ELSEVIER

Alzheimer's & Dementia: Translational Research & Clinical Interventions 1 (2015) 63-71

Alzheimer's & Dementia

Effect of sample size re-estimation in adaptive clinical trials for Alzheimer's disease and mild cognitive impairment

Guoqiao Wang[a,*], Richard E. Kennedy[b], Gary R. Cutter[a], Lon S. Schneider[c]

[a]*Department of Biostatistics, University of Alabama-Birmingham, Birmingham, AL, USA*
[b]*Department of Medicine, University of Alabama-Birmingham, Birmingham, AL, USA*
[c]*Departments of Psychiatry and Neurology, University of Southern California Keck School of Medicine, Los Angeles, CA, USA*

Abstract

Introduction: The sample size re-estimation (SSR) adaptive design allows interim analyses and resultant modifications of the ongoing trial to preserve or increase power. We investigated the applicability of SSR in Alzheimer's disease (AD) trials using a meta-database of clinical studies.
Methods: Based on six studies, we simulated clinical trials using Alzheimer's Disease Assessment Scale-cognitive subscale (ADAS-Cog) as primary outcome. A single SSR based on effect sizes or based on variances was conducted at 6 months and 12 months. Resultant power improvement and sample size adjustments were evaluated.
Results: SSR resulted in highly variable outcomes for both sample size increases and power improvement. The gain in power after SSR varies by initial sample sizes, trial durations, and effect sizes.
Conclusions: SSR adaptive designs can be effective for trials in AD and mild cognitive impairment with small or medium initial sample sizes. However, SSR in larger trials (>200 subjects per arm) generates no major advantages over the typical randomized trials.

Keywords: Alzheimer's disease; Alzheimer's Disease Assessment Scale; Mild cognitive impairment; Sample size re-estimation; Adaptive design

1. Introduction

The number of individuals with Alzheimer's disease (AD) continues to grow worldwide with the aging of the population [1]. Although a handful of modestly effective symptomatic treatments have been developed using the typical randomized clinical trial (RCT) design, clinical trials to identify effective disease-modifying treatments have been uniformly negative [2]. There are several potential causes of these negative trials, including the lack of efficacy in the treatments, insensitivity of the primary outcome to treatment changes, and low power due to the inaccurate pretrial estimates of the treatment effect. Therefore, clinical trial designs allowing interim analyses and the resultant modification of the ongoing trial to increase power have been recommended [3]. One such approach is the sample size re-estimation (SSR) adaptive design, which allows sample size adjustment based on the comparison between the interim treatment effect (or the interim variance) to the pretrial treatment effect (or the pretrial variance) [4].

The typical RCT design starts with a prespecified sample size, and modifications would not be allowed after the trial has started. In the absence of dropouts, the trial would end with the same sample size as specified at the beginning. The SSR adaptive design allows the sample size to increase when the pretrial treatment effect size was overestimated or the pretrial variance of the outcome was underestimated, leading to a trial that concludes using a larger sample size to retain the power specified at the beginning. It can allow early stopping or an overall decrease in the sample size when the pretrial treatment effect size was underestimated or the pretrial variance was overestimated, leading to a trial with the prespecified power but a smaller sample size,

*Corresponding author. Tel.: +1-205-826-8967; Fax: +1-205-975-2541.
E-mail address: guoqiao@uab.edu

http://dx.doi.org/10.1016/j.trci.2015.03.002

although this is rare. This flexibility not only improves efficacy, but also provides other advantages over the RCT design, such as minimizing the number of patients exposed to inferior treatments, avoiding long-term trials for drugs with limited efficacy, and better using the most recent external or internal information of the ongoing trial. Potential concerns about using SSR include the reliability in estimating the overall treatment effect based on relatively small interim samples (or, for longitudinal trials, the precision in predicting the final treatment effect using only the early measurements), and the trade-off between the gain in power versus the burden to recruit more subjects. The former concern is particularly relevant for AD trials, as heterogeneity in the course of the disease may introduce significant inaccuracies in estimating the final treatment effect based on interim analyses. This study used simulations based on real patient data to investigate the SSR adaptive designs for AD trials.

2. Methods

2.1. Study overview

Of the 19 studies in our meta-database [5], we excluded seven studies that did not collect Alzheimer's Disease Assessment Scale-cognitive subscale (ADAS-Cog) data, four trials with duration less than 18 months, one trial enrolling only normal subjects, and one trial enrolling only moderate AD, yielding six studies that were used for the simulation (Supplementary Table 1). The primary outcome was the ADAS-Cog, which evaluates memory, reasoning, orientation, praxis, language, and word finding difficulty, and is scored from 0 to 70 errors, with higher scores indicating greater impairment [6]. Clinical assessments were obtained at 6-month intervals over the first 2 years.

2.2. Simulation methods

Simulations were conducted under a detailed protocol [7], similar to our previously published approach [5,8], to reflect clinical trials for an experimental drug for AD or MCI with one treatment group and one placebo group, 1:1 allocation ratio, and parameters for the distribution of ADAS-Cog were selected to be consistent with previously published trials and Alzheimer's Disease Neuroimaging Initiative (ADNI) ADNI [9,10].

Clinical trials with sample sizes of 50, 100, 200, 300, and 400 per group, trial durations of 12 months or 18 months for AD and of 18 months or 24 months for MCI, and dropout rates of 20% or 40% in both groups, were simulated. For each scenario, subjects were randomly selected from the meta-database with replacement, i.e., subjects from the data set could be present more than once in the same or different treatment groups. The placebo group outcome was the score for the subject at the specified time point in the meta-database, with normally distributed random error with mean 0 and standard deviation 1 added to minimize ties in the outcome. For each subject in the treatment group, effect sizes of 0.15 and 0.25 (representing treatment effects of small to medium size) were used to compute simulated treatment results. The individual treatment effect was randomly generated from a χ^2 distribution with a mean equal to the expected treatment effect (effect size times the pooled group standard deviation) to allow for a more realistic distribution of declines over time, where a few patients may fail or worsen more markedly than would be predicted by a normal distribution. This method introduced some extreme measurements but not to the extent of violating the homoscedasticity assumption of the analysis models. As successful treatments would lead to smaller increases on the ADAS-Cog than placebo, the individual treatment effect was shifted by subtracting two times the expected treatment effect, then adding the resultant to the patient's score at the specified time point in the database. For example, if a is the ADAS-Cog score at a given time point, then $a+\chi_z^2-2*z$, is the corresponding score in the simulated treatment group, where $z=effect\ size * SD$ and SD is the sample standard deviation of the change in ADAS-Cog from baseline. In this example, if a = 24, effect size is 0.25, SD is 8, and the randomly generated treatment effect from the χ_z^2 is 3, then the ADAS-Cog score in the simulation would be 23. With this added treatment effect, the mean difference in ADAS-Cog between treatment arms and its standard deviation increased over time (Supplementary Table 2).

2.3. Time points used for SSR

Patients' enrolment times vary in a typical trial, leading to different numbers of available measurements for each patient at the interim analysis. In this study, for trials with given initial sample sizes, "SSR at 12 months" means that all the patients had enrolled and had been measured for up to at least 12 months. We truncated the follow-up at the specified time point so our results would not depend on the recruitment rate.

2.4. Estimation methods used for SSR

SSR based on interim variances (henceforth, referred as "variance only method") and SSR based on interim effect sizes (henceforth, referred as "effect size method") were used, and both methods assumed equal variances for both treatment arms. The "variance only method" assumes that the pretrial estimate of the mean difference between treatment arms is accurate, and only the variance is uncertain and needs re-estimation. At the interim analysis, the variance of ADAS-Cog was estimated and compared with the pretrial estimate, and then the sample size was adjusted based on the following equation [11]:

$$N=\frac{\widehat{\sigma}_i^2}{\widehat{\sigma}_0^2}N_0,$$

where N is the re-estimated sample size, N_0 is the initial sample size, and $\widehat{\sigma}_i^2$ and $\widehat{\sigma}_0^2$ are the interim and the estimated