.13±0.68 vs 5.50±0.49 mL; \( P < .001 \)) and repeat scanning (mean±SD, 3.96±0.71 vs 5.48±0.48 mL; \( P < .001 \)) and significantly increased rates of total HC atrophy as compared with controls (mean±SD, 4.49%±2.91% vs 0.37%±0.93% per year; \( P < .001 \)) (Table).

### STATISTICAL ANALYSIS

Medial temporal lobe atrophy measures and MMSE scores of the test subset of subjects with AD and controls were compared using 2-sample \( t \) tests. The receiver operating characteristic curves were plotted for the 3 medial atrophy measures at baseline and repeat scanning and their annualized rates of decline. For comparison purposes, the sensitivity of each medial temporal lobe atrophy measure was calculated with specificity set at 85%. The relationships between medial temporal atrophy measures were assessed within patients with AD and controls separately, using the Pearson correlation coefficient (\( r \)). For illustrative purposes, a comparison was made between the sample sizes needed to power a clinical trial using HC atrophy rate and decline in the ATLAS score, based on the assumption of a treatment effect corresponding to a 20% reduction in the annualized rate of change of the outcome measure, 90% power, and a 2-sided \( \alpha = 0.05 \) level of significance.

### VISUAL RATING SCALE OF MEDIAL TEMPORAL ATROPHY

Patients with AD had significantly higher mean±SD visual rating scores than controls at baseline (2.2±0.7 vs 0.7±0.6; \( P < .001 \)) and repeat scanning (2.4±0.7 vs 0.9±0.5; \( P < .001 \)) (Table). However, there was no statistically significant difference in the annualized decline of the visual rating score between patients and controls (mean±SD, 0.15±0.41 vs 0.20±0.43 points/y; \( P = .70 \)) (Table).

### ATLAS SCORES

Patients with AD had significantly lower mean±SD ATLAS scores than controls at baseline (0.70±0.05 vs 0.77±0.04; \( P < .001 \)) and repeat scanning (0.68±0.05 vs 0.77±0.04; \( P < .001 \)) (Table). In addition, patients with AD had significantly increased rates of decline in ATLAS score compared with controls (mean±SD, 2.90%±2.00% vs 0.57%±1.59% per year; \( P < .001 \)) (Table). There were no significant differences in the HC formation–pons intensity ratio between patients with AD and controls at baseline (mean±SD, 0.79±0.03 vs 0.78±0.02; \( P = .09 \)) and repeat scanning (mean±SD, 0.78±0.03 vs 0.77±0.02; \( P = .19 \)) or in their rates of decline (0.54%±2.10% vs 0.09%±1.59% per year; \( P = .42 \)). In addition, there were no statistically significant differences in the mean±SD absolute T1-weighted image intensity within the standardized pontine volume between patients with AD and controls at baseline (844±77 vs 806±98; \( P = .12 \)) and repeat scanning (1736±128 vs 1711±135; \( P = .52 \)).

### SENSITIVITY AND SPECIFICITY

**Figure 2** shows the receiver operating characteristic curves for the 3 medial temporal lobe atrophy measures.
at baseline and repeat scanning and their rates of atrophy. At baseline, for a specificity of 85%, the sensitivities of HC volume measurement and visual rating scale were similar (84% vs 86%), whereas the sensitivity of the ATLAS measure was lower at 73%. However, at repeat scanning, all 3 measures had similar sensitivities between 86% and 87%. Rate of HC atrophy had a higher sensitivity than that for rate of decline in the ATLAS score (81% vs 68%). Rate of decline in the visual rating scale had a low sensitivity of 15%.

RELATIONSHIP BETWEEN VARIOUS MEDIAL TEMPORAL ATROPHY MEASURES

In cross-sectional analyses of the data, there were significant correlations between ATLAS scores and visual rating scores at baseline (among patients with AD, \(r = 0.73; P < .001\); among controls, \(r = 0.51; P = .03\)) and repeat scanning (among patients with AD, \(r = 0.60; P = .001\); among controls, \(r = 0.70; P = .001\)) and between ATLAS scores and HC volumes at baseline (among patients with AD, \(r = 0.46; P = .004\); among controls, \(r = 0.51; P = .03\) and among patients with AD, \(r = 0.44; P = .006\); among controls, \(r = 0.53; P = .02\)). The correlations between the visual rating scores and HC volumes at baseline were \(r = 0.51\) and \(P < .001\) among patients with AD and \(r = 0.55\) and \(P = .02\) among controls at repeat scanning.

Rates of ATLAS score decline were not significantly correlated with rates of HC atrophy among patients (\(r = 0.12; P = .48\)) or controls (\(r = 0.19; P = .45\)). In addition, annualized decline in the visual rating scale score was not correlated with rates of HC atrophy or rates of ATLAS score decline in both subject groups (all \(P\) values > .05).

SAMPLE SIZE CALCULATIONS

Based on the assumption of a treatment effect corresponding to a 20% change in outcome measure, 90% power, and a 2-sided .05 level of significance, rates of HC atrophy or ATLAS score decline required similar numbers of subjects per treatment arm to provide sufficient statistical power (220 vs 250, respectively). In comparison, rates of MMSE score decline would require 1583 subjects.

COMMENT

We describe a simple automated measure of medial temporal lobe atrophy based on measurement of signal intensity of a standardized volume centered on the HC and adjusted to the intensity of a standardized pontine volume. This ratio measure is largely driven by the amount of hypointense cerebrospinal fluid (CSF) (low intensity on T1-weighted MR scans) relative to gray and white matter within the standardized perihippocampal region.

The ATLAS measure was compared with 2 of the most widely used quantitative measures of medial temporal lobe atrophy: HC volume measurement and the Scheltens visual rating scale. Cross-sectionally, all 3 measures showed significant differences between patients with AD and controls. There were significant correlations between all measures, with the strongest being between the ATLAS measure and the visual rating scale, probably reflecting the strong influence of the relative amount of CSF spaces...
around the HC on both measures. Hippocampal volume measurement and the visual rating scale performed the best at group discrimination, with sensitivity and specificity around 85% for mild to moderate AD (mean MMSE score, approximately 19 of 30).

On serial scanning, significant reductions in HC volume measurement and ATLAS scores were seen when patients with AD were compared with controls. However, rates of ATLAS score decline were not significantly correlated with rates of HC atrophy among patients or controls. This may be due to the fact that the 2 measures reflect slightly different aspects of medial temporal lobe atrophy: HC atrophy rate reflects volume changes within a well-delineated structure, whereas the ATLAS score takes into account, in addition to changes in intensity within the HC, the relative change in intensity in other surrounding medial temporal lobe structures and CSF spaces. The small number of subjects included in this analysis may also be a source of error.

Sample size calculations estimate that using either measure required a similar number of subjects per treatment arm to power a clinical trial. However, as a diagnostic measure, rate of decline in the ATLAS score was less sensitive than rate of HC atrophy in differentiating patient and control groups. As expected, the visual rating scale—a quantized measure—was not sensitive to change over time (1-year interval) in the AD group.

The ATLAS method is simple, quick, and automated and so avoids intraindividual and interindividual variability inherent of the semiautomated segmentation techniques. By contrast, manual volumetric measures of each HC region take around 45 minutes, and thus, segmenting the left and right HC regions on baseline and repeat scans of the 55 subjects in the test subset of this study took approximately 165 hours. This compares with just less than 2 hours to generate the ATLAS scores on the same data set, which in turn is similar to the time needed to apply the visual rating scale.

The reduction in the ATLAS score among patients with AD is thought to reflect the reduction of the relative amount of brain tissue (HC, parahippocampal regions) to CSF within the standardized perihippocampal region. We hypothesized that there may be additional contribution due to a reduction in the signal intensity within the HC in patients with AD compared with controls as a direct result of the neurodegenerative process. However, the ratio of intensity of the manually segmented HC region to that of the standardized pontine region was not significantly different between patients with AD and controls.

The ATLAS method relies on correcting the intensity of a standardized perihippocampal region to that of a pontine region. The standardized perihippocampal region was generated from uniting segmented HC regions on separate samples of patients with AD and controls, with all scans placed in the same standard space. The region is 4 to 5 times the volume of an average HC, with the rest of the volume containing adjacent medial temporal lobe structures and CSF spaces, such as the temporal horn and ambient cistern. Because the intensity of a T1-weighted image is meaningless on its own, a standardized pontine region was used as an internal reference point. The pons was chosen because it is relatively unaffected by AD pathological features, apart from the locus coeruleus and pontine raphe, which are small pontine structures. There were no statistically significant differences in the mean intensity within the standardized pontine volume between patients with AD and controls at baseline and repeat scanning. This supports the use of the pons as an internal reference point because absolute T1-weighted image intensity values are meaningless on their own. The close anatomical proximity to the HC on coronal MR acquisition makes the pons subject to similar local magnetic field variations as the HC, thus reducing image intensity errors arising from magnetic field inhomogeneity (Figure 1).

The standardized perihippocampal volume generated aims to encompass the HC region proper of all the subjects, when scans are placed in the same standard space. The exact choice of standardized volume could be expanded to allow for greater variation in the HC location in other study populations or optimized to focus on the subregions that offer either greatest diagnostic discrimination or, alternatively, greatest sensitivity to change.

The rate of HC atrophy found in this study was similar to that reported by Jack et al in a multicentered clinical trial setting of patients with mild to moderate AD. The median annualized rate of HC atrophy was 4.9% (range, −0.5% to 15.2%) and similar to ours (mean ± SD, 4.49% ± 2.91%). This was associated with a median decline of 1.9 points on the MMSE (range, −7.2 to 18.1), similar to our findings (mean ± SD, 1.9 ± 3.3) over a 1-year period.

Several methods of automated HC segmentation have previously been described. Most of these rely on either the manual identification of several HC landmarks on each scan or a detailed algorithm program based on the intensities and spatial anatomical relationship of different brain structures to guide HC outlining. Webb et al devised an automated method to detect HC atrophy in patients with temporal lobe epilepsy based on the analysis of the image intensity differences between patients and controls within a volume of interest centered on the HC. Thompson et al generated color-coded maps to visualize the HC atrophy rate using 3-dimensional parametric mesh models of manually segmented HC regions on serial scans. In addition, automated quantification of HC atrophy rates has been derived using regional fluid registration or by calculating the regional boundary shift integral. However, both methods require manual segmentation of the baseline HC region. Rusinek et al used the boundary shift integral analysis applied to a volume of interest centered on the HC to calculate the rate of medial temporal lobe atrophy. Compared with these automated methods of HC or medial temporal lobe atrophy, ATLAS requires relatively little image postprocessing and prerequisites for automation. Further research is necessary to investigate the relevance of the ATLAS method for longitudinal clinical and research purposes.

In summary, we report a simple technique for assessing medial temporal lobe atrophy based on intensity measurement in a standardized perihippocampal volume using established T1-weighted volumetric scans. The measure significantly differentiates patients with AD from con-
controls at cross-sectional and longitudinal levels. We show that this differentiation is largely driven by the reduction of the brain tissue relative to surrounding CSF spaces rather than intrinsic signal loss within the HC. Like the visual rating scale, the technique is simple to use and so may be of value in clinical practice but, in addition, has the ability to track disease progression on serial imaging without the need for expert assessment or labor-intensive manual measures.

Acknowledgment: We are grateful to Dave MacManus, MSc, and Philippa Bartlett, DCR, for their help with this study. We particularly thank the subjects who participated in this study.

REFERENCES