International Psychogeriatrics (2012), 24:7, 1065–1075 © International Psychogeriatric Association 2012
doi:10.1017/S1041610212000191

Cross-validation of brain structural biomarkers and cognitive aging in a community-based study

James T. Becker,[1,2,3] Ranjan Duara,[4,5,6] Ching-Wen Lee,[1,7] Leonid Teverovsky,[8] Beth E. Snitz,[2] Chung-Chou H. Chang[7,9] and Mary Ganguli[1,2,10]

[1] *Department of Psychiatry, University of Pittsburgh, Pittsburgh, Pennsylvania, USA*
[2] *Department of Neurology, University of Pittsburgh, Pittsburgh, Pennsylvania, USA*
[3] *Department of Psychology, University of Pittsburgh, Pittsburgh, Pennsylvania, USA*
[4] *Wien Center for Alzheimer's Disease & Memory Disorders, Mt. Sinai Medical Center, Miami Beach, Florida, USA*
[5] *Miller School of Medicine, University of Miami, Coral Gables, Florida, USA*
[6] *Wertheim College of Medicine, Florida International University, Miami, Florida, USA*
[7] *Department of Biostatistics, University of Pittsburgh, Pittsburgh, Pennsylvania, USA*
[8] *Department of Computer Science, Carnegie Mellon University, Pittsburgh, Pennsylvania, USA*
[9] *Department of Medicine, University of Pittsburgh, Pittsburgh, Pennsylvania, USA*
[10] *Department of Epidemiology, University of Pittsburgh, Pittsburgh, Pennsylvania, USA*

ABSTRACT

Background: Population-based studies face challenges in measuring brain structure relative to cognitive aging. We examined the feasibility of acquiring state-of-the-art brain MRI images at a community hospital, and attempted to cross-validate two independent approaches to image analysis.

Methods: Participants were 49 older adults (29 cognitively normal and 20 with mild cognitive impairment (MCI)) drawn from an ongoing cohort study, with annual clinical assessments within one month of scan, without overt cerebrovascular disease, and without dementia (Clinical Dementia Rating (CDR) < 1). Brain MRI images, acquired at the local hospital using the Alzheimer's Disease Neuroimaging Initiative protocol, were analyzed using (1) a visual atrophy rating scale and (2) a semi-automated voxel-level morphometric method. Atrophy and volume measures were examined in relation to cognitive classification (any MCI and amnestic MCI vs. normal cognition), CDR (0.5 vs. 0), and presumed etiology.

Results: Measures indicating greater atrophy or lesser volume of the hippocampal formation, the medial temporal lobe, and the dilation of the ventricular space were significantly associated with cognitive classification, CDR = 0.5, and presumed neurodegenerative etiology, independent of the image analytic method. Statistically significant correlations were also found between the visual ratings of medial temporal lobe atrophy and the semi-automated ratings of brain structural integrity.

Conclusions: High quality MRI data can be acquired and analyzed from older adults in population studies, enhancing their capacity to examine imaging biomarkers in relation to cognitive aging and dementia.

Key words: aging, Alzheimer's disease, cognitive disorders, dementia, imaging, neuroepidemiology

Introduction

A valid biomarker for Alzheimer's disease (AD) will help to distinguish between cognitively impaired individuals with and without AD pathology. It will also help to identify those cognitively healthy individuals likely to develop cognitive decline over time (Breteler, 2011). Biomarkers are not tied to the clinical spectrum of the disease but instead reflect the presence of disease pathology. Therefore, they can both confirm the clinical diagnosis and provide a measure of pathological staging. Although not currently a requirement for diagnosis of AD, brain imaging data are increasingly regarded as biomarkers supporting the clinical diagnosis (Dubois *et al.*, 2009; McKhann *et al.*, 2011). Neuroimaging data also provide evidence of the brain substrate upon which the neurodegenerative process takes place, thus determining the rate of clinical expression of the underlying pathology.

Although brain imaging is less commonly performed in population studies than in clinical

Correspondence should be addressed to: James T. Becker, PhD, Neuropsychology Research Program, Department of Psychiatry, University of Pittsburgh, Suite 830, 3501 Forbes Avenue, Pittsburgh, PA 15213, USA. Phone: +1 412-246-6970; Fax: +1 412-246-6873. Email: BeckerJT@upmc.edu. Received 8 Nov 2011; revision requested 13 Dec 2011; revised version received 26 Dec 2011; accepted 12 Jan 2012. First published online 16 March 2012.

https://doi.org/10.1017/S1041610212000191 Published online by Cambridge University Press

studies, the issue of underlying brain structural integrity may be particularly important in studies that draw their samples from the community at large. Being more representative of the base population than patients typically encountered in specialty memory disorder clinics, community study participants tend to be older, sicker, and often have less access to medical care. These differences can potentially attenuate the relationships between clinical deficits and structural abnormalities. Thus, it is essential to complement imaging biomarker studies from specialty clinical centers by replicating them at the community level. The Cardiovascular Health Study (CHS), for example, which conducts scans at academic medical centers, has provided valuable population-based data showing risk of incident dementia (Carmichael *et al.*, 2007) and the independent contributions of exercise and vascular risk to brain structure and dementia (Erickson *et al.*, 2010).

However, in many population-based studies, it is not feasible for participants to utilize imaging facilities at major medical centers. Here, we report a case-control study within a newer population-based cohort. We performed MRI brain scans at a local community hospital, using the state-of-the-art Alzheimer's Disease Neuroimaging Initiative (ADNI) image acquisition protocol (Mueller *et al.*, 2005), and applying two independent approaches to analyze the images remotely. The first goal of this study was to determine the feasibility of incorporating a magnetic resonance (MR) imaging protocol, acquired at a *non-research* imaging center, into a community-based study of brain aging. The second goal was to cross-validate the two independent image analytic approaches against each other as well as against the clinical assessments performed during the preceding month.

Methods

Study area, sampling, and recruitment

The study cohort, named the Monongahela–Youghiogheny Healthy Aging Team (MYHAT), is an age-stratified random sample of individuals aged 65 years or over, sampled from voter registration lists for a small-town region of southwestern Pennsylvania, USA (Ganguli *et al.*, 2009). The University of Pittsburgh Institutional Review Board approved the community outreach, recruitment, and assessment protocols. Recruitment criteria were (a) age 65 years or over, (b) living within the selected towns, and (c) not already in long-term care institutions. Individuals were ineligible if (a) they were too ill to participate, (b) had severe vision or hearing impairments, or (c) were decisionally incapacitated. Over a two-year period, 2036 individuals were recruited and enrolled. When asked whether they would consider participating in an MRI research study, 63% of the study participants responded affirmatively.

Overview of assessment

We first screened participants with the age/education-corrected Mini-Mental State Examination (Folstein *et al.*, 1975; Mungas *et al.*, 2005). We classified 54 individuals (2.7%) with scores below 21 as too severely impaired for a study of mild cognitive impairment (MCI) and did not assess them further. We then performed a detailed neurobehavioral assessment on the remaining 1982 participants, usually in their homes. We repeated this assessment annually thereafter. At each assessment, we classified participants according to several criteria for MCI. Here we focus on two definitions of MCI: a functional definition, using the Clinical Dementia Rating (CDR) (Hughes *et al.*, 1982; Morris *et al.*, 1993), and a purely cognitive classification based on neuropsychological test performance relative to norms (Ganguli *et al.*, 2010a).

Neuropsychological assessment and classification

We measured cognitive functioning with a detailed test neuropsychological battery (Ganguli *et al.*, 2010b). We categorized each test according to the principal cognitive domain that it assessed (attention/processing speed, executive function, language, memory, and visuospatial function) and created a composite score for each domain. We used these data for the previously described "cognitive classification" of participants as having (a) normal cognition (all domain composites within one standard deviation (SD) of the appropriate mean); (b) severe cognitive impairment (at least two domain composites worse than two SDs below the appropriate mean); or (c) MCI (at least one domain composite between one and two SDs of the appropriate mean) (Ganguli *et al.*, 2010a). We further sub-classified those with MCI, based on the presence or absence of a memory deficit, into amnestic and non-amnestic MCI.

Clinical dementia rating

We also rated participants on the CDR scale, based on a standardized assessment protocol focused on cognitively driven everyday functioning, and which we rated independently of cognitive test performance (Ganguli *et al.*, 2010a). We selected participants for the MRI study only if their summary CDR ratings were <1 (i.e. no dementia); thus, those

https://doi.org/10.1017/S1041610212000191 Published online by Cambridge University Press

included in the current study are all classified as CDR = 0 (normal) or CDR = 0.5 (very mild MCI).

Etiological classification

A Web-based, online diagnostic process developed for the MYHAT study (Weir *et al.*, 2011) allowed expert raters (JB, RD, BS, MG) to view relevant data for study participants on a secure website and render a diagnostic impression about likely etiology. Participants were etiologically classified into neurodegenerative (including AD), vascular, mixed degenerative and vascular, other (including depression, metabolic causes), and none (normal). Raters did not review the MRI data before rendering the etiologic classification. We selected the consensus or modal diagnosis for each participant. For the current analyses, we compared only the etiological groups designated as neurodegenerative and none (normal) as the other subgroups were too small in size.

Selection of participants for the MRI study

The time window for the MRI scan was no greater than one month after the nearest annual MYHAT clinical/cognitive assessment. Each time we recruited an MCI participant, we randomly selected two cognitively normal individuals matched on gender and age (±5 years) to serve as controls for the MCI case. Inclusion criteria for both cases and controls included willingness to undergo a brain MRI scan for research purposes, clinical study assessment within one month, absence of contra-indications to MRI, absence of clinical cerebrovascular disease (history of stroke or TIA), and CDR < 1. The first 52 participants (17 cases and 35 controls) who met the inclusion criteria underwent scanning; three controls were later reclassified as MCI based on revised norms; image datasets from three other controls were subsequently discarded because they did not meet quality control standards. The final study sample included 29 cognitively normal and 20 cognitively impaired participants.

MRI image acquisition

Participants underwent scanning on a 1.5T GE scanner at the MR Center of UPMC McKeesport, the community hospital located in the heart of the study area. The Department of Radiology at the hospital downloaded the Spoiled Gradient Recall (SPGR) sequence appropriate for their machine and software from the ADNI website (http://www.loni.ucla.edu/ADNI/Research/Cores/). Briefly, the image parameters were: 3D acquisition, sagittal, TR = 6.48 ms, TE = 1.488 ms, TI = 1000 ms, 192 slices of 1.2-mm thickness, flip = 8°, FOV = 24 × 24 cm^2, image matrix = 256 × 256 × 1. Copies of all of the scans were sent to the University of Pittsburgh on CDs, and stored for later processing.

MR image analysis

Visual Rating Scale (VRS)

Individual SPGR scans labeled with the participant's age, sex, and study identification numbers, loaded on CDs, were provided to one of us (RD) for evaluation using a validated visual rating system (VRS) (Urs *et al.*, 2009). VRS ratings were performed on a standard coronal slice of 1.2-mm thickness, perpendicular to the AC-PC line, intersecting the midpoint of the mammillary bodies. Standardized rating of the degree of atrophy seen in the hippocampus and entorhinal and perirhinal cortices in each target image was compared to a set of reference images provided by the software in VRS that was developed for this purpose. The reference images depict five possible levels of atrophy for each structure to be rated (i.e. "0" for no atrophy and "4" for the most severe atrophy). The reliability for individual structures ranges between 0.75 and 0.94 for inter-rater reliability, and 0.87 and 0.93 for intra-rater reliability (Urs *et al.*, 2009). Only two MRI scans were rejected from this study because they had sufficient movement or positioning artifact to compromise reliable VRS ratings. A medial temporal atrophy score for each side of the brain was created by taking the average of the ratings for these three structures. Additional standardized visual ratings included cortical atrophy, the size of the lateral and third ventricles, and the extent of periventricular and subcortical white matter hyperintensities.

Semi-automated Volumetric Analysis (SVA)

Each of the image data sets was run through a processing pipeline that included bias field correction, intra- and inter-slice intensity normalization, rigid body registration to the Colin27 template, and affine and fully deformable registration to the Colin27 template (Teverovsky *et al.*, 2011). Total intracranial volume was calculated following an atlas-based brain extraction from the skull, with the subsequent addition of the cerebrospinal fluid (CSF) volume between the surface of the brain and the inner table of the skull.

Each brain image was processed using the FMRIB's Integrated Registration and Segmentation Tool (FIRST, http://www.fmrib.ox.ac.uk/fsl/first/index.html) (Patenaude *et al.*, 2011), which provides model-based segmentation and registration (see Figure 1). The output from the brain extraction was processed using FMRIB's

https://doi.org/10.1017/S1041610212000191 Published online by Cambridge University Press

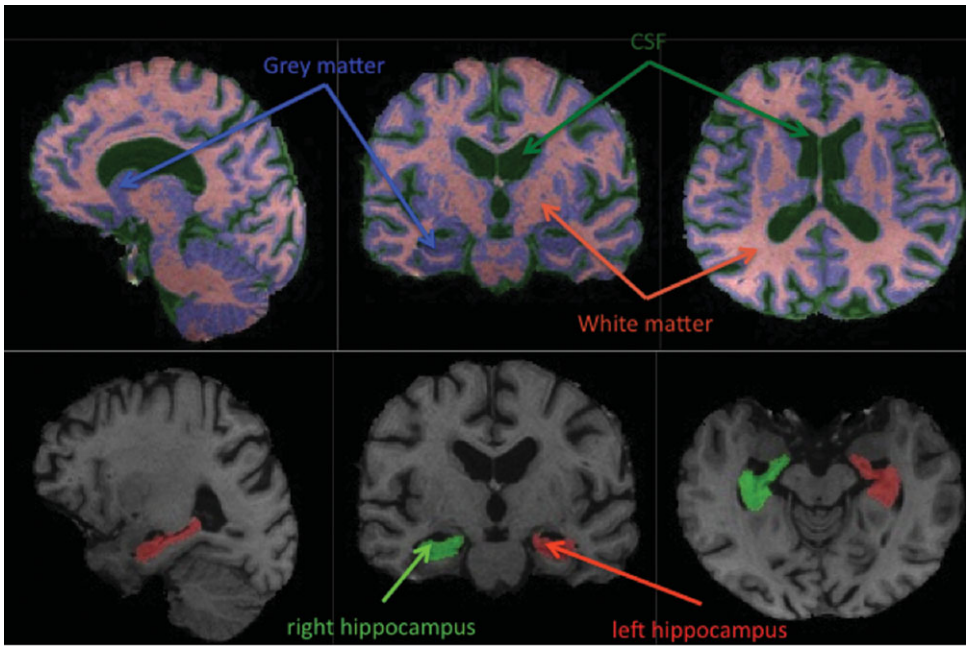


Figure 1. Sample segmented images. (Top) Gray matter/white matter/CSF segmentation; (bottom) hippocampus segmentation.

Automated Segmentation Tool (FAST v4) (http://www.fmrib.ox.ac.uk/fsl/fast4/index.html) to identify the relevant structures, including the hippocampus, amygdala, nucleus accumbens, thalamus, putamen, caudate nucleus, globus pallidus, total gray matter (GM), total white matter (WM), and total CSF. The raw data were converted to percent of total intracranial volume (TIV) using the data acquired from the initial pipeline processing. The hippocampus segmentation was of inadequate quality for three scans, as identified by visual inspection.

Statistical analysis

We compared demographic and cognitive characteristics of the cases and controls using χ^2 or Fisher's Exact tests, and *t*-tests or Wilcoxon two-sample test, as appropriate to the distribution of the given variable. Although our cases and controls were selected based on their cognitive classification, we also classified them according to the CDR and etiologic classification, which are independent of, and do not map entirely to, the cognitive classification. Mean VRS and SVA MRI measures were separately compared between these groups. Rank/order correlations were calculated between the overall distribution of the atrophy ratings and automated volumetrics.

Results

Of the 49 participants with complete data, 29 had been selected as cognitively normal controls for 20 individuals classified as MCI (14 of them amnestic MCI) based on the purely cognitive classification approach. On the functionally based CDR, 33 were rated as CDR = 0 (normal) and 16 as CDR = 0.5 (very mild impairment). On the web-based etiologic classification rendered by clinical raters, 28 were classified as none (normal) and 12 as neurodegenerative; the remaining, much smaller, etiological groups were not included in analysis. Since these three ratings were independent of one another, they do not overlap totally; three individuals who were clinically impaired, and one who was diagnosed as neurodegenerative, scored in the normal ranges on the neuropsychological tests.

The 20 MCI cases and 29 cognitively normal controls were similar in age (mean age of 78.6 years, with SD of 5.0 and 5.6, respectively), and gender (35% and 27.6% male). Difference in educational level (70.0% and 37.9% with greater than high school education) reached borderline significance ($p = 0.065$). The two groups were not significantly different in self-reported diabetes mellitus, hypertension, and current smoking status.

Cognitive classification (MCI vs. normal cognition) (Table 1)

Looking first at VRS, the overall cortical atrophy ratings, as well as the ratings of ventricular expansion, differed significantly between the MCI and normal groups. Measures of structural integrity in the medial temporal lobe, as well as the summary measures of medial temporal atrophy, significantly discriminated between the MCI and normal groups.

https://doi.org/10.1017/S1041610212000191 Published online by Cambridge University Press

https://doi.org/10.1017/S1041610212000191 Published online by Cambridge University Press

Table 1. MRI measurements as a function of study group

	NORMAL (n = 29)		ALL MCI (n = 20)					AMNESTIC MCI (n = 14)				
	MEAN	SD	MEAN	SD	STATISTIC[+]	p-VALUE[+]	EFFECT SIZE (HEDGES' G)	MEAN	SD	STATISTIC[+]	P-VALUE[+]	EFFECT SIZE (HEDGES' G)
Semi-automated volumetric (SVA) measures[†]												
Gray matter	37.213	0.753	37.033	1.038	0.7	0.487	−0.20	37.151	0.994	0.2	0.823	−0.07
White matter	26.544	0.604	26.358	0.689	1.0	0.321	−0.29	26.401	0.783	0.7	0.512	−0.20
CSF	21.497	0.607	21.719	0.722	−1.2	0.250	0.33	21.864	0.787	−1.7	0.100	0.51
Hippocampus right	0.217	0.025	0.201	0.034	390.0	0.031*	−0.55	0.193	0.020	194.0	0.005*	−0.97
Hippocampus left	0.204	0.026	0.185	0.032	394.0	0.037*	−0.68	0.176	0.027	203.0	0.010*	−0.99
Hippocampus both	0.421	0.049	0.385	0.059	372.0	0.013*	−0.66	0.369	0.036	176.0	0.001*	−1.08
Duara atrophy ratings												
Cortical atrophy	1.672	0.685	2.100	0.788	584.0	0.074	0.58	2.214	0.802	383.5	0.043*	0.70
Lat. ventricle, Rt.	1.414	0.825	1.900	0.852	588.5	0.063	0.57	2.000	0.961	380.5	0.056	0.63
Lat. ventricle, Lft.	1.638	0.766	2.050	0.759	593.5	0.048*	0.53	2.143	0.864	384.0	0.044*	0.59
3rd ventricle	1.241	0.830	1.800	0.894	601.0	0.035*	0.64	2.000	0.877	402.5	0.014*	0.84
PWMH	1.517	0.829	1.950	0.887	579.0	0.097	0.50	2.000	0.784	369.0	0.102	0.56
SWMH	1.379	0.942	1.650	0.933	544.5	0.341	0.28	1.857	0.864	362.5	0.140	0.49
Hippocampus, Rt.	1.379	0.820	2.100	1.119	603.5	0.033*	0.74	2.286	1.204	399.5	0.018*	0.89
Hippocampus, Lft.	1.586	1.053	2.100	1.119	573.0	0.133	0.47	2.500	1.019	401.0	0.017*	0.82
Entorhinal cortex, Rt.	1.138	1.217	1.950	1.099	611.5	0.024*	0.68	2.214	0.975	410.0	0.010*	0.88
Entorhinal cortex, Lft.	1.379	1.237	2.150	1.137	606.5	0.030*	0.63	2.500	0.941	415.5	0.006*	0.91
Parahippocampal gyrus, Rt.	0.931	1.163	1.450	1.099	572.0	0.127	0.45	1.643	1.082	377.0	0.066	0.59
Parahippocampal gyrus, Lft.	1.034	1.239	1.750	1.251	592.0	0.056	0.57	2.000	1.177	395.0	0.024*	0.74
Medial temporal atrophy, Rt.	1.149	0.941	1.835	1.012	612.0	0.027*	0.69	2.049	1.003	409.5	0.012*	0.88
Medial temporal atrophy, Lft.	1.334	1.084	2.001	1.014	602.5	0.042*	0.62	2.334	0.887	412.5	0.010*	0.91

[+]Appropriately tested by the *t*-test or Wilcoxon two-sample test.
*$p < 0.05$.
[†]Expressed as a percentage of intra-cranial volume.
MCI = mild cognitive impairment; CSF = cerebrospinal fluid; Lat = lateral; RT = right; Lft = left; PWMH = periventricular white matter hyperintestine; SWMH = subcortical white matter abnormalities.

Table 2. MRI measurements as a function of CDR

	CDR 0 (N = 33)		CDR 0.5 (N = 16)				
	MEAN	SD	MEAN	SD	STATISTIC[+]	P-VALUE[+]	EFFECT SIZE (HEDGES' G)
Semi-automated volumetric measures[†]							
Gray matter	37.152	0.823	37.114	1.001	0.1	0.890	−0.04
White matter	26.424	0.638	26.560	0.652	−0.7	0.492	0.20
CSF	21.471	0.639	21.827	0.652	−1.8	0.075	0.52
Hippocampus, right	0.217	0.030	0.196	0.024	298.0	0.035*	−0.69
Hippocampus, left	0.204	0.028	0.181	0.028	277.0	0.012*	−0.77
Hippocampus, both	0.421	0.057	0.377	0.042	273.0	0.010*	−0.79
Visual rating scale atrophy measures							
Cortical atrophy	1.742	0.730	2.063	0.772	457.0	0.201	0.40
Lat. ventricle, Rt.	1.576	0.830	1.688	0.946	422.0	0.628	0.12
Lat. ventricle, Lft.	1.773	0.782	1.875	0.806	426.0	0.562	0.12
3rd ventricle	1.333	0.890	1.750	0.856	463.5	0.160	0.44
PWMH	1.576	0.830	1.938	0.929	460.5	0.182	0.39
SWMH	1.424	0.902	1.625	1.025	431.5	0.481	0.20
Hippocampus, Rt.	1.485	0.906	2.063	1.124	489.5	0.052	0.55
Hippocampus, Lft.	1.636	1.084	2.125	1.088	470.0	0.132	0.42
Entorhinal cortex, Rt.	1.182	1.185	2.063	1.124	507.5	0.023*	0.71
Entorhinal cortex, Lft.	1.394	1.223	2.313	1.078	515.5	0.014*	0.73
Perirhinal cortex, Rt.	0.879	1.083	1.688	1.138	500.0	0.028*	0.69
Perirhinal cortex, Lft.	0.970	1.212	2.063	1.124	528.0	0.006*	0.87
Medial temporal atrophy, Rt.	1.182	0.936	1.938	1.020	510.0	0.023*	0.74
Medial temporal atrophy, Lft.	1.334	1.058	2.166	0.981	514.5	0.018*	0.75

[+]Appropriately tested by the *t*-test or Wilcoxon two-sample test.
*$p < 0.05$.
[†]Expressed as a percentage of intra-cranial volume.
CDR = Clinical Dementia Rating; CSF = cerebrospinal fluid; Lat = lateral; RT = right; Lft = left.

Next, looking at SVA, there were significant differences between the volumes of the right hippocampus, the left hippocampus, and the total hippocampal volume (bilaterally) in those with normal cognition compared with those with any MCI, and specifically those with amnestic MCI. By contrast, there were no significant differences in total GM, WM and CSF measures (expressed as a percentage of TIV) as a function of cognitive classification.

Thus, the expert visual ratings were more sensitive than the semi-automated volumetrics to the presence of MCI, as indicated by the greater number of brain regions that differed significantly in their atrophy ratings between groups.

Clinical dementia rating (Table 2)

Similar patterns of association were found when we compared brain volumes as a function of CDR rating; the hippocampal regions and the total hippocampal volume successfully discriminated between the unimpaired (CDR = 0) and the mildly impaired (CDR = 0.5) participants. The summary VRS variables reflecting medial temporal atrophy (both right and left), as well as measures of the entorhinal and perirhinal cortices, were significantly different between groups. The SVA volumes of the left and right hippocampus, however, were not significantly different between groups.

Etiological classification

With regard to the VRS, a similar pattern was seen as (Table 3) for the CDR, which included the entorhinal and perirhinal cortices, and the medial temporal atrophy summary score. Periventricular white matter hyperintensities were present in a significantly greater proportion in individuals with presumed neurodegenerative etiologies than in the cognitively normal controls.

When comparing SVA between the individuals with presumed neurodegenerative etiology and those classified as normal, the hippocampal volume measurements, and in particular the right hippocampus and bilateral hippocampal volumes as assessed using the semi-automated techniques, significantly differentiated between the two groups.

https://doi.org/10.1017/S1041610212000191 Published online by Cambridge University Press

Table 3. MRI measurements as a function of etiologic classification

	NONE (N = 28)		NEURODEGEN (N = 12)				
	MEAN	SD	MEAN	SD	STATISTIC[+]	P-VALUE[+]	EFFECT SIZE (HEDGES' G)
Semi-automated volumetric measures[†]							
Gray matter	37.105	0.803	36.776	1.022	213.0	0.343	−0.35
White matter	26.467	0.622	26.381	0.801	226.0	0.568	−0.12
CSF	21.429	0.601	21.812	0.788	297.0	0.144	0.53
Hippocampus, Rt.	0.220	0.030	0.192	0.019	146.0	0.006*	−0.95
Hippocampus, Lft.	0.205	0.029	0.180	0.024	167.0	0.026*	−0.83
Hippocampus, both	0.425	0.057	0.372	0.038	150.0	0.008*	−0.93
Visual rating scale atrophy measures							
Cortical atrophy	1.696	0.685	2.167	0.835	299.5	0.100	0.59
Lat. ventricle, Rt.	1.429	0.836	1.833	1.030	284.0	0.250	0.41
Lat. ventricle, Lft.	1.661	0.770	2.000	0.853	292.5	0.155	0.39
3rd ventricle	1.250	0.844	1.667	0.888	287.5	0.207	0.45
PWMH	1.393	0.832	2.083	0.793	317.0	0.033*	0.77
SWMH	1.321	0.983	1.750	0.866	284.5	0.237	0.41
Hippocampus, Rt.	1.357	0.826	2.333	1.073	337.5	0.007*	0.99
Hippocampus, Lft.	1.500	1.036	2.250	1.055	311.0	0.056	0.66
Entorhinal cortex, Rt.	1.107	1.166	2.250	0.965	337.5	0.009*	0.94
Entorhinal cortex, Lft.	1.321	1.219	2.583	0.900	346.0	0.004*	1.02
Perirhinal cortex, Rt.	0.857	1.113	1.917	0.900	332.5	0.011*	0.92
Perirhinal cortex, Lft.	0.929	1.215	2.250	0.965	345.0	0.004*	1.05
Medial temporal atrophy, Rt.	1.107	0.912	2.168	0.927	344.5	0.006*	1.06
Medial temporal atrophy, Lft.	1.251	1.045	2.362	0.823	343.0	0.007*	1.03

[+]Appropriately tested by the *t*-test or Wilcoxon two-sample test.
*$p < 0.05$.
[†]Expressed as a percentage of intra-cranial volume.
CSF = cerebrospinal fluid; Lat = lateral; RT = right; Lft = left.

Comparison of semi-automated and visual ratings

The correlations between the two approaches to measure CNS integrity showed consistent patterns between the two methods (Tables 4 and 5). The automated method for determining overall gray matter and white matter volume did not yield values that were significantly correlated with any of the visual atrophy ratings (Table 4). The total volume of CSF was, however, significantly correlated with the size of the third ventricle and that of the lateral ventricle on the right. However, no other correlations were statistically significant. The volumes of the hippocampal formation correlated significantly with the corresponding atrophy ratings (Table 5).

Discussion

In this study we have demonstrated the feasibility of acquiring research-quality MRI data at local community facilities, and shown that the resulting information can be analyzed remotely using the same image analytic tools available at academic medical centers. We were able to distinguish cognitively normal individuals from those with MCI, using both volumetric and visual rating methods to measure atrophy of medial temporal structures. Our method of acquiring and analyzing the anatomical MRIs did not rely on the availability of new, higher field, and research center-based scanners. As such, our experience as reported in this study is potentially relevant to large-scale research projects devoted to understanding factors associated with the development of clinically significant abnormalities in cognitive functions.

The findings that we report here are not new; atrophy in the hippocampus and expansion of the ventricles have been noted for decades from research on aging and cognition using computed tomography (de Leon *et al.*, 1984; de Leon *et al.*, 1989) and MRI scans (de Leon *et al.*, 1993). However, our data were acquired at a community hospital, which is rarely the case in neuroimaging research. This hospital was not only conveniently located for our study participants but also familiar to them, thus maximizing the

https://doi.org/10.1017/S1041610212000191 Published online by Cambridge University Press

Table 4. Spearman rank-order correlations between visual ratings and quantitative measures of atrophy and ventricular expansion

	CORTICAL ATROPHY	LATERAL VENTRICLE (RIGHT)	LATERAL VENTRICLE (LEFT)	THIRD VENTRICLE	PERIVENTRICULAR WHITE MATTER LESIONS	SUBCORTICAL WHITE MATTER LESIONS
Gray matter	−0.266	−0.138	−0.164	0.062	−0.049	0.124
White matter	−0.182	−0.179	−0.212	−0.098	−0.104	0.110
CSF	0.166	0.290*	0.170	0.367*	0.248	0.133

*$p < 0.05$.

likelihood of their consenting to undergo research scans. The alternative, requiring them to travel from the small mill towns of southwestern Pennsylvania into the city of Pittsburgh to be scanned at the University's MR Research Center, would have been viewed as an unreasonable burden. Here, the barrier to participation is more cultural than physical, and is related to the historic separation of the communities. In other regions of the USA (and perhaps elsewhere), actual physical distance and lack of available transportation may be additional limiting factors. If only a select few participants take part in an imaging study, the resulting sample bias can obviate some of the benefits of generalizability from using a true population-based sample. However, before utilizing convenient and familiar facilities for the imaging procedures, it is critical to demonstrate that the data that are acquired from such sites can be acquired and analyzed using the standard neuroimaging tools.

In terms of logistics and feasibility, both visual rating and volumetric methods of analysis worked smoothly and cost-effectively. In this study, the expert visual ratings showed greater sensitivity to the various clinical classifications than did this particular semi-automated volumetric approach to image analysis. The visual ratings can be easily completed by trained raters in any setting (e.g. study field office), and are constrained by fewer problems with image quality than are the mechanistic approaches. In other words, a visual rating of medial temporal lobe atrophy may be possible even when the image quality may preclude (in the absence of additional post-processing) automated analysis. As further external validation of our approach, the measures of medial temporal lobe atrophy that we obtained here did not differ significantly from previously reported findings in amnestic MCI in a clinic sample, using the same visual rating method (Duara *et al.*, 2008). More importantly, the measures of medial temporal atrophy among our cognitively normal community-based participants were significantly greater (i.e. showed more atrophy) in both the right ($t(33) = 3.76$, $p < 0.001$) and left ($t(32) = 4.09$, $p < 0.001$) hemispheres than those observed in the previous clinical study. This discrepancy reinforces the importance of selection factors, suggesting that the brains (or at least the temporal lobes) of putatively normal participants in a community-based study, representative of the general population, have suffered more damage, likely secondary to medical comorbidities, than have the brains of volunteers in academic research centers. Of course, inherent methodological differences between clinical and population studies might also have contributed to the differences in results.

A major advantage of the semi-automated volumetric approach is that large numbers of scans can be analyzed. Our data processing pipeline utilized state-of-the-art methodologies in order to "clean" the data, and then to reliably segment and classify brain regions. We used publicly available software on a distributed network within the research group, which optimized the time to complete the analysis. While other methods (e.g. Freesurfer) are increasingly popular because of the quality of the data, these techniques take a large amount of time to process an individual scan. Our pipeline was able to accomplish the entire process, start to finish, in less than one hour per scan. We were capable of running 6–9 scans at any given time, which means that the total amount of time taken to extract critical data from this data set was on of the order of 10 hours (i.e. overnight). A test-run of nearly 1000 brains was completed in less than one week. Alterations in brain structure are often small, and may coexist with other CNS abnormalities (e.g. small vessel disease and white matter hyperintensities). Therefore, studies must include a sufficient number of participants to provide sufficient power to disentangle the effects of the various independent variables on the expression of cognitive dysfunction. In large cohort studies (e.g. >2000 participants), where the MRI data do not need to be analyzed until all scans are acquired, an automated method will

https://doi.org/10.1017/S1041610212000191 Published online by Cambridge University Press

Table 5. Spearman rank-order correlations* between visual and quantitative measures of medial temporal (MT) lobe

	RIGHT HIPPOCAMPUS	LEFT HIPPOCAMPUS	RIGHT ENTORHINAL CORTEX	LEFT ENTORHINAL CORTEX	RIGHT PERIRHINAL CORTEX	LEFT PERIRHINAL CORTEX	RIGHT MT ATROPHY	LEFT MT ATROPHY
Right hippocampus	−0.394	−0.433	−0.601	−0.565	−0.418	−0.411	−0.533	−0.521
Left hippocampus	−0.306	−0.512	−0.581	−0.683	−0.515	−0.575	−0.541	−0.649
Total hippocampus	−0.387	−0.502	−0.631	−0.663	−0.494	−0.534	−0.577	−0.628

*All correlation coefficients significant at $p < 0.05$.

have advantages. Automated methods also have the advantage that no observer/rater training is necessary, and there is no need for intra- and inter-rater reliability checks (or measures of rater drift).

In this validation study, we examined the relative merits of semi-automated techniques for the measurement of both large and small structures in the brains of participants in an epidemiological study. We found that the measured volume of CSF, expressed as a proportion of the total intracranial volume, was significantly correlated with the size of the third ventricle and the size of the right lateral ventricle. We did not find correlations between the measures of overall cortical atrophy and the measured volumes of gray matter and white matter. However, with regard to the small structures of the medial temporal lobe, and in particular the hippocampus, we found large and statistically significant correlations between the semi-automated and visual rating systems. This means that, even for a structure as small as the hippocampus, the semi-automated technique is capable of reliably measuring volume of the structure as compared to the rated atrophy. Further, concurrent validity is provided by the fact that both the visual ratings and the semi-automated measurements were significantly associated with clinical classification, CDR, and presumed etiology. We have previously reported on the relationship between a different automated measure of ventricular expansion and a different visual rating of ventricular enlargement (Carmichael *et al.*, 2005a, 2005b). We found that the correlation between the two measures was high, but there were significant outliers, some of which reflected problems with the semi-automated technique, and some of which reflected idiosyncratic ratings by radiologists.

Advances in neuroimaging have led to major improvements in understanding brain aging and the ways in which altered brain structure and function lead to cognitive impairment and dementia. While neuroimaging is standard in the diagnosis of vascular dementia (Chui *et al.*, 1992; Roman *et al.*, 1993; Gorelick *et al.*, 2011), only the recently revised National Institute on Aging and the Alzheimer's Association (NIA-AA) criteria for the diagnosis of AD include structural MRI as a potential biomarker to confirm diagnosis and identify stage (McKhann *et al.*, 2011). Further, it is now clear that data on brain structure (and preferably also brain function) are required to understand the impact of various comorbidities on the clinical expression of a neurodegenerative condition like AD (e.g. Raji *et al.*, 2009; 2010; 2012). Comorbid conditions such as

https://doi.org/10.1017/S1041610212000191 Published online by Cambridge University Press

cerebrovascular disease alter the substrate onto which an AD-type or related neurodegenerative process is expressed, decreasing the extent of cognitive or brain reserve (Satz, 1991; Stern *et al.*, 1992; Satz *et al.*, 1993) and accelerating the clinical expression of the underlying neuropathology.

Thus, clinical research into cognitive impairment and dementia increasingly includes neuroimaging, and such research from tertiary care medical centers advances our understanding of these neurodegenerative processes. However, patients who participate in specialty clinic studies have many characteristics that set them apart from the rest of the community, potentially introducing selection bias into the research. Hence, results generated in specialty clinic studies must be replicated in population-based studies, allowing investigators to see the entire spectrum of the disease as it exists in the population at large. Conversely, the ability of population studies to perform high-quality assessments and diagnoses may increasingly depend on the availability of imaging data (if not of fluid biomarkers such as CSF) to confirm the diagnosis of concurrent disorders and identify predictors of future disorders. Yet, neuroimaging within representative population samples brings the challenge of implementing a clinical imaging protocol in a study cohort of individuals who are randomly selected from the population, are not seeking services, and often undergo their research assessments in their homes. The logistics of any imaging study must be sufficiently simple and flexible to engage the interest of such individuals and maximize their participation. It may be possible in the near future to design studies to process brain imaging data efficiently, and to integrate that information relatively quickly for subsequent analysis and identification of biomarkers of impending clinical disorder. When this has been accomplished, it will also become relevant to the care of patients in areas remote from tertiary medical care centers.

Conflict of interest

None.

Description of authors' roles

James T. Becker drafted the paper, provided critical revision of the paper for important intellectual content, participated in etiologic classification, and was responsible for volumetric analyses, analysis and interpretation of data, and statistical analysis, conception, and design. Ranjan Duara provided critical revision of the paper for important intellectual content, participated in etiologic classification, was responsible for visual ratings, analysis, and interpretation of data, and was involved in conception and design. Beth E. Snitz provided critical revision of the paper for important intellectual content, participated in etiologic classification, and was involved in conception and design. Ching-Wen Lee provided analysis and interpretation of data and was involved in statistical analysis. Chung-Chou (Joyce) Chang provided critical revision of the paper for important intellectual content, participated in statistical analysis, and was involved in conception and design. Leonid Teverovsky provided critical revision of the paper for important intellectual content and was involved in statistical analysis. Mary Ganguli provided critical revision of the paper for important intellectual content, participated in etiologic classification, carried out analysis and interpretation of data, and was involved in conception and design.

Acknowledgments

The authors thank the 52 MYHAT participants who underwent scans, and also the MYHAT research staff, particularly Amy Sinot, Kathryn McMichael, and Diane Kurta; Denise Davis and Daniel Galeza at the University of Pittsburgh Medical Center; as well as Annette Volbers and Dr. Shaikh Rashid at the MRI Center of UPMC McKeesport.

This research was supported in part by funds from the National Institute on Aging to M.G. (R01AG023651 and K24AG022035). Additional support came from the Neuroimaging Core of the Alzheimer's Disease Research Center at the University of Pittsburgh (AG05133), and we utilized analysis techniques developed in the context of the Cardiovascular Health Study–Cognition Study (AG020098; HC055222).

References

Breteler, M. M. (2011). Mapping out biomarkers for Alzheimer disease. *JAMA*, 305, 304–305.

Carmichael, O. T. *et al.* (2005a). Atlas-based hippocampus segmentation in Alzheimer's disease and mild cognitive impairment. *Neuroimage*, 27, 979–990.

Carmichael, O. T. *et al.* (2005b). Dementia-associated ventricular volume changes in a community cohort. *Neuroimage*, 26, S47.

Carmichael, O. T. *et al.* (2007). Cerebral ventricular changes associated with transitions between normal cognitive function, mild cognitive impairment, and dementia. *Alzheimer Disease and Related Disorders*, 21, 14–24.

https://doi.org/10.1017/S1041610212000191 Published online by Cambridge University Press

Chui, H. C., Victoroff, J. I., Margolin, D., Jagust, W., Shankle, R. and Katzman, R. (1992). Criteria for the diagnosis of ischemic vascular dementia proposed by the State of California Alzheimer's Disease Diagnostic and Treatment Centers. *Neurology*, 42, 473–480.

de Leon, M. J. *et al.* (1984). Positron emission tomography and computed tomography assessments of the aging human brain. *Journal of Computer Assisted Tomography*, 8, 88–94.

de Leon, M. J., George, A. E. and Reisberg, B. (1989). Alzheimer's disease: longitudinal CT studies of ventricular change. *American Journal of Roentgenology*, 152, 1257–1262.

de Leon, M. J. *et al.* (1993). The radiologic prediction of Alzheimer disease: the atrophic hippocampal formation. *American Journal of Neuroradiology*, 14, 897–906.

Duara, R. *et al.* (2008). Medial temporal lobe atrophy on MRI scans and the diagnosis of Alzheimer disease. *Neurology*, 71, 1986–1992.

Dubois, B., Picard, G. and Sarazin, M. (2009). Early detection of Alzheimer's disease: new diagnostic criteria. *Dialogues in Clinical Neurosciences*, 11, 135–139.

Erickson, K. I. *et al.* (2010). Physical activity predicts gray matter volume in late adulthood: the cardiovascular health study. *Neurology*, 75, 1415–1422.

Folstein, M. F., Folstein, S. E. and McHugh, P. R. (1975). "Mini-mental state": a practical method grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, 12, 189–198.

Ganguli, M., Snitz, B., Vander Bilt, J. and Chang, C. C. (2009). How much do depressive symptoms affect cognition at the population level? The Monongahela-Youghiogheny Healthy Aging Team (MYHAT) study. *International Journal of Geriatric Psychiatry*, 24, 1277–1284.

Ganguli, M., Chang, C. C., Snitz, B. E., Saxton, J. A., Vanderbilt, J. and Lee, C. W. (2010a). Prevalence of mild cognitive impairment by multiple classifications: the Monongahela-Youghiogheny Healthy Aging Team (MYHAT) project. *American Journal of Geriatric Psychiatry*, 18, 674–683.

Ganguli, M., Snitz, B. E., Lee, C. W., Vanderbilt, J., Saxton, J. A. and Chang, C. C. (2010b). Age and education effects and norms on a cognitive test battery from a population-based cohort: the Monongahela-Youghiogheny Healthy Aging Team. *Aging and Mental Health*, 14, 100–107.

Gorelick, P. B. *et al.* (2011). Vascular contributions to cognitive impairment and dementia: a statement for healthcare professionals from the American Heart Association/American Stroke Association. *Stroke*, 42, 2672–2713.

Hughes, C. P., Berg, L. and Danzinger, W. L. (1982). A new clinical scale for the staging of dementia. *British Journal of Psychiatry*, 140, 566–572.

McKhann, G. M. *et al.* (2011). The diagnosis of dementia due to Alzheimer's disease: recommendations from the National Institute on Aging and the Alzheimer's Association workgroup. *Alzheimer's and Dementia*, 7, 263–269.

Morris, J. C. *et al.* (1993). The Consortium to Establish a Registry for Alzheimer's Disease (CERAD). Part IV. Rates of cognitive change in the longitudinal assessment of probable Alzheimer's disease. *Neurology*, 43, 2457–2465.

Mueller, S. G. *et al.* (2005). The Alzheimer's disease neuroimaging initiative. *Neuroimaging Clinics of North America*, 15, 869–877.

Mungas, D., Reed, B. R., Tomaszewski-Farias, S. and DeCarli, C. (2005). Criterion-referenced validity of a neuropsychological test battery: equivalent performance in elderly Hispanics and non-Hispanic Whites. *Journal of the International Neuropsychological Society*, 11, 620–630.

Patenaude, B., Smith, S. M., Kennedy, D. N. and Jenkinson, M. (2011). A Bayesian model of shape and appearance for subcortical brain segmentation. *Neuroimage*, 56, 907–922.

Raji, C. A., Lopez, O. L., Kuller, L. H., Carmichael, O. T. and Becker, J. T. (2009). Age, Alzheimer disease, and brain structure. *Neurology*, 73, 1899–1905.

Raji, C. A. *et al.* (2010). Brain structure and obesity. *Human Brain Mapping*, 31, 353–364.

Raji, C. A. *et al.* (2012). White matter lesions and brain gray matter volume in cognitively normal elders. *Neurobiology of Aging*, 33, 834.e7–834.e16. Epub 2011 Sep 23.

Román, G. C. *et al.* (1993). Vascular dementia: diagnostic criteria for research studies: report of the NINDS-AIREN International Workshop. *Neurology*, 43, 250–260.

Satz, P. (1991). Brain reserve capacity on symptom onset after brain injury: a formulation and review of evidence of threshold theory. *Neuropsychology*, 7, 273–295.

Satz, P. *et al.* (1993). Low education as a possible risk factor for early cognitive abnormalities in HIV1: new findings from the Multicenter AIDS Cohort Study (MACS). *Journal of Acquired Immune Deficiency Syndromes*, 6, 503–511.

Stern, Y., Alexander, G. E., Prohovnik, I. and Mayeux, R. (1992). Inverse relationship between education and parietotemporal perfusion deficit in Alzheimer's disease. *Annals of Neurology*, 32, 371–375.

Teverovsky, L. *et al.* (2011). Brain extraction by deformable registration, Presented at the 17th Annual Meeting of the Ogranizagtion for Human Brain Mapping, Quebec City, June 26–30.

Urs, R. *et al.* (2009). Visual rating system for assessing magnetic resonance images: a tool in the diagnosis of mild cognitive impairment and Alzheimer disease. *Journal of Computer Assisted Tomography*, 33, 73–78.

Weir, D. R. *et al.* (2011). Reducing case ascertainment costs in US population studies of Alzheimer's disease, dementia, and cognitive impairment–part 1. *Alzheimer's and Dementia*, 7, 94–109.

https://doi.org/10.1017/S1041610212000191 Published online by Cambridge University Press