Neurobiology of Aging 36 (2015) S69–S80

Contents lists available at ScienceDirect

Neurobiology of Aging

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

High-dimensional morphometry

Empowering imaging biomarkers of Alzheimer's disease

Boris A. Gutman [a], Yalin Wang [b], Igor Yanovsky [c], Xue Hua [a], Arthur W. Toga [h], Clifford R. Jack Jr [d], Michael W. Weiner [e,f,g], Paul M. Thompson [a,h,*], for the Alzheimer's Disease Neuroimaging Initiative [1]

[a] USC Imaging Genetics Center, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA
[b] School of Computing, Informatics, and Decision Systems Engineering, Arizona State University, Tempe, AZ, USA
[c] UCLA Joint Institute for Regional Earth System Science and Engineering, Los Angeles, CA, USA
[d] Mayo Clinic, Rochester, MN, USA
[e] Department of Radiology and Biomedical Imaging, UC San Francisco, San Francisco, CA, USA
[f] Department of Medicine, UC San Francisco, San Francisco, CA, USA
[g] Department of Psychiatry, UC San Francisco, San Francisco, CA, USA
[h] Institute for Neuroimaging and Informatics, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA

ARTICLE INFO

Article history:
Received 6 May 2013
Received in revised form 22 May 2014
Accepted 23 May 2014
Available online 27 August 2014

Keywords:
Linear discriminant analysis
Shape analysis
Tensor-based morphometry
ADNI
Lateral ventricles
Alzheimer's disease
Mild cognitive impairment
Biomarker
Drug trial
Machine learning

ABSTRACT

In a previous report, we proposed a method for combining multiple markers of atrophy caused by Alzheimer's disease into a single atrophy score that is more powerful than any one feature. We applied the method to expansion rates of the lateral ventricles, achieving the most powerful ventricular atrophy measure to date. Here, we expand our method's application to tensor-based morphometry measures. We also combine the volumetric tensor-based morphometry measures with previously computed ventricular surface measures into a combined atrophy score. We show that our atrophy scores are longitudinally unbiased with the intercept bias estimated at 2 orders of magnitude below the mean atrophy of control subjects at 1 year. Both approaches yield the most powerful biomarker of atrophy not only for ventricular measures but also for all published unbiased imaging measures to date. A 2-year trial using our measures requires only 31 (22, 43) Alzheimer's disease subjects or 56 (44, 64) subjects with mild cognitive impairment to detect 25% slowing in atrophy with 80% power and 95% confidence.

1. Introduction

Imaging biomarkers of Alzheimer's disease (AD) must offer sufficient power to detect brain atrophy in subjects scanned repeatedly over time (Cummings, 2010; Ross et al., 2012; Wyman et al., 2012). The expected cost of a drug trial may be prohibitively high, unless we can reasonably expect disease-slowing effects to be detected quickly enough and with reasonably few subjects. Imaging measures from standard structural magnetic resonance imaging (MRI) show considerable promise. Their use stems from the premise that longitudinal changes may be more precisely and reproducibly measured with MRI than comparable changes in clinical, cerebrospinal fluid (CSF), or proteomic assessments; clearly, whether that is true depends on the measures used. The use of MRI in a drug trial has some caveats; most MR studies from published drug trials have detected no effect or even a small, and possibly irrelevant but significant, increase in atrophy in the treatment group. Brain measures that are helpful for diagnosis, such as positron emission tomography (PET) scanning, may not be stable for large multicenter (N = several hundred) longitudinal trials that aim to slow disease progression. Other markers, such as CSF measures of amyloid and tau proteins to assess brain amyloid, may suffer the opposite problem of showing too little change during the clinical AD period. As a result, there is interest in testing the reproducibility of biomarkers, as well as methods to optimally combine them (Yuan et al., 2012).

Recent studies have tested the reproducibility and accuracy of a variety of MRI-derived measures of brain change. Several of these are highly correlated with clinical assessments and can predict

* Corresponding author at: Laboratory of Neuro Imaging, Department of Neurology, UCLA School of Medicine, Imaging Genetics Center, Neuroscience Research Building 225E, 635 Charles Young Drive, Los Angeles, CA 90095-1769, USA. Tel.: +1 310 206 2101; fax: +1 310 206 5518.
E-mail address: pthomp@usc.edu (P.M. Thompson).

[1] Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database.

0197-4580/$ – see front matter © 2015 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.neurobiolaging.2014.05.038

future decline on their own or in combination with other relevant measures. Although not the only important consideration, some analyses have assessed which MRI-based measures show greatest effect sizes for measuring brain change over time, while avoiding issues of bias and asymmetry that can complicate longitudinal image analysis (Fox et al., 2011; Holland et al., 2011; Hua et al., 2013), and while avoiding removing scans from the analysis that may lead to unfairly optimistic sample size estimates (Hua et al., 2013; Wyman et al., 2012). Promising MRI-based measures include the brain boundary shift integral (Leung et al., 2012; Schott et al., 2010), the ventricular boundary shift integral (Schott et al., 2010), and measures derived from anatomic segmentation software such as Quarc or FreeSurfer, some of which have been recently modified to handle longitudinal data more accurately (Fischl and Dale, 2000; Holland and Dale, 2011; Reuter et al., 2012; Smith et al., 2002).

Although several power estimates are possible, the analysis advocated by the Alzheimer's Disease Neuroimaging Initiative (ADNI) Biostatistics Core (Beckett, 2000) is to estimate the minimal sample size required to detect, with 80% power, a 25% reduction in the mean annual change, using a 2-sided test and standard significance level $\alpha = 0.05$ for a hypothetical 2-arm study (treatment vs. placebo). The estimate for the minimum sample size is computed from the formula below. $\hat{\beta}$ denotes the annual change (average across the group) and $\hat{\sigma}_D^2$ refers to the variance of the annual rate of change.

$$n = \frac{2\hat{\sigma}_D^2\left(z_{1-\alpha/2} + z_{power}\right)^2}{\left(0.25\hat{\beta}\right)^2} \quad (1)$$

Here, z_α is the value of the standard normal distribution for which $P[Z < z_\alpha] = \alpha$ the sample size required to achieve 80% power is commonly denoted by n80. Typical n80s for competitive methods are under 150 AD subjects and under 300 mild cognitive impairment (MCI) subjects; the larger numbers for MCI reflect the fact that brain changes tend to be slower in MCI than AD, and MCI is an etiologically more heterogeneous clinical category. For this reason, it is harder to detect a modification of changes that are inherently smaller, so greater sample sizes are needed to guarantee sufficient power to detect the slowing of disease.

Many algorithms can detect localized or diffuse changes in the brain, creating detailed 3D maps of changes (Avants et al., 2008; Leow et al., 2007; Shi et al., 2009), but the detail in the maps they produce is often disregarded when making sample size estimates according to Equation 1 as the formula expects a single univariate measure of change. In other words, it requires a single number or "numeric summary" to represent all the relevant changes occurring within the brain. To mitigate this problem, Hua et al. (2009) defined a "statistical ROI" based on a small sample of AD subjects by thresholding the *t*-statistic of each feature (voxel) and summing the relevant features over the ROI; this approach was initially advocated in the FDG-PET literature to home in on regions that show greatest effects (Chen et al., 2010). In spirit, the statistical ROI is a rudimentary supervised learning approach, as it finds regions that show detectable effects in a training sample and uses them to empower the analysis of future samples; the samples used are nonoverlapping and independent to avoid circularity. However, a simple threshold-based masking is known to potentially eliminate useful features as binarization loses a lot of the information present in continuous weights (Duda et al., 2001). Although many studies have used machine learning to predict the progression of neurodegenerative diseases and differentiate diagnostic groups such as AD, MCI, and controls (Kloppel et al., 2012; Kohannim et al., 2010; Vemuri et al., 2008), we found no attempts in the literature that used learning to directly optimize power to detect brain change.

To address this issue, we observed that minimizing Equation 1 is exactly analogous to one-class linear discriminant analysis (LDA). We applied the method to surface-based longitudinal expansion rates of the ventricular boundary (Gutman et al., 2013), achieving the lowest sample size estimates of any ventricle-based measure of AD to date, both in terms of absolute and control-adjusted atrophy. Here, we apply the LDA-based weighting to recently reported maps of whole brain volume change based on tensor-based morphometry (Hua et al., 2013). Further, we combine ventricular surface and tensor-based morphometry (TBM) volume measures into one combined atrophy score. Our results show a marked improvement over the stat-ROI approach, achieving substantively lower sample size estimates than any ADNI-based report to date.

2. Methods

2.1. Alzheimer's Disease Neuroimaging Initiative

Data used in the preparation of this article were obtained from the ADNI database (adni.loni.ucla.edu). The ADNI was launched in 2003 by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, the Food and Drug Administration, private pharmaceutical companies and nonprofit organizations as a $60 million, 5-year public-private partnership. The primary goal of ADNI has been to test whether serial MRI, PET, other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of MCI and early AD. 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. Weiner, MD, VA Medical Center and University of California, San Francisco. ADNI is the result of efforts of many coinvestigators from a broad range of academic institutions and private corporations, and subjects have been recruited from over 50 sites across the United States and Canada. The initial goal of ADNI was to recruit 800 adults, aged 55–90 years, 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.

Longitudinal brain MRI scans (1.5 Tesla) and associated study data (age, sex, diagnosis, genotype, and family history of AD) were downloaded from the ADNI public database (http://www.loni.ucla.edu/ADNI/Data/) on July 1, 2012. The first phase of ADNI, that is, ADNI-1, was a 5-year study launched in 2004 to develop longitudinal outcome measures of Alzheimer's progression using serial MRI, PET, biochemical changes in CSF, blood, and urine, and cognitive and neuropsychological assessments acquired at multiple sites similar to typical clinical trials.

All subjects underwent thorough clinical and cognitive assessment at the time of scan acquisition. All AD patients met NINCDS/ADRDA criteria for probable AD (McKhann et al., 1984). The ADNI protocol lists more detailed inclusion and exclusion criteria (Mueller et al., 2005a, 2005b), available online (http://www.alzheimers.org/clinicaltrials/fullrec.asp?PrimaryKey=208). The study was conducted according to the Good Clinical Practice guidelines, the Declaration of Helsinki and the United States, 21 CFR Part 50-Protection of Human Subjects and Part 56-Institutional Review Boards. Written informed consent was obtained from all