NeuroImage 66 (2013) 648–661

Contents lists available at SciVerse ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg

Unbiased tensor-based morphometry: Improved robustness and sample size estimates for Alzheimer's disease clinical trials

Xue Hua [a], Derrek P. Hibar [a], Christopher R.K. Ching [a], Christina P. Boyle [a], Priya Rajagopalan [a], Boris A. Gutman [a], Alex D. Leow [b,c], Arthur W. Toga [a], Clifford R. Jack Jr. [d], Danielle Harvey [e], Michael W. Weiner [f,g,h], Paul M. Thompson [a,i,*] and the Alzheimer's Disease Neuroimaging Initiative [1]

[a] Imaging Genetics Center, Laboratory of Neuro Imaging, Dept. of Neurology, UCLA School of Medicine, Los Angeles, CA, USA
[b] Dept. of Psychiatry, University of Illinois at Chicago, College of Medicine, Chicago, IL, USA
[c] Dept. of Bioengineering, University of Illinois at Chicago, Chicago, IL, USA
[d] Mayo Clinic, Rochester, MN, USA
[e] Dept. of Public Health Sciences, UC Davis School of Medicine, Davis, CA, USA
[f] Dept. of Radiology and Biomedical Imaging, UC San Francisco, San Francisco, CA, USA
[g] Dept. of Medicine, UC San Francisco, San Francisco, CA, USA
[h] Dept. of Psychiatry, UC San Francisco, San Francisco, CA, USA
[i] Dept. of Psychiatry, Semel Institute, UCLA School of Medicine, Los Angeles, CA, USA

ARTICLE INFO

Article history:
Accepted 30 October 2012
Available online 12 November 2012

Keywords:
Alzheimer's disease
Mild cognitive impairment
Aging
ADNI
Tensor-based morphometry
Drug trial

ABSTRACT

Various neuroimaging measures are being evaluated for tracking Alzheimer's disease (AD) progression in therapeutic trials, including measures of structural brain change based on repeated scanning of patients with magnetic resonance imaging (MRI). Methods to compute brain change must be robust to scan quality. Biases may arise if any scans are thrown out, as this can lead to the true changes being overestimated or underestimated. Here we analyzed the full MRI dataset from the first phase of Alzheimer's Disease Neuroimaging Initiative (ADNI-1) from the first phase of Alzheimer's Disease Neuroimaging Initiative (ADNI-1) and assessed several sources of bias that can arise when tracking brain changes with structural brain imaging methods, as part of a pipeline for tensor-based morphometry (TBM). In all healthy subjects who completed MRI scanning at screening, 6, 12, and 24 months, brain atrophy was essentially linear with no detectable bias in longitudinal measures. In power analyses for clinical trials based on these change measures, only 39 AD patients and 95 mild cognitive impairment (MCI) subjects were needed for a 24-month trial to detect a 25% reduction in the average rate of change using a two-sided test ($\alpha = 0.05$, *power* = 80%). Further sample size reductions were achieved by stratifying the data into Apolipoprotein E (ApoE) ε4 carriers versus non-carriers. We show how selective data exclusion affects sample size estimates, motivating an objective comparison of different analysis techniques based on statistical power and robustness. TBM is an unbiased, robust, high-throughput imaging surrogate marker for large, multi-site neuroimaging studies and clinical trials of AD and MCI.

Introduction

Alzheimer's disease (AD) affects 5.4 million people in the U.S. alone, and over 24 million people worldwide (Ferri et al., 2005). New treatments to slow or delay Alzheimer's disease progression must be rapidly and efficiently evaluated to alleviate a growing public health crisis. A related condition is mild cognitive impairment (MCI); people with MCI are at greatly increased risk of developing AD. As many as 10–25% of MCI subjects progress to probable AD per year (Petersen, 2000, 2003a, 2003b). Numerous therapeutic trials are underway to test novel compounds. Some of these trials use neuroimaging measures to assess treatment effects on brain measures, such as amyloid levels in the brain or rates of atrophy (Petersen, 2003a; Ross et al., 2012).

A wide variety of neuroimaging measures may be useful in tracking the progression of AD and MCI. The Alzheimer's Disease

Abbreviations: AD, Alzheimer's disease; ADNI, Alzheimer's Disease Neuroimaging Initiative; ApoE, Apolipoprotein E; BBSI, Brain boundary shift integral; FSL, FMRIB Software Library; MRI, Magnetic resonance imaging; MCI, Mild cognitive impairment; MDT, Minimal deformation target; MI, Mutual information; Quarc, Quantitative anatomical regional change; SPM, Statistical parametric mapping; stat-ROI, Statistical region-of-interest; TBM, Tensor-based morphometry; VBM, Voxel-based morphometry; TR, Repetition time; TE, Echo time; TI, Inversion time; QC, Quality control.

* Corresponding author at: Imaging Genetics Center, Laboratory of Neuro Imaging, Dept. of Neurology, UCLA School of Medicine, Neuroscience Research Building 225E, 635 Charles Young Drive, Los Angeles, CA 90095-1769, USA. Fax: +1 310 206 5518.
E-mail address: thompson@loni.ucla.edu (P.M. Thompson).

[1] Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.ucla.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data, but only some participated in the analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.ucla.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf.

1053-8119/$ – see front matter © 2012 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.neuroimage.2012.10.086

Neuroimaging Initiative (ADNI) was set up as one of several multicenter studies worldwide to develop and validate novel biomarkers to characterize, detect and track AD (Frisoni and Weiner, 2010; Mueller et al., 2005a, 2005b; Trojanowski et al., 2010; Weiner et al., 2010, 2012). In the first phase of ADNI (ADNI-1), 817 subjects received screening scans, including 188 early Alzheimer's patients, 400 subjects with MCI, and 229 healthy controls, who were studied at 6- or 12-month intervals for up to 36 months (Wyman et al., 2012). The entire dataset is publicly available (http://adni.loni.ucla.edu), offering a large test dataset to develop, validate, and compare biomarkers for disease classification and prognosis. A summary of approximately 200 published ADNI papers is provided in a recent review (Weiner et al., 2012).

High-resolution structural MRI is one of several imaging methods used to track AD, and numerous MRI-derived biomarkers have been thoroughly investigated, including but not limited to: (1) hippocampal volume (Jack et al., 1999, 2002; Morra et al., 2009a, 2009b; Schuff et al., 2009), (2) lateral ventricular volumes (Carmichael et al., 2006; Chou et al., 2008, 2009; Thompson et al., 2004), (3) gray matter volume or density, as measured using voxel-based morphometry (VBM) in the statistical parametric mapping (SPM) software package (Ashburner and Friston, 2000; Baron et al., 2001; Chetelat et al., 2005), (4) a measure of brain change over time known as the brain boundary shift integral (BBSI) (Fox et al., 2000; Freeborough and Fox, 1997), (5) automated methods for computing a variety of regional subvolumes, such as longitudinal FreeSurfer (Reuter et al., 2012), and the FMRIB Software Library (FSL) (Smith et al., 2007), (6) data-driven measures of temporal lobe atrophy using tensor-based morphometry (TBM) (Hua et al., 2009, 2010), and (7) measures of volume change in the entorhinal cortex, hippocampus, and whole brain using the commercial software known as quantitative anatomical regional change (Quarc) (Holland and Dale, 2011; Holland et al., 2009). Some of these methods also derive statistical *maps* of brain changes over time, as well as numeric summaries of atrophy from anatomically and statistically defined regions of interest. Earlier research applying pattern recognition and machine learning to medical image analysis has resulted in significant improvements in diagnostic accuracy and the specificity of AD imaging biomarkers (Davatzikos et al., 2008; Vemuri et al., 2008). In these studies, the goal was to create a tool to discriminate between diagnostic groups, rather than to optimize the efficiency of a biomarker. In other words, applying these algorithms to explicitly minimize required sample sizes will require modifications. These modifications will likely lead to greatly reduced sample size requirements in a clinical trial. Some evidence for this can be found in a recent paper by (Gutman et al., 2012), where numeric summaries are computed from signals weighted using linear discriminant analysis, and others, like Hobbs et al. (2010), which use a linear support vector machine classifier.

With several imaging biomarkers currently being considered for therapeutic trials to track brain degeneration (Cummings, 2010), different approaches need to be compared. Ideally, biomarkers would show excellent effect sizes for detecting longitudinal changes, avoid sources of bias, and not fail on a substantial fraction of the data, as a real clinical trial would not allow the selective exclusion of data (Fox et al., 2011).

In the current paper, we had 3 goals: first, to report improved and highly competitive sample size estimates for TBM, showing that no bias is present. Second, to develop and test several new efforts to improve the robustness of TBM, making the results robust to outliers in the data and poor quality scans. Third, to test whether standard enrichment methods – preferential selection of subjects based on Apolipoprotein E (ApoE) ε4 genotype or family history – could further reduce the required sample sizes when used in conjunction with the proposed improvements. We also studied the effect of selective data exclusion on the sample size estimates, suggesting that sample size estimates may be unduly optimistic if any removal of outliers is allowed.

Materials and methods

Overall design

We employed TBM to analyze the full ADNI-1 dataset, including all available 1.5 Tesla MR images scanned at screening, with follow-up scans at 6, 12, 18, 24, and 36 months ($N = 3314$), available for download at March 20, 2012. Numerical summaries were derived from a statistical region-of-interest (stat-ROI) inside the temporal lobes to quantify cumulative brain degeneration over time, and these were later used to compute sample size estimates for hypothetical clinical trials. We used a subgroup of healthy subjects with completed scan series at screening, 6, 12, and 24 months to assess whether our method was biased (in the sense of over- or under-estimating the true rate of change), and to confirm the biological plausibility of atrophy measures. We hypothesized that the healthy aging group would exhibit an essentially linear trend of minimal brain atrophy, with a zero intercept for the regression line fitted through all time points. We conducted power analyses to estimate sample size requirements for hypothetical clinical trials employing imaging outcome measures. We further tested the added effect of performing more standard drug trial enrichment strategies using ApoE status and family history of dementia. Finally, we conducted a simulation to demonstrate how sample size estimates were influenced by selective data removal, an effort that suggests reasonable recommendations for fair comparisons of methods in the future (cf. Wyman et al., 2012).

Alzheimer's Disease Neuroimaging Initiative

Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative database (adni.loni.ucla.edu). The ADNI was launched in 2003 by the National Institute on Aging (NIA), the National Institute of Biomedical Imaging and Bioengineering (NIBIB), the Food and Drug Administration (FDA), private pharmaceutical companies and non-profit organizations, as a $60 million, 5-year public–private partnership. The primary goal of ADNI has been to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment and early Alzheimer's disease. Determination of sensitive and specific markers of very early AD progression is intended to aid researchers and clinicians to develop new treatments and monitor their effectiveness, as well as lessen the time and cost of clinical trials.

The Principal Investigator of this initiative is Michael W. MD, VA Medical Center and University of California–San Francisco. ADNI is the result of efforts of many co-investigators from a broad range of academic institutions and private corporations, and subjects have been recruited from over 50 sites across the U.S. and Canada. The initial goal of ADNI was to recruit 800 adults, ages 55 to 90, to participate in the research, approximately 200 cognitively normal older individuals to be followed for 3 years, 400 people with MCI to be followed for 3 years and 200 people with early AD to be followed for 2 years. For up-to-date information, see www.adni-info.org.

MRI acquisition and image correction

All subjects were scanned with a standardized MRI protocol developed for ADNI (Jack et al., 2008). Briefly, high-resolution structural brain MRI scans were acquired at 59 ADNI sites using 1.5 Tesla MRI scanners (GE Healthcare, Philips Medical Systems, or Siemens). Using a sagittal 3D MP-RAGE scanning protocol, the typical acquisition parameters were repetition time (TR) of 2400 ms, minimum full echo time (TE) of 3 ms, inversion time (TI) of 1000 ms, flip angle of 8°, 24 cm field of view, $192 \times 192 \times 166$ acquisition matrix in the x-, y-, and z- dimensions, yielding a voxel size of $1.25 \times 1.25 \times 1.2$ mm^3, later reconstructed to 1 mm