Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring 2 (2016) 113-122

Alzheimer's & Dementia

Cognitive & Behavioral Assessment

Novel verbal fluency scores and structural brain imaging for prediction of cognitive outcome in mild cognitive impairment

David Glenn Clark[a,b,*], Paula M. McLaughlin[c], Ellen Woo[d], Kristy Hwang[e], Sona Hurtz[f], Leslie Ramirez[f], Jennifer Eastman[g], Reshil-Marie Dukes[a], Puneet Kapur[h], Thomas P. DeRamus[i], Liana G. Apostolova[j]

[a]*Department of Neurology, Medical University of South Carolina, Charleston, SC, USA*
[b]*Department of Neurology, Ralph H. Johnson VA Medical Center, Charleston, SC, USA*
[c]*Ontario Neurodegenerative Disease Research Initiative, Schulich School of Medicine and Dentistry, Western University, London, Ontario, Canada*
[d]*Department of Neurology, David Geffen School of Medicine at UCLA, Los Angeles, CA, USA*
[e]*Oakland University William Beaumont School of Medicine, Rochester, MI, USA*
[f]*Drexel University College of Medicine, Philadelphia, PA, USA*
[g]*Northwestern University Feinberg School of Medicine, Chicago, IL, USA*
[h]*Department of Neurology, SUNY Upstate Medical University, Syracuse, NY, USA*
[i]*Department of Psychology, University of Alabama at Birmingham, Birmingham, AL, USA*
[j]*Department of Neurology, Indiana University, Indianapolis, IN, USA*

Abstract

Introduction: The objective of this study was to assess the utility of novel verbal fluency scores for predicting conversion from mild cognitive impairment (MCI) to clinical Alzheimer's disease (AD).
Method: Verbal fluency lists (animals, vegetables, F, A, and S) from 107 MCI patients and 51 cognitively normal controls were transcribed into electronic text files and automatically scored with traditional raw scores and five types of novel scores computed using methods from machine learning and natural language processing. Additional scores were derived from structural MRI scans: region of interest measures of hippocampal and ventricular volumes and gray matter scores derived from performing ICA on measures of cortical thickness. Over 4 years of follow-up, 24 MCI patients converted to AD. Using conversion as the outcome variable, ensemble classifiers were constructed by training classifiers on the individual groups of scores and then entering predictions from the primary classifiers into regularized logistic regression models. Receiver operating characteristic curves were plotted, and the area under the curve (AUC) was measured for classifiers trained with five groups of available variables.
Results: Classifiers trained with novel scores outperformed those trained with raw scores (AUC 0.872 vs 0.735; $P < .05$ by DeLong test). Addition of structural brain measurements did not improve performance based on novel scores alone.
Conclusion: The brevity and cost profile of verbal fluency tasks recommends their use for clinical decision making. The word lists generated are a rich source of information for predicting outcomes in MCI. Further work is needed to assess the utility of verbal fluency for early AD.

Keywords: Alzheimer's disease; Cognitive neuropsychology; Dementia; MCI (mild cognitive impairment); Machine learning; MRI (magnetic resonance imaging); Natural language processing

1. Introduction

Alzheimer's disease (AD) is a major socioeconomic crisis for the 20th century, with a projected 14 million cases by the year 2050 [1]. The dominant hypothesis for the

*Corresponding author. Tel.: +1-843-792-7446; Fax: +1-843-792-8626.
E-mail address: clarkda@musc.edu

http://dx.doi.org/10.1016/j.dadm.2016.02.001

pathogenesis of AD involves early deposition of beta-amyloid in the brain, but clinical trials targeting amyloid during the past decade have not met primary endpoints [2–4]. There is now evidence that beta-amyloid accumulates in the brain >10 years before the onset of cognitive symptoms [5,6]. Although cognitive symptoms appear late, studies in autosomal dominant AD suggest that individuals with mutations have measurable changes in cognition several years before onset of symptoms, when compared to mutation-free individuals [5]. These discoveries raise the concern that if treatments targeting beta-amyloid are to work, it might be necessary to implement them before the onset of symptoms. Two practical challenges arise. First, what is the best way to design clinical trials to ensure that pathophysiological changes of AD are occurring, if the individuals we would like to enroll are asymptomatic and cannot be expected to seek medical attention? Second, if a beta-amyloid targeted treatment is proven to work in the asymptomatic or early symptomatic stages of the disease, how can we identify individuals in the general population who will benefit from them?

There is an imminent need for new methods of detecting the earliest changes of AD, as the available biological methods are expensive, invasive, or entail exposure to radiation. Candidate methods under investigation include brief neuropsychological tests, ocular imaging, speech signal analysis, and computerized assessment of gait [7]. Structural MRI has been evaluated for this purpose, but the typical approach requires more than one image over a 6–12-month period, making it relatively expensive for a single patient [8]. Another approach is to use machine learning methods to train a classifier to discern between amyloid-positive and amyloid-negative individuals using available predictor variables, such as demographic data, structural brain imaging, cognitive tests, and blood tests. This approach has shown some success in a recent analysis of individuals with mild cognitive impairment (MCI—a condition thought to be a risk state for dementia [9,10]) who were studied in the AD Neuroimaging Initiative study [11]. The classifier achieved 0.78 area under the receiver operating characteristic curve (AUC) on a test set and 0.76 AUC when predicting conversion from MCI to AD.

The current work focuses on verbal fluency tasks—very brief cognitive measures in which the participant is given 1 minute to generate as many words as possible within a certain category, such as animals, or that start with a specific letter. The traditional method of scoring these tests is to simply count the number of unique, valid items in the list. The raw score obtained has proven clinical value [12–14], and as a result, verbal fluency tasks are performed in many research studies on AD or other cognitive disorders. However, there is strong evidence that careful examination of the words produced during the tasks may have additional clinical value, apart from or in addition to the raw score. The classic method for studying the explicit word content of these lists is to identify clusters of consecutively listed words that are related in some way (e.g., they have similar meaning, rhyme, or start with same two letters) [15–18]. The average length of these clusters is termed the clustering score and is thought to relate to spreading activation in a semantic or lexical network. The number of transitions between clusters is termed the switching score and is thought to relate to an individual's ability to deliberately change the subcategory of items that is currently being searched. Investigators have found that each of these scores has value for predicting dementia in longitudinal studies [19,20]. Some investigators have made use of unsupervised learning methods either on a corpus of fluency word lists [21] or on large English-language corpora [19,22] to improve prognostications.

In the present study, we develop new models for estimating risk of dementia conversion in MCI patients using measures derived from structural brain images and novel verbal fluency scores. In the statistical sub-discipline of machine learning, classification is often enhanced through expansion of the set of predictive features. This approach differs from the traditional inferential statistical approach, in which the primary goal is to identify statistically significant relationships between the independent and dependent variables. A potential weakness of the traditional approach is that one may develop a model with very poor predictive accuracy although it contains only statistically significant predictors. In machine learning, there is no immediate ambition to explain the relationship between the outcome and individual predictors—instead, the model is justified through the quality of predictions it yields. With this goal in mind, we developed a large set of novel predictive scores for five fluency tasks (categories animals and vegetables, and letters F, A, and S). Some of these novel predictive scores have roots in previous work (e.g., based on clustering, switching, or independent components analysis [ICA]), whereas others were developed specifically for this project. Some of the new scores are based on fundamental lexical qualities, such as syllable counts or frequencies of the words generated. Both of these quantities have good face validity and are easy to obtain, and the machine learning approach permits us to consider several possible ways of using them, such as (for a given fluency word list) taking the average, taking the sum, or subtracting the minimum value from the maximum value (i.e., metric range).

We based several novel scores on graph theory, a branch of discrete mathematics that provides techniques for analyzing networks. For this approach, we viewed the words in each list as nodes in a network and created weighted graphs by assigning numerical values to the edges or connections between the nodes. These values corresponded to the semantic, orthographic, or phonologic similarity between the two words being connected. Several scores were derived directly from these weighted graphs. The computation of other measures depended on conversion of each weighted graph into an unweighted graph by first identifying a threshold of the similarity metric and then creating a new