Alzheimer's & Dementia: Translational Research & Clinical Interventions 3 (2017) 360-366

Alzheimer's & Dementia

Featured Article

Power analysis to detect treatment effects in longitudinal clinical trials for Alzheimer's disease

Zhiyue Huang[a,*], Graciela Muniz-Terrera[b], Brian D. M. Tom[a], for the Alzheimer's Disease Neuroimaging Initiative[1]

[a]MRC Biostatistics Unit, University of Cambridge, UK
[b]Centre for Dementia Prevention, University of Edinburgh, UK

Abstract

Introduction: Assessing cognitive and functional changes at the early stage of Alzheimer's disease (AD) and detecting treatment effects in clinical trials for early AD are challenging.
Methods: Under the assumption that transformed versions of the Mini–Mental State Examination, the Clinical Dementia Rating Scale–Sum of Boxes, and the Alzheimer's Disease Assessment Scale–Cognitive Subscale tests'/components' scores are from a multivariate linear mixed-effects model, we calculated the sample sizes required to detect treatment effects on the annual rates of change in these three components in clinical trials for participants with mild cognitive impairment.
Results: Our results suggest that a large number of participants would be required to detect a clinically meaningful treatment effect in a population with preclinical or prodromal Alzheimer's disease. We found that the transformed Mini–Mental State Examination is more sensitive for detecting treatment effects in early AD than the transformed Clinical Dementia Rating Scale–Sum of Boxes and Alzheimer's Disease Assessment Scale–Cognitive Subscale. The use of optimal weights to construct powerful test statistics or sensitive composite scores/endpoints can reduce the required sample sizes needed for clinical trials.
Conclusion: Consideration of the multivariate/joint distribution of components' scores rather than the distribution of a single composite score when designing clinical trials can lead to an increase in power and reduced sample sizes for detecting treatment effects in clinical trials for early AD.
© 2017 The Authors. Published by Elsevier Inc. on behalf of the Alzheimer's Association. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

Keywords: Power analysis; Clinical trial; Sample size; Multivariate linear mixed-effects model; Composite score; Alzheimer's disease

1. Introduction

Much effort has been devoted to developing disease-modifying treatments that intervene in the pathobiologic processes involved in the early stage of Alzheimer's disease (AD). Any therapy that is effective at treating this early manifestation of the dementia process may provide an opportunity for managing the disease while patient function is relatively preserved [1]. Standard instruments used to quantify cognitive and functional decline in AD are relatively insensitive to the changes at early AD [2]. This raises challenges for assessing the early changes in cognition and function across the spectrum of AD [3] and makes detecting treatment effects in clinical trials for early AD even harder [2].

Power analysis is standard when designing clinical trials for detecting treatment effects. Ard *et al.* [4] provide a comprehensive review for clinical trials in AD. Misalignment of the power analysis can lead to possible errors in

[1]Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at http://adni.loni.usc.edu/wp-content/uploads/how_ to_ apply/ADNI_ Acknowledgement_ List.pdf.

*Corresponding author. Tel.: +44 (0)1223 330300; Fax: +44 (0)1223 330365.

E-mail address: robin.huang@mrc-bsu.cam.ac.uk

http://dx.doi.org/10.1016/j.trci.2017.04.007
2352-8737/ © 2017 The Authors. Published by Elsevier Inc. on behalf of the Alzheimer's Association. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

decisions regarding sample size. Too large samples may waste time, resources, and money and may unnecessarily expose some participants to inferior treatment if a treatment could have been shown to be more effective with fewer participants. Significant underestimation of the sample size may be a waste of time as it would unlikely lead to conclusive findings and therefore be unfair to all participants taking part in the trial. In this article, we are interested in the power/sample size to detect the treatment effects on the component scores in clinical trials for early AD.

In the literature of early AD, many researchers have used composite scores as single endpoints for performing power analysis [4]. A composite score is typically a linear combination of the scores of sensitive instruments. It provides a univariate summary of the component scores, avoids the multiple-hypothesis testing problem when each component score is considered separately, and reduces the impact of measurement error [5]. Furthermore, it may be more sensitive to the cognitive and functional decline than its separate components [6].

The construction of a composite score involves the selection and weighting of the component scores. Typically, the selection of the component scores may be based on a broad literature review regarding sensitivity to decline of candidate components [7], with equal weighting tending to be applied, possibly naively, to the chosen components. However, more statistically driven approaches can be used to derive the weights to construct more sensitive composite scores [2,6,8–12].

We therefore classify the statistical strategies used for the construction of a composite score into two major classes. The first is focused principally on selecting the most informative composite components and using prespecified weights not derived from statistical considerations; for example, Raghavan *et al.* [8] identify the informative component instruments based on standardized mean of 2-year change from baseline for a mild cognitive impairment (MCI) cohort and summed them to create a new composite score. The other is focused on "optimizing" the weights assigned to component scores based on an appropriate optimality criterion and is therefore more data driven; for example, some previous proposals find composite weights, which are sensitive to the clinical decline, by fitting linear mixed-effect models (LMMs) to the longitudinal composite scores [2,6,9]. Xiong *et al.* [6] propose composite weights that maximize the probability of observing a decline in one participant over a unit interval of time. Their weights can be considered as a special case of the composite weights proposed by Ard *et al.*, who use the power to detect the time effect in a clinical trial as their criterion and obtain the component weights by maximizing this criterion [2]. Ard *et al.*'s approach is applied to construct a composite atrophy index [9]. Another approach within this class is to base the estimation of the composite weights on a criterion that looks at the mean to standard deviation ratio of change over time [10,11]. Wang *et al.* [12] propose another composite score construct by using a linear clinical decline equation to select and reweight the component scores simultaneously.

In general, using composite scores as single endpoints may lose information to detect the changes in components [3]; for example, a large change in one component can be masked by small changes on other component scores. Data-driven composite scores have been further criticized [7]. Firstly, they may lose clinical interpretation. It is possible that a clinically meaningful component score has small weights in a data-driven composite score [7]. In addition, they may not be consistent across different data sets. Donohue *et al.* [7] apply cross-validation to quantify the out-of-sample performance of optimal composite scores and conclude that the overall performance of the optimal composite scores is worse than those composite scores derived without optimization.

A limited amount of the literature in AD has considered power analysis with multiple endpoints, although multiple endpoints are commonplace in AD. Under the assumption that the component scores are jointly from a multivariate linear mixed-effects model (MLMM), we compare three approaches with regard to their power to detect the treatment effects on component scores. Two of them are with multiple endpoints, whereas the other is with a single-composite endpoint.

2. Methods

2.1. MLMM for component scores

Mixed-effect models are from a class of useful statistical models for analyzing longitudinal data [13]. They allow a subset of the regression parameters (random effects) to vary randomly between participants and thereby characterize the natural heterogeneity in the target population in these parameters. Fixed effects are used to refer to the regression parameters, which are fixed but unknown and need to be estimated.

Assuming that all possible covariates are balanced (as would be assumed in a clinical trial through randomization), we model the component scores using an MLMM with a random intercept, fixed time, and time by treatment interaction effects. (The addition of further covariates can be easily incorporated if deemed necessary.) Such a model is able to simultaneously characterize the correlations between the component scores at each time t and the correlations across time for each component score.

Let Y_{ntj} be the j-th component score of the n-th participant at visit time t, where $n = 1,\ldots,N$, $t = 1,\ldots,T_n$, and $j = 1,\ldots,J$. Here, the number of visits T_n is a positive integer depending on the n-th participant, and the number of component scores J is prespecified. We use a linear function to link the component scores with the mixed effects

$$Y_{ntj} = \beta_{j0} + \gamma_j \times (\text{Treatment} \times \text{Time}) + \beta_{j2} \times \text{Time} + b_{nj} + \varepsilon_{ntj},$$

where γ_j is the j-th component treatment effect, b_{nj} is the random intercept that is unique to the j-th component score

Download English Version:

https://daneshyari.com/en/article/8680560

Download Persian Version:

https://daneshyari.com/article/8680560

Daneshyari.com