Alzheimer's & Dementia: Translational Research & Clinical Interventions 2 (2016) 177-181

Featured Article

Proof of concept demonstration of optimal composite MRI endpoints for clinical trials

Steven D. Edland[a,b,*], M. Colin Ard[b], Jaiashre Sridhar[c], Derin Cobia[d], Adam Martersteck[c,e], M.-Marsel Mesulam[c,f], Emily J. Rogalski[c]

[a]Division of Biostatistics, Department of Family Medicine & Public Health, University of California San Diego, La Jolla, CA, USA
[b]Department of Neurosciences, University of California San Diego, La Jolla, CA, USA
[c]Cognitive Neurology and Alzheimer's Disease Center, Northwestern University Feinberg School of Medicine, Chicago, IL, USA
[d]Department of Psychiatry and Behavioral Sciences, Northwestern University Feinberg School of Medicine, Chicago, IL, USA
[e]Department of Radiology, Northwestern University Feinberg School of Medicine, Chicago, IL, USA
[f]Department of Neurology, Northwestern University Feinberg School of Medicine, Chicago, IL, USA

Abstract

Introduction: Atrophy measures derived from structural MRI are promising outcome measures for early phase clinical trials, especially for rare diseases such as primary progressive aphasia (PPA), where the small available subject pool limits our ability to perform meaningfully powered trials with traditional cognitive and functional outcome measures.

Methods: We investigated a composite atrophy index in 26 PPA participants with longitudinal MRIs separated by 2 years. Rogalski *et al.* [5] previously demonstrated that atrophy of the left perisylvian temporal cortex (PSTC) is a highly sensitive measure of disease progression in this population and a promising endpoint for clinical trials. Using methods described by Ard *et al.* [1], we constructed a composite atrophy index composed of a weighted sum of volumetric measures of 10 regions of interest within the left perisylvian cortex using weights that maximize signal-to-noise and minimize sample size required of trials using the resulting score. Sample size required to detect a fixed percentage slowing in atrophy in a 2-year clinical trial with equal allocation of subjects across arms and 90% power was calculated for the PSTC and optimal composite surrogate biomarker endpoints.

Results: The optimal composite endpoint required 38% fewer subjects to detect the same percent slowing in atrophy than required by the left PSTC endpoint.

Conclusions: Optimal composites can increase the power of clinical trials and increase the probability that smaller trials are informative, an observation especially relevant for PPA but also for related neurodegenerative disorders including Alzheimer's disease.

Keywords: Primary progressive aphasia; PPA; Clinical trial; Sample size; Power calculations; Composite endpoint; Alzheimer's disease; Structural magnetic resonance imaging; MRI; Region of interest

1. Introduction

Clinical trials of interventions to slow the course of chronic progressive neurodegenerative diseases typically use cognitive neuropsychometric and functional (activities of daily living) outcome measures to demonstrate efficacy. Treatment efficacy is difficult to demonstrate with these endpoints, because decline is subtle during the relatively short span of observation of a clinical trial, and because there is substantial random variability in these measures from person to person and from observation to observation within a person. Volumetric magnetic resonance imaging (MRI) on the other hand has been shown to have good signal-to-noise

*Corresponding author. Tel: +1-848-246-1250; Fax: +1-858-246-1287.
E-mail address: sedland@ucsd.edu

http://dx.doi.org/10.1016/j.trci.2016.05.002

properties in this context. For example, for amnestic dementia of the Alzheimer type (AD), a substantial literature has consistently demonstrated that volumetric MRI endpoints could reduce required sample size in AD treatment trials and secondary prevention trials of mild cognitive impairment by 75% or more compared to a standard cognitive function outcome [1]. The need for efficient endpoints is especially critical for rare subtypes of disease where the pool of subjects available for recruitment limits our ability to even perform large-scale phase 3 trials using neuropsychometric and functional outcome measures.

Primary progressive aphasia (PPA) is a clinical dementia syndrome characterized by an initially isolated and progressive decline in language function and is associated with peak atrophy within the left hemisphere perisylvian language network [2–4]. Rogalski *et al.* [5] demonstrated that atrophy of the left perisylvian temporal cortex in particular is a highly sensitive measure of disease progression and a promising endpoint for clinical trials. Using this endpoint, clinical trials as small as 10 participants per arm would have 80% power to detect a 40% slowing of atrophy [5]. Efficiency of trials may be improved beyond these impressive levels by efficient utilization of the richness of data obtained by MRI. Xiong *et al.* [6] proposed that "composite" endpoints calculated as weighted averages of volumetric region of interest (ROI) substructures may outperform simple sums. These methods were operationalized by Ard *et al.* [7], who derived algorithms for determining optimal composite measures that maximize statistical power when used as an endpoint for clinical trials. In this brief communication, we demonstrate the potential utility of composite atrophy measures for clinical trials of neurodegenerative diseases with a prominent atrophy component.

2. Methods

Study subjects and imaging techniques have been previously described [5]. Briefly, study subjects included 26 individuals with a root diagnosis of PPA [2–4] (8 PPA logopenic, 10 PPA agrammatic, 8 PPA semantic). For the purposes of this article, the three clinical subtypes of PPA were combined to insure sufficient sample size to estimate parameters required for calculating weights. Hence, this analysis is best interpreted as a proof-of-concept demonstration of the potential utility of composite volumetric measures, rather than derivation of an endpoint appropriate for use in a clinical trial. Mean age at baseline was 63.7 years (SD = 6.7), 58% were women, mean Boston Naming Test score was 39.5 (SD = 20.9), and mean Western Aphasia Battery Aphasia Quotient score was 86.8 (SD = 8.0). All subjects received a baseline structural MRI and follow-up MRI approximately 2 years later (mean interval 2.0 years).

Structural MRIs were processed using the cross-sectional [8] and longitudinal [9] pipelines from FreeSurfer, version 5.1.0. Ten regions of interest (ROIs) within the left perisylvian temporal cortex region (Fig. 1) taken from the automated Desikan–Killiany cortical parcellation atlas were the components of a composite outcome measure [10]. The composite was calculated as the optimally weighted sum of these ROIs using weights that maximize the signal-to-noise ratio of rate of change on the composite, as previously described [7]. Clinical trail endpoints with high signal-to-noise ratio, also called the mean to standard deviation ratio (MSDR), are more sensitive to treatment effects and optimize the power of a trial. We used relative efficiency to compare the performance of different outcome measure, where relative efficiency is defined as the ratio of sample size required for trials using the respective outcomes calculated using the standard formula for a two-sample *t* test:

$$n/arm = 2\left(z_{1-\alpha/2} + z_{1-\beta}\right)^2 \sigma_d^2 \Big/ \Delta^2$$

where σ_d^2 is the within group variance of the outcome measure being compared across treatments, in this case, the change from baseline to 2-year follow-up, Δ is the treatment effect size under the alternative, and $z_{1-\alpha/2}$ and $z_{1-\beta}$ are the usual quantiles of the standard normal distribution, with α

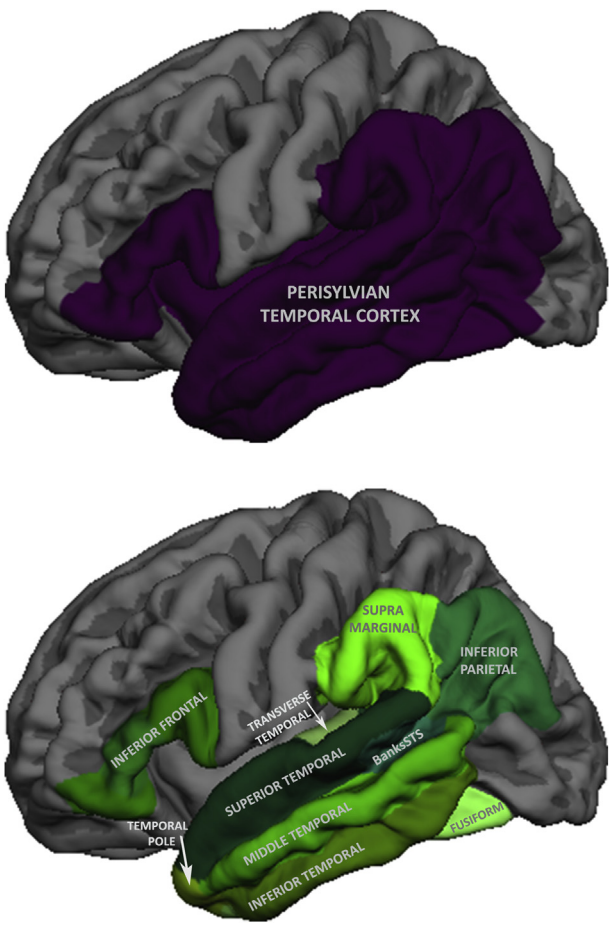

Fig. 1. Regions of interest used to examine longitudinal cortical atrophy in PPA. Top: The perisylvian temporal cortex region of interest defined in Rogalski *et al.* [5]. Bottom: the regions of interest used to create the composite outcome measure.