Published in final edited form as:
Comput Stat Data Anal. 2014 January 1; 69: 101–113. doi:10.1016/j.csda.2013.07.013.

Optimizing parameters in clinical trials with a randomized start or withdrawal design

Chengjie Xiong[a,b,*], **Jingqin Luo**[a], **Feng Gao**[a], and **John C. Morris**[b,c]

[a]Division of Biostatistics, Washington University, St. Louis, USA

[b]Knight Alzheimer's Disease Research Center, Washington University, St. Louis, MO

[c]Department of Neurology, Pathology and Immunology, Washington University, St. Louis, MO

Abstract

Disease-modifying (DM) trials on chronic diseases such as Alzheimer's disease (AD) require a randomized start or withdrawal design. The analysis and optimization of such trials remain poorly understood, even for the simplest scenario in which only three repeated efficacy assessments are planned for each subject: one at the baseline, one at the end of the trial, and the other at the time when the treatments are switched. Under the assumption that the repeated measures across subjects follow a trivariate distribution whose mean and covariance matrix exist, the DM efficacy hypothesis is formulated by comparing the change of efficacy outcome between treatment arms with and without a treatment switch. Using a minimax criterion, a methodology is developed to optimally determine the sample size allocations to individual treatment arms as well as the optimum time when treatments are switched. The sensitivity of the optimum designs with respect to various model parameters is further assessed. An intersection-union test (IUT) is proposed to test the DM hypothesis, and determine the asymptotic size and the power of the IUT. Finally, the proposed methodology is demonstrated by using reported statistics on the placebo arms from several recently published symptomatic trials on AD to estimate necessary parameters and then deriving the optimum sample sizes and the time of treatment switch for future DM trials on AD.

Keywords

Alzheimer's disease; Disease-modifying trials; Intersection-union test; Minimax criterion; Random intercept and slope models; Randomized start design

1. Introduction

Many clinical trials require a switch of treatments in the middle of the studies, and yet are not conducted through a standard crossover design. Examples include disease-modifying (DM) trials on chronic diseases such as Alzheimer's disease (AD) that strive not only to establish the symptomatic efficacy for novel treatments in improving cognition, function, and global measures or deferring decline over a relatively short period of time, but also to

*Corresponding Author: Chengjie Xiong, PhD, Division of Biostatistics, Washington University, St. Louis, MO 63110, (314) 362 3635 (Office), (314) 362 2693 FAX, chengjie@wubios.wustl.edu.

demonstrate that the course of AD has been altered and the rate of disease progression has been slowed (Cummings 2006, Aisen 2006, Citron 2004, Mani 2004) over a relatively long period of time. DM trials have been widely discussed in the AD research community (Leber 1997; Sampaio 2006; Whitehouse *et al.* 1998, Cummings *et al.* 2007). Clinicians have conceptualized designs of DM trials to allow a definite distinction from the symptomatic trials (Cummings 2006, Aisen 2006, Citron 2004). These designs in general require a switch of treatments in the middle of trials for at least a proportion of subjects originally randomized to either placebo or active treatment. One such design is the randomized start design (Mani 2004). All patients in the design eventually will receive the active treatment, but are randomized to two treatment groups that begin the active drug at different times. During the initial time period of the study one group receives active drug and the other placebo. After an interval of time sufficient to demonstrate the symptomatic efficacy for the active drug, the placebo group switches to the active drug. If the patients who begin active drug late 'catch up' with those who begin the active drug at baseline, the treatment effect is assumed to be symptomatic. If there is no 'catch-up', it is assumed that the effect of the drug is DM. Another DM design is the randomized withdrawal design (Mani 2004), which differs from the randomized start design only in that subjects initially on active drug are switched to placebo in the second phase. If the group withdrawn from the active drug maintains gains on the efficacy measure relative to the placebo group, it is assumed that the drug is DM. In both DM designs, in order to preserve the blinding, a second randomization must be performed to the initial placebo (or active drug) arm so that a proportion of patients will maintain on placebo (or active drug) throughout the trial. Figure 1 presents the expected longitudinal cognitive growth profiles of a randomized start design on AD. To date, all FDA-approved treatments to AD have been entirely based on their efficacy for treating symptoms (Kryscio *et al.* 2004, Ringman *et al.* 2009, Andrieu *et al.* 2006), partly because the design and analyses to establish the DM effectiveness of these treatments as well as many emerging ones (Salloway *et al.* 2008) have not been fully established.

This paper focus on the DM trial with a randomized start or withdrawal design for which only three repeated efficacy assessments are planned: one at the baseline, one at the end of the trial, and the other at the time when the treatments are switched. We formulate the DM hypothesis and propose a method to test the hypothesis, derive optimal sample size allocations to different treatment arms and the optimum time of treatment switch, and assess the sensitivity of these designs. Finally, we present optimum designs of future DM trials on AD using the reported statistics from recently published symptomatic trials.

2. A model for DM trials with a randomized start design

2.1. Estimating DM efficacy and testing the DM hypothesis

We will discuss our methodology by focusing on the DM trials on AD with a randomized start design, and point out that our proposed methods apply to many similar clinical trials on other diseases that require a switch of treatments in the middle of the trial (i.e., randomized withdrawal design). We use Y to denote the primary efficacy outcome in DM trials on AD (Cummings 2008). The Alzheimer's Disease Assessment Scale-cognitive subscale (ADAS-cog, Rosen *et al.* 1984) has been the most commonly used primary cognitive outcome in symptomatic trials of AD. Recently, many modalities of biomarkers have now shown promising ability to track the disease progression, including magnetic resonance imaging-based brain volumes (Storandt *et al.* 2009), diffusion tensor imaging-based measures of white matter microstructure (Head *et al.* 2004), cerebrospinal fluid (Fagan *et al.* 2006), and molecular imaging of cerebral fibrillar amyloid with positron emission tomography using the [^{11}C] benzothiazole tracer, Pittsburgh Compound-B (Mintun *et al.* 2006).

We use *u=tt* and *pp* to represent the group of subjects who are in the treatment arm and placebo arm throughout the trial, respectively, and use *u=pt* to represent the group of subjects who initially receive the placebo and then switch to the active treatment. Assume that $y_j^u(t)$ is the observation of Y at time t for the j-th subject from treatment group u. Let t_1, t_2, t_3 be the three time points for repeated measures of Y in a DM trial with a randomized start design, where $t_1 = 0$ is the baseline, t_3 is the time of the final assessment, and t_2 is the time for the treatment switch. Let $y_{ji}^u = y_j^u(t_i)$, $i = 1,2,3$, and let $Y_j^u = (y_{j1}^u, y_{j2}^u, y_{j3}^u)^t$ (superscript t means matrix transpose) be the vector of longitudinal measurements of the j-th subject from treatment group u. Because the DM design requires the establishment of both symptomatic efficacy prior to the treatment switch and the DM efficacy after the treatment switch, the first objective is to compare the change from the baseline between the treatment and the placebo before the treatment switch, and the second objective is to compare the change in efficacy outcome between subjects treated throughout the trial and those receiving the delayed treatment.

For subject j from treatment arm *u, u=tt, pp,* and *pt*, we assume a trivariate distribution for $Y_j^u = (y_{j1}^u, y_{j2}^u, y_{j3}^u)^t$ whose first and second moments exist. Notice that here we do not assume a specific parametric family such as normal distributions for the joint distribution of Y_j^u. This more general assumption on the distributions is justified, especially in the future clinical trials on AD for which a wide range of cognitive and CSF and imaging biomarkers (Storandt *et al.* 2009, Head *et al.* 2004, Fagan *et al.* 2006, Mintun *et al.* 2006) with highly skewed distributions could potentially serve as efficacy endpoints. Let $\mu^u = (\mu_1^u, \mu_2^u, \mu_3^u)^t$ be the mean vector of Y_j^u for group u. We assume that the duration of the trial, t_3, is pre-specified, but the time for the treatment switch, t_2, can be designed. Let

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} \\ & \sigma_2^2 & \sigma_{23} \\ & & \sigma_3^2 \end{pmatrix} \quad (1)$$

be the shared covariance matrix of Y_j^u for *u=tt, pp,* and *pt*. We point out that our proposed methods described below can also be generalized to the case when the variance and covariance parameters in Σ vary as a function of t_2. Because of the randomization at the baseline as well as the second randomization at the treatment switch time for those initially assigned to the placebo, we can assume that $\mu_1^{pp} = \mu_1^{tt} = \mu_1^{pt}$ and $\mu_2^{pp} = \mu_2^{pt}$. Further, we assume that a linear trend adequately describes the longitudinal progression of the efficacy outcome for the treated group and the placebo group throughout the trial, and let

$$\beta^u = \frac{\mu_2^u - \mu_1^u}{t_2 - t_1} = \frac{\mu_3^u - \mu_2^u}{t_3 - t_2}$$

be the mean slopes for *u=tt* and *pp*. Let

$$\beta^{pt} = \frac{\mu_3^{pt} - \mu_2^{pt}}{t_3 - t_2}$$

be the slope of efficacy outcome after the treatment switch. The comparative nature of $\mu^u = (\mu_1^u, \mu_2^u, \mu_3^u)^t$, *u=tt, pp,* and *pt*, determines whether the novel treatment is DM. More specifically, before the treatment switch, it is expected that the symptomatic efficacy for

treating AD will be established. This implies that the mean decline since baseline for the treated and the placebo arms should be different, with the treated arm decline more slowly before the time when the placebo is switched to the active treatment. This implies that $\mu_2^{tt} - \mu_1^{tt} > \mu_2^{pp} - \mu_1^{pp}$, i.e., $\beta^{tt} > \beta^{pp}$. After the treatment switch, the efficacy for modifying the disease can only be established by the fact that the subjects whose treatment is delayed (i.e., *u=pt*) can never 'catch up' those whose treatment is not delayed if the trial continues indefinitely. Mathematically, this implies that, after the treatment switch, the rate of decline for the delayed treatment arm has to be equal to or faster than that for the treated arm throughout the trial, i.e., $(\mu_3^{tt} - \mu_2^{tt})/(t_3 - t_2) \geq (\mu_3^{pt} - \mu_2^{pt})/(t_3 - t_2)$, or equivalently, $\beta^{tt} \geq \beta^{pt}$. Therefore, an appropriate statistical hypothesis for establishing the DM efficacy of the novel treatment is $\beta^{tt} \geq \beta^{pt} > \beta^{pp}$. In order to test this hypothesis, we have to estimate two differences on slopes: one is between the treated arm throughout the trial and the delayed treatment arm, i.e., $\delta = \beta^{tt} - \beta^{pt}$, and the other is between the delayed treatment arm and the placebo arm throughout the trial, i.e., $\Delta = \beta^{pt} - \beta^{pp}$.

Let n_u be the sample size within group *u*. Let $n = n_{tt} + n_{pp} + n_{pt}$ be the total sample size. Let $\lambda_u = n_u / n$ be the proportion (i.e., allocation) of the total sample size to each treatment group *u=pp, tt,* and *pt*. It is clear that $\lambda_{pp} + \lambda_{tt} + \lambda_{pt} = 1$. Let $d_1 = t_2 - t_1$, $d_2 = t_3 - t_2$, $s_1 = \sigma_1^2 - 2\sigma_{12}$, $s_2 = \sigma_2^2 - 2\sigma_{23}$, $s_3 = \sigma_{13} - \sigma_{12}$, $s_4 = \sigma_2^2 + \sigma_3^2 - 2\sigma_{23}$. Let

$$\hat{\mu}_1^{tt} = \hat{\mu}_1^{pp} = \hat{\mu}_1^{pt} = \frac{\sum_{j=1}^{n_{tt}} y_{j1}^{tt} + \sum_{j=1}^{n_{pp}} y_{j1}^{pp} + \sum_{j=1}^{n_{pt}} y_{j1}^{pt}}{n}.$$

Because β^{tt} can be estimated by two unbiased estimators

$$\hat{\beta}_1^{tt} = \frac{\sum_{j=1}^{n_{tt}} y_{j2}^{tt}/n_{tt} - \hat{\mu}_1^{tt}}{d_1},$$

and

$$\hat{\beta}_2^{tt} = \frac{\sum_{j=1}^{n_{tt}} (y_{j3}^{tt} - y_{j2}^{tt})}{d_2 n_{tt}},$$

an optimum estimate to β^{tt} is

$$\hat{\beta}^{tt} = c\hat{\beta}_1^{tt} + (1-c)\hat{\beta}_2^{tt}, \quad (2)$$

for some constant *c* such that the variance of $\hat{\beta}^{tt}$, $\sigma^2_{\hat{\beta}^{tt}}$, as given in row 2 of Table 1, is minimized over the choice of *c*. A simple calculus reveals the optimum *c* as given by

$$c = \frac{s_4 d_1^2/\lambda_{tt} - [(\sigma_{23} - \sigma_2^2)/\lambda_{tt} - s_3] d_1 d_2}{[\sigma_2^2/\lambda_{tt} + s_1] d_2^2 + s_4 d_1^2/\lambda_{tt} - 2[(\sigma_{23} - \sigma_2^2)/\lambda_{tt} - s_3] d_1 d_2}.$$

Let

$$\hat{\mu}_2^{pp}=\hat{\mu}_2^{pt}=\frac{\sum_{j=1}^{n_{pp}} y_{j2}^{pp}+\sum_{j=1}^{n_{pt}} y_{j2}^{pt}}{n_{pp}+n_{pt}}.$$

β^{pp} can also be estimated by two unbiased estimators. One involves data from $t_1 = 0$ to t_2 for both the placebo arm throughout the trial and the delayed treatment arm,

$$\hat{\beta}_1^{pp}=\frac{\hat{\mu}_2^{pp}-\hat{\mu}_1^{pp}}{d_1}.$$

The other involves data from t_2 to t_3 for the placebo arm throughout the trial,

$$\hat{\beta}_2^{pp}=\frac{\sum_{j=1}^{n_{pp}} y_{j3}^{pp}/n_{pp}-\hat{\mu}_2^{pp}}{d_2}.$$

An optimum estimate to β^{pp} is

$$\hat{\beta}^{pp}=f\hat{\beta}_1^{pp}+(1-f)\hat{\beta}_2^{pp}, \quad (3)$$

for some constant f such that the variance of $\hat{\beta}^{pp}$, $\sigma^2_{\hat{\beta}^{pp}}$, as given in row 3 of Table 1, is minimized over the choice of f. Another simple calculus reveals the optimum f as

$$f=\frac{[(\sigma_3^2/\lambda_{pp}+s_2/(\lambda_{pp}+\lambda_{pt})]d_1^2-[(\sigma_{23}-\sigma_2^2)/(\lambda_{pp}+\lambda_{pt})-s_3]d_1d_2}{[\sigma_2^2d_2^2+s_2d_1^2-2(\sigma_{23}-\sigma_2^2)d_1d_2]/(\lambda_{pt}+\lambda_{pp})+[s_1d_2^2+2s_3d_1d_2]+\sigma_3^2d_1^2/\lambda_{pp}}.$$

Next, the slope β^{pt} after the treatment switch can be estimated by

$$\hat{\beta}^{pt}=\frac{\sum_{j=1}^{n_{pt}} y_{j3}^{pt}/n_{pt}-\hat{\mu}_2^{pt}}{d_2}.$$

The variance of $\hat{\beta}^{pt}$, $\sigma^2_{\hat{\beta}^{pt}}$, is given by row 5 of Table 1. Finally, based on the optimum estimator $\hat{\beta}^{tt}$, $\delta = \beta^{tt} - \beta^{pt}$ can be estimated by $\hat{\delta} = \hat{\beta}^{tt} - \hat{\beta}^{pt}$. The variance of $\hat{\delta}$, $\sigma^2_{\hat{\delta}}$, is given by row 8 of Table 1, where the covariance of $\hat{\beta}^{tt}$ and $\hat{\beta}^{pt}$, cov($\hat{\beta}^{tt}$, $\hat{\beta}^{pt}$), is given by row 6 of Table 1. Likewise, based on the optimum estimator $\hat{\beta}^{pp}$, $\Delta = \beta^{pt} - \beta^{pp}$ can be estimated by $\hat{\Delta} = \hat{\beta}^{pt} - \hat{\beta}^{pp}$. The variance of $\hat{\Delta}$, $\sigma^2_{\hat{\Delta}}$, is given by row 9 of Table 1, where cov($\hat{\beta}^{pt}$, $\hat{\beta}^{pp}$) is given by row 7 of Table 1. It is clear that under standard regularity conditions required for the joint distribution of Y_j^u in the Central Limit Theorem, the estimates, ($\hat{\delta}$,$\hat{\Delta}$), follow an asymptotic bivariate normal distribution with mean (δ,Δ) and covariance matrix

$$\Psi=\begin{pmatrix} \sigma^2_{\hat{\delta}} & \sigma_{\hat{\delta}\hat{\Delta}} \\ & \sigma^2_{\hat{\Delta}} \end{pmatrix},$$

where the covariance of $(\hat{\delta},\hat{\Delta})$, $\sigma_{\hat{\delta}\hat{\Delta}}$, is given by row 10 of Table 1 with the covariance of $\hat{\beta}^{tt}$ and $\hat{\beta}^{pp}$, cov($\hat{\beta}^{tt}$, $\hat{\beta}^{pp}$), given by row 4 of Table 1.

Table 1 summarizes the variance and covariance of all estimated parameters for testing DM efficacy as described above. Notice that the covariance in the third column of Table 1 becomes the variance when two estimates are the same.

To test the DM efficacy of the active treatment, we propose to test the null hypothesis $H_0 : \delta < 0$ or $\Delta \le 0$ against the alternative $H_1 : \delta \ge 0$ and $\Delta > 0$. The null hypothesis is the union of two null hypotheses $H_{0\delta} : \delta < 0$ and $H_{0\Delta} : \Delta \le 0$, and the alternative is the intersection of two alternative hypotheses $H_{1\delta} : \delta \ge 0$ and $H_{1\Delta} : \Delta > 0$. For each individual set of null and alternative hypotheses, let $z^{\delta} = \hat{\delta} / \sigma_{\hat{\delta}}$ and $z^{\Delta} = \hat{\Delta} / \sigma_{\hat{\Delta}}$ be the corresponding test statistic. If $\delta = 0$ or $\Delta = 0$, the corresponding test statistic follows an asymptotic standard normal distribution. To test the null hypothesis $H_0 : \delta < 0$ or $\Delta \le 0$ against the alternative $H_1 : \delta \ge 0$ or $\Delta > 0$, an intersection-union test (IUT, Berger and Sinclair 1984, Berger 1989, Liu and Berger 1995) rejects the null hypothesis when both $z^{\delta} > M$ and $z^{\Delta} > M$ for some constant M. In order for the test to have a size of α ($0 < \alpha < 1$), M has to be chosen such that

$$\sup_{H_0} \lim_{\min_u n_u \to \infty} \Pr(z^{\delta} > M, z^{\Delta} > M) = \alpha.$$

Notice that

$$p(\delta, \Delta, M) = \lim_{\min_u n_u \to \infty} \Pr\left(z^{\delta} > M, z^{\Delta} > M\right) = \lim_{\min_u n_u \to \infty} \int_{m_\delta}^{\infty} \int_{m_\Delta}^{\infty} \frac{(2\pi)^{-1}}{\sqrt{|\Psi_1|}} \exp\left[-0.5 Z \Psi_1^{-1} Z^t\right] dz^{\delta} dz^{\Delta}, \quad (4)$$

where $Z = (z^{\delta}, z^{\Delta})$, $m_\delta = M - \delta / \sigma_{\hat{\delta}}$ and $m_\Delta = M - \Delta / \sigma_{\hat{\Delta}}$, and

$$\Psi_1 = \begin{pmatrix} 1 & \sigma_{\hat{\delta}\hat{\Delta}} / (\sigma_{\hat{\delta}} \sigma_{\hat{\Delta}}) \\ \sigma_{\hat{\delta}\hat{\Delta}} / (\sigma_{\hat{\delta}} \sigma_{\hat{\Delta}}) & 1 \end{pmatrix}.$$

Because $\lim_{\delta \to +\infty} m_\delta = -\infty$, it follows that if $\Delta = 0$, i.e., still staying within H_0, then, as $\min_u n_u \to \infty$,

$$\alpha = \sup_{H_0} \Pr(z^{\delta} > M, z^{\Delta} > M) \ge \lim_{\delta \to +\infty} P(\delta, \Delta, M) = \Pr(z^{\Delta} > M | \Delta = 0).$$

Therefore, when $M = z_\alpha$, the 100α upper percentile of the standard normal distribution, the rejection region $z^{\delta} > M$ and $z^{\Delta} > M$ for the IUT provides asymptotic size α for testing $H_0 : \delta < 0$ or $\Delta \le 0$ against the alternative $H_1 : \delta \ge 0$ and $\Delta > 0$. The corresponding approximate power function for the IUT is then given by

$$p(\delta, \Delta) = \Pr(z^{\delta} > z_\alpha, z^{\Delta} > z_\alpha). \quad (5)$$

Thus, the sample sizes required to achieve an asymptotic statistical power of $(1-\eta)$ ($0 < \eta < 1$) are the solutions to n_{tt}, n_{pp}, and n_{pt} such that $P(\delta,\Delta) = 1 - \eta$.

If the total spacing t_3 has been determined, given the simple DM trial we consider here, an important design question is how the treatment switch time, t_2, should be chosen to give the best estimate to the efficacy parameters δ and Δ. Another important set of design parameters to be optimized is the sample size allocations λ_u, $u = pp, tt, pt$, subject to $\lambda_u > 0$,, and $\lambda_{pp} +$

$\lambda_{tt} + \lambda_{pt} = 1$. Because the DM design requires simultaneous estimates to two parameters, (δ,Δ), an optimum design must simultaneously minimize the variances associated with both estimates, $(\hat{\delta},\hat{\Delta})$. Hence, it is necessary to find the maximum possible variance from any linear combinations of the two estimates. Let $\hat{L}(\tau_1, \tau_2) = \tau_1\hat{\delta} + \tau_2\hat{\Delta}$ be a linear combination of the two estimators with weights (τ_1, τ_2) subject to $\tau_1^2 + \tau_2^2$. The variance of the linear combination is

$$\sigma^2_{\hat{L}(\tau_1,\tau_2)} = (\tau_1, \tau_2)\Psi(\tau_1, \tau_2)^t. \quad (6)$$

The maximum variance from all possible linear combinations is the largest eigenvalue of Ψ (Noble and Daniel, 1977) as given by

$$\max_{\tau_1^2+\tau_2^2=1}\sigma^2_{\hat{L}(\tau_1,\tau_2)} = 0.5(\sigma^2_{\hat{\delta}} + \sigma^2_{\hat{\Delta}} + \sqrt{(\sigma^2_{\hat{\delta}} - \sigma^2_{\hat{\Delta}})^2 + 4\sigma^2_{\hat{\delta}\hat{\Delta}}}). \quad (7)$$

Notice that $\max_{\tau_1^2+\tau_2^2=1}\sigma^2_{\hat{L}(\tau_1,\tau_2)}$ depends on both individual variances of $(\hat{\delta},\hat{\Delta})$ and the covariance between them, $\sigma_{\hat{\delta}\hat{\Delta}}$. In the extreme case when two estimators are independent, i.e., $\sigma_{\hat{\delta}\hat{\Delta}} = 0$, $\max_{\tau_1^2+\tau_2^2=1}\sigma^2_{\hat{L}(\tau_1,\tau_2)}$ is the maximum of the variances from $(\hat{\delta},\hat{\Delta})$. To optimize the DM design, one criterion is to use the 'minimax' criterion, i.e., to choose all design parameters to minimize the maximum variance, $\max_{\tau_1^2+\tau_2^2=1}\sigma^2_{\hat{L}(\tau_1,\tau_2)}$, from all possible linear combinations of $(\hat{\delta},\hat{\Delta})$. Notice that $n\max_{\tau_1^2+\tau_2^2=1}\sigma^2_{\hat{L}(\tau_1,\tau_2)}$ (n is the total sample size) is a function of λ_u, $u = pp, tt, pt,$ and t_2. Assuming that the duration t_3 of the simple DM trial is pre-specified, we locate the optimum treatment switch time t_2 and the sample size allocations λ_u, $u = pp, tt, pt,$ by minimizing $n\max_{\tau_1^2+\tau_2^2=1}\sigma^2_{\hat{L}(\tau_1,\tau_2)}$. Notice that the focus of the DM trial with a randomized start design is on the delayed and non-delayed treatment arms, i.e., data from these two arms alone already allow the test of the DM efficacy. Subjects randomized into the placebo again from the second randomization mainly serve to preserve the blinding of subjects and investigators to the active treatment, albeit they have to participate in the efficacy analyses based on the 'intent-to-treat' principle (Montori and Guyatt 2001, Heritier *et al.* 2003). Hence, it is practical for the investigators to pre-specify a small portion of subjects (denoted by λ_0) to be randomized to placebo from the second randomization, i.e., $u=pp$. The optimum treatment switch time t_2 and sample size allocations are then the solutions for minimizing $n\max_{\tau_1^2+\tau_2^2=1}\sigma^2_{\hat{L}(\tau_1,\tau_2)}$, subject to $0 = t_1 < t_2 < t_3$ and $0 < \lambda_{tt} = 1 - \lambda_0 - \lambda_{pt} < 1$. Mathematically, $n\max_{\tau_1^2+\tau_2^2=1}\sigma^2_{\hat{L}(\tau_1,\tau_2)}$, as a function of t_2 and λ_{pt}, is complicated. The minimization over t_2 and λ_{pt} has no closed form, and can be done by Newton–Raphson method (Bonnans *et al.* 2006) to solve the system of equations

$$\begin{cases} \frac{\partial(n\max_{\tau_1^2+\tau_2^2=1}\sigma^2_{\hat{L}(\tau_1,\tau_2)})}{\partial t_2} = 0 \\ \frac{\partial(n\max_{\tau_1^2+\tau_2^2=1}\sigma^2_{\hat{L}(\tau_1,\tau_2)})}{\partial \lambda_{pt}} = 0 \end{cases}. \quad (8)$$

2.2. Specification of covariance matrix for the trivariate distribution

It is important to point out that the optimum design we proposed above does not assume a specific parametric family of trivariate distribution (e.g., normal distribution) for Y_j^u. However, the proposed methodology does require the existence of the mean vector $\mu^u = (\mu_1^u, \mu_2^u, \mu_3^u)^t$ and the covariance matrix for Y_j^u. The common covariance structure

assumed for *u=pp, pt*, and *tt* has to be specified to derive the optimum design parameters for the DM trials. Because general linear mixed effects models have been very successful to fit the longitudinal data from many of the outcomes in AD studies (Johnson *et al.* 2009, Storandt *et al.* 2006), here we focus on the specific covariance structure that can be derived from the random intercept and random slope models (Laird and Ware 1982) on $y_j^u(t)$, 0 = t_1 ≤ t ≤ t_3, for group *u=pp* or *tt*. Assuming a linear trend over time for $y_j^u(t)$, a random intercept and slope model assumes that

$$y_j^u(t)=\beta_{0j}^u+\beta_{1j}^u t+\varepsilon_j^u(t), \quad (9)$$

where $(\beta_{0j}^u, \beta_{1j}^u)$ is subject-specific vector of latent intercept and slope, and $\varepsilon_j^u(t)$ is the within-subject random error over time. We further assume a bivariate (not necessarily normal) distribution for $(\beta_{0j}^u, \beta_{1j}^u)$ across subjects with a 2 by 2 covariance matrix

$$\Omega=\begin{bmatrix} \omega_1^2 & \omega_{12} \\ & \omega_2^2 \end{bmatrix}, \quad (10)$$

and a stationary error process for $\varepsilon_j^u(t)$, 0 = t_1 ≤ t ≤ t_3, with $E[\varepsilon_j^u(t)]=0$ and the autocovariance function given by

$$\rho(h)=\mathrm{cov}[\varepsilon_j^u(t), \varepsilon_j^u(t+h)], \quad (11)$$

for $h \geq 0$. For example, if ρ(*h*) is a constant when $h>0$, it corresponds to the compound symmetry covariance structure in longitudinal models (Diggle *et al.*, 2002). If

$$\rho(h)=\tau^2\rho^h, \quad (12)$$

for $h>0$, it corresponds to the autoregressive covariance structure in longitudinal models (Diggle *et al.*, 2002). Finally, with the standard assumption of independence between $(\beta_{0j}^u, \beta_{1j}^u)$ and $\varepsilon_j^u(t)$, 0 = t_1 < t_2 < t_3 in the linear mixed model, the covariance matrix for $Y_j^u=(y_j^u(t_1), y_j^u(t_2), y_j^u(t_3))^t$ is given by

$$\Sigma=\begin{pmatrix} \omega_1^2+\gamma(0) & \omega_1^2+t_2\omega_{12}+\gamma t_2 & \omega_1^2+t_3\omega_{12}+\gamma t_3 \\ & \omega_1^2+2t_2\omega_{12}+t_2^2\omega_2^2+\gamma(0) & \omega_1^2+(t_2+t_3)\omega_{12}+t_2t_3\omega_2^2+\gamma(t_3-t_2) \\ & & \omega_1^2+2t_3\omega_{12}+t_3^2\omega_2^2+\gamma(0) \end{pmatrix}. \quad (13)$$

To assess how different parameters in the covariance structure affect the optimum choices for the treatment switch time and the sample size allocations to different treatment arms, we compute the optimum design parameters in a hypothetical DM trial with a duration of unit 1, i.e., t_1 = 0, t_3 = 1. We assume the following variance/covariance parameter for $(\beta_{0j}^u, \beta_{1j}^u)$ in Equation (10): $\omega_1^2=\omega_2^2=4$, ω_{12} = 4γ, where γ is the correlation between the latent intercept and the slope across subjects. We also assume an autoregressive covariance structure for $\varepsilon_j^u(t)$ in Equation (12) with τ^2 = 6 and ρ, the correlation between $\varepsilon_j^u(t)$ and $\varepsilon_j^u(t+1)$. For a wide range of γ and ρ, assuming 10% subjects are randomized to the placebo throughout the trial, Table 2 presents the optimum time of treatment switch (t_2) and optimum sample size allocations (λ_{tt}, λ_{pt}) to treatment groups. Results in Table 2 suggest that the optimum sample size allocations (λ_{tt}, λ_{pt}) are remarkably stable and robust as a function of the two correlations γ and ρ. Further, the correlation γ only minimally affects the optimum choice of treatment switch time t_2, whereas the correlation ρ in the autoregressive covariance structure

for $\varepsilon_j^u(t)$ seems to play a moderately bigger role in affecting the optimum treatment switch time, i.e., as ρ increases, the optimum treatment switch time t_2 increases moderately.

3. Application: optimum DM trials on AD

We demonstrate the proposed methodology to design a future DM trial on AD using the latest available information from the literature. AD is characterized by an insidious onset of memory deterioration, progressive cognitive deterioration, emergence of neuropsychiatric symptoms and behavioral disturbances, impairment of activities of daily living, and loss of independent function. The looming public health crisis due to AD mandates a fast development of novel DM treatments for the disease. Currently, several compounds have been approved by the FDA to enhance cognition and global function of AD patients through symptomatic trials alone. On the other hand, recent advances in understanding AD pathogenesis have led to the development of numerous compounds that might modify the disease process. A wide array of antiamyloid and neuroprotective therapeutic approaches are under investigation on the basis of the hypothesis that amyloid beta (Abeta) protein plays a pivotal role in disease onset and progression and that secondary consequences of Abeta generation and deposition, including tau hyperphosphorylation and neurofibrillary tangle formation, oxidation, inflammation, and excitotoxicity, contribute to the disease process (Salloway *et al.* 2008). Interventions in these processes with agents that reduce amyloid production, limit aggregation, or increase removal might block the cascade of events comprising AD pathogenesis. Reducing tau hyperphosphorylation, limiting oxidation and excitotoxicity, and controlling inflammation might be beneficial disease-modifying strategies. Potentially neuroprotective and restorative treatments such as neurotrophins, neurotrophic factor enhancers, and stem cell-related approaches are also under investigation (Salloway *et al.* 2008). It is anticipated that these promising treatments will soon be tested for their ability to modify the disease process of AD through well designed DM clinical trials.

Unlike symptomatic trials for which a single randomization at baseline is generally implemented, DM trials require an initial randomization followed by a re-randomization of patients in either the placebo or treatment arm. Although the design and analysis of DM trials on AD have been extensively discussed in the AD literature (Cummings 2006, Aisen 2006, Citron 2004, Mani 2004), no DM trials on patients with AD have been reported to date. We therefore have to obtain necessary estimates to important model parameters through recently reported symptomatic trials on AD. These parameters include between and within subject variances and covariance ω_s^2, $s = 1,2$, ω_{12} in Equation (10) and τ^2 and ρ (in the autoregressive function of $\varepsilon_j^u(t)$, $t_1 \le t \le t_3$). Using these estimates, here we provide optimum design parameters for future DM trials on AD by applying our proposed methodology to a variety of design scenarios. We assume a randomized start design in which 10% subjects will be randomized to receiving placebo throughout the trial and then optimize the sample size allocations to the treatment arm and the delayed treatment arm. We also optimize the time of treatment switch for the delayed treatment arm. The optimum weights c and f as given in Equations (2) and (3) are used to estimate β^{tt} and β^{pp}, as well as subsequently in the IUT of the DM efficacy.

More specifically, we conducted a comprehensive literature review on symptomatic clinical trials on AD. Most recently reported symptomatic trials on AD used ADAS-cog as the primary efficacy outcome measure. We therefore assumed that the future DM trials will also use the same cognitive outcome as the primary efficacy endpoint. Essentially all published symptomatic trials on AD reported the efficacy analyses using the change of ADAS-cog score from the baseline. These published symptomatic treatment trials on AD followed

patients for a duration ranging from 4 weeks to 1.5 year (Qizilbash *et al.* 1998; Kaduszkiewicz *et al.* 2005, Sano *et al.* 2011), and therefore the reported variance for the change from baseline on ADAS-cog score also spanned a wide range. None of the published symptomatic trials on AD directly reported estimates to model parameters ω_s^2, $s = 1,2$, ω_{12}, τ^2, and ρ as given in Equations (10) and (12). We hence combined the reported statistics from multiple published symptomatic trials on AD to obtain estimates necessary for optimizing DM trials with three outcome assessments. Three recently published trials (two published in 2011) that were reasonably large in sample size and long in follow-up duration were identified. These three trials specifically reported the variance associated with the change of ADAS-cog score from the baseline for the placebo arm (i.e., $u=pp$). Sano *et al.* (2011) reported the effect of simvastatin in treating mild to moderate AD for an 18 months trial from which 202 subjects were randomized to the placebo. Rafii *et al.* (2011) reported the effect of huperzine A in treating mild to moderate AD for a 16 weeks trial from which 73 subjects were randomized to the placebo. Rogers *et al.* (1998) reported the effects of Donepezil in treating AD for a 24-weeks trial from which 162 subjects were randomized to placebo. Because of variable length of longitudinal follow-up for these trials, the reported variance associated with the change of ADAS-cog score from the baseline is a function of the length of follow-up. However, if model (9) is appropriate for the placebo arm of the reported symptomatic trials, i.e., assuming a linear growth pattern of ADAS-cog over time, the annual rate of change on ADAS-cog (i.e., the slope) in the placebo arm can be estimated (sometimes through extrapolations with less than 1 year follow-up) by the reported mean difference from baseline divided by the length of follow-up (in years). Therefore, the standard deviation for the annual rate of change can be estimated by the reported standard deviation on the change from baseline divided by the number of years in follow-up. Let D be the duration of a reported symptomatic trial on AD. We linked the reported statistics on the change score of ADAS-cog from the baseline with our proposed covariance structure in Equation (13). Whereas these published trials did assess subjects at more than two time points, our approach is likely the only practical one because none of these trials reported relevant statistics on the efficacy at middle time points from the baseline to the final assessment. Specifically, the model given by Equation (9) implies

$$\frac{y_j^{pp}(D) - y_j^{pp}(0)}{D} = \beta_{1j}^{pp} + \frac{\varepsilon_j^{pp}(D) - \varepsilon_j^{pp}(0)}{D}.$$

It then follows that

$$\frac{\sigma_d^2}{D^2} = \frac{2\tau^2(1-\rho^D)}{D^2} + \omega_2^2, \quad (14)$$

where σ_d^2 is the reported variance for the change score of ADAS-cog from the baseline in the placebo arm. In addition, all three published trials reported basic statistics of ADAS-cog at baseline for the placebo arm, which can be used to estimate the variance parameters in our covariance structure associated with the baseline outcome measure (Equation (13)). In fact, at baseline, i.e., $t = t_1 = 0$, $\sigma_b^2 = \tau^2 + \omega_1^2$, where σ_b^2 is the reported variance for the baseline ADAS-cog. The reported statistics on the placebo arm from the three trials are given in Table 3.

From the reported statistics of the three trials (Rogers *et al.* 1998, Rafii *et al.* 2011, and Sano *et al.* 2011) along with Equation (14), we have the following system of equations:

$$\begin{cases} 6.06^2/0.46^2 = 2\tau^2(1-\rho^{0.46})/0.46^2 + \omega_2^2 \\ 5.17^2/0.31^2 = 2\tau^2(1-\rho^{0.31})/0.31^2 + \omega_2^2 \\ 8.70^2/1.50^2 = 2\tau^2(1-\rho^{1.50})/1.50^2 + \omega_2^2 \end{cases}.$$

Solving the system of equations revealed that ρ = 0.319, $\omega_2^2 = 1.125$, τ^2 = 44.627. Finally, using the baseline variance estimate from the largest trial in Table 3 (i.e., Sano *et al.* 2011), we estimated ω_1^2 by $10.5^2 = \tau^2 + \omega_1^2$, resulting $\omega_1^2 = 65.624$.

We have now obtained estimates to variance and covariance parameters ω_s^2, s = 1,2, τ^2 and ρ in the covariance structure as given by Equation (13) except for ω_{12}, the covariance between the latent intercept and the slope across subjects in the random intercept and random slope model (Equation (9)). No information was directly reported in published symptomatic trials on AD that can be used to obtain an estimate to ω_{12}. Because $\omega_{12} = \gamma\omega_1\omega_2$, we therefore searched for the optimum design parameters for future DM trials on AD as a function of γ, the correlation between the latent intercept and the slope across subjects. For a range of γ, λ_{pp} = 10%, t_3 = 1.5 years or 2 years or 2.5 years, Table 4 reports the optimum sample size allocations to different arms as well as the optimum time of treatment switch for future simple DM trials on AD with three longitudinal assessments. Results in Table 4 confirm what has been found in Table 2, i.e., the optimum sample size allocations (λ_{tt}, λ_{pt}) and the optimum choice of treatment switch (t_2) are very stable and robust as a function of the correlation γ between the latent intercept and the slope.

Assuming a randomized start design with three efficacy assessments for a duration of 1.5 years, 2 years and 2.5 years, for a selected set of effect sizes typically reported in the AD literature, Table 5 presents the sample sizes for different treatment arms (i.e., *u=pt* and *u=tt*) required to detect the effect sizes with a statistical power of 80%. These individual sample sizes are based on the optimum sample size allocations (in %) to group *u=pt* and *u=tt* (i.e., λ_{pt} and λ_{tt}) and optimum treatment switch time as presented in Table 4. 10% subjects are assumed to be randomized into the placebo arm throughout the trial for preserving the blinding. Because the correlation (γ) between the latent intercept and the slope across subjects does not appreciably affect the optimum design parameters (i.e., sample size allocations λ_{pt} and λ_{tt}, and treatment switch time t_2) as evidenced in Table 4, we assumed that γ = 0.5 in Table 5. Power function in Equation (5) was evaluated by SAS function PROBBNRM (SAS Institute Inc., 1990) in Table 5.

4. Discussion

In order to design optimum clinical trials for establishing the DM efficacy of potential novel treatments on chronic disease such as AD, we proposed a methodology to analyze the rate of change for efficacy outcome variable in a randomized start trial. We focused on the simplest longitudinal design with only three repeated outcome assessments from which the middle one serves as the time of treatment switch. We assumed a trivariate distribution without a specific parametric structure for the three repeated measures on the same subjects with a covariance matrix that was derived from a random intercept and random slope model (Laird and Ware 1982). Based on these assumptions, we first formulated the appropriate DM hypothesis by comparing the rate of change in efficacy outcome between the treated group throughout the trial and the delayed treatment group. Because of the second randomization to the subjects who are initially randomized to the placebo in a DM trial, a third treatment arm in which subjects are randomized to the placebo throughout the trial is available. The third treatment arm complicated the statistical test of DM efficacy due to the ‘intent-to-treat’

principle that requires that data from all randomized and treated subjects participate in the efficacy analysis. Because multiple unbiased estimators were available to the rate of change for the placebo and the treatment arms, we obtained optimum unbiased estimates to the rates of change by minimizing the variance of linear combinations of multiple estimators. With these optimum estimates, we developed a methodology to optimally determine the sample size allocations to different treatment arms as well as the time for treatment switch for subjects whose treatment was delayed. After the design parameters were optimally chosen, we proposed an intersection-union test to assess the efficacy of potential DM agents. We studied the size and the power of the IUT, and provided a method of determining the sample sizes to adequately power the test of DM efficacy.

The simple DM trials with a randomized start design we considered here differ from the standard crossover designs (Chi 1992) in the sense that the former allows some of the subjects receiving only placebo throughout the trials. Further, whereas most crossover trials are designed for testing the symptomatic efficacy of a novel treatment, the simple DM trials aim to establish DM efficacy which mandates the symptomatic efficacy as a prerequisite during the first phase of the trial. Therefore, the time of treatment switch becomes an important parameter in designing DM trials. Our proposed methodology provided an optimum choice of treatment switch time. Jarjoura (2003) considered the efficiency of a clinical trial design which did allow crossing control to treatment (i.e., similar to our designs), but optimal switching time for crossing controls to active treatment was only discussed under the assumption that the treatment effect is constant across time (please see page 311 of Jarjoura (2003)). Here, we have focused on the DM efficacy which requires not only a symptomatic treatment efficacy that is time-varying prior to treatment switching time but also a DM treatment efficacy that depends on whether the treatment is delayed or not. In addition, our analytic approaches also differed from those of other authors (Jarjoura 2003). The randomized start design we discussed here allowed the third arm of continued control throughout the trial (to protect the blinding after second randomization), which complicated the parameter estimations much more considerably than the two arm scenario that was discussed in Jarjoura (2003). We have assumed a trivariate distribution without a specific parametric structure for the three repeated measures on the same subjects whose mean vectors across different arms were used to define the DM hypothesis. The covariance structure of the trivariate distribution, however, was derived from a standard random intercept and random slope model (Laird and Ware 1982) which assumed a bivariate distribution for the latent intercept and slope and incorporated an autoregressive structure on the error process. We assumed a linear trend over time for subjects whose treatment was not switched in the middle of the trials. For individuals whose treatment was delayed, the simple DM trials we considered here forced a piecewise linear pattern. Our approach allowed potentially differential rate of change for subjects receiving delayed treatment as compared to those receiving the treatment throughout the trial, as well as a correlation on the rates of change before and after the treatment switch. Through a hypothetical DM trial, we found that the optimum sample size allocations were very stable and robust against different choices of the correlation (i.e., γ) between the latent intercept and the slope and the correlation (i.e., ρ) in the autoregressive covariance structure for $\varepsilon_j^u(t)$ in model (11). We also found that the correlation ρ in the autoregressive covariance structure for $\varepsilon_j^u(t)$ played a moderately bigger role in determining the optimum treatment switch time than the correlation γ.

The current definition of DM efficacy in clinical trials with a randomized start design requires two conditions to be met simultaneously: one is the symptomatic efficacy before the treatments are switched, and the other is the fact that the delayed treatment arm will never ‘catch up’ the continuously treated arm. It is therefore very intuitive to use some type of

adaptive design in which the first condition (i.e., the symptomatic efficacy) is tested using data prior to the treatment switch, and then the second condition tested after the treatment switch. This type of adaptive design is called adaptive-hypotheses design (Chow and Chang 2008), and can be especially appealing for testing a novel treatment for which no symptomatic efficacy has ever been established, because the adaptive testing can prevent the trial from continuing (and thus saving cost) if the test of the first condition (symptomatic efficacy) is not even successful. On the other hand, even if the test of the first condition is successful, it can not establish the DM efficacy that requires both conditions simultaneously. Because our goal is to optimize design parameters such as treatment switch time and sample size allocations to different arms in a future DM trial and the cost of the clinical trial is not considered in the optimum design of the DM trials for the current work, we have hence chosen to use a randomized start design to test the DM efficacy by allowing a treatment switch in the middle of the trial. These are especially relevant to the companies whose drugs have already been approved by the FDA due to their successful symptomatic trials on the symptomatic efficacy. What they would like to do next is to prove that their drugs have additional DM efficacy. To achieve this, they would need to design a future DM trial involving a treatment switch. Our proposed methods provide them parameters to optimally design such trials.

We applied our proposed methodology to design future simple DM trials on AD. Because no DM trials on AD have been reported, we conducted a literature review on published symptomatic trials on AD, and located three recently reported symptomatic trials on AD that were relatively large in sample size and long in follow-up and also reported the variance associated with the change of ADAS-cog score from the baseline for the placebo arm (Rogers *et al*. 1998, Rafii *et al*. 2011, and Sano *et al*. 2011). Given that none of the three trials directly reported the estimates to parameters in the covariance structure in Equation (13), we proposed a novel approach to obtain pilot estimates to these important parameters by linking our proposed covariance structure with the reported statistics. More specifically, we solved a system of three equations that were derived from the reported variances associated with the change of ADAS-cog score from the baseline in the three trials. After obtaining estimates to these parameters, we computed the optimum design parameters (i.e., sample size allocations, and treatment switch time) for future simple DM trials on AD, and provided the sample sizes into different treatment arms required to detect a selected set of effect sizes with a statistical power of 80%. Our results indicate that clinical trials for DM treatments on AD can be adequately powered and optimized. Because a DM treatment, by definition, requires symptomatic efficacy, it is no surprise that a DM trial on AD requires a much larger sample size than symptomatic trials, as demonstrated by our findings. Another contributing factor to the relatively large sample sizes in future DM trials on AD is the relatively small effect sizes we have intentionally assumed in Table 5. This is based on the fact that, although ADAS-cog has shown a wide range of effect sizes in published symptomatic trials, smaller effect sizes are expected for future novel drugs on AD because any future clinical trials on AD will have to be conducted on the patients who have already been using approved drugs, i.e., the efficacy of future drugs on AD will have to be based on the incremental effect from the already approved AD drugs. Further, Table 5 suggests that, if the effect sizes are fixed, then the biggest effect on the required sample size is the trial duration. Whereas we have specifically considered the simplest longitudinal design of the DM trials with 3 observations per subject here, the issue of dropouts and missing data will play an important role in determining the final sample sizes of future DM trials. Given that a longer trial duration is in general associated with a higher likelihood of subject dropouts, further research is needed to fully understand the trade offs in final sample sizes between dropouts and the required three observations per subject. Finally, the relatively large sample size in Table 5 is also partly due to the fact that the rate of change on ADAS-cog is subject to a large variation as evidenced by Table 3 and therefore may not be an ideal efficacy

outcome in future DM trials. More sensitive novel biomarkers will be needed to design future DM trials on AD with a much smaller sample size. For example, MRI-based brain volumes (Storandt *et al*. 2009), DTI-based measures of white matter microstructure (Head *et al*. 2004), CSF-based biomarkers (Fagan *et al*. 2006), and molecular imaging of cerebral fibrillar amyloid with PET using the [^{11}C] benzothiazole tracer, Pittsburgh Compound-B (PIB, Mintun *et al*. 2006), are all potential candidates of efficacy outcomes for future DM trials on AD. Our study has some limitations. First, we have assumed a linear growth or decline pattern over time for the efficacy outcome in group $u=tt$ and $u=pp$ and derived the optimum design parameters. Although our proposed method can be generalized to address a nonlinear pattern from time 1 to 2 to 3 in group $u=tt$ and $u=pp$, the robustness of our optimum design parameters with respect to the potentially nonlinear growth patterns remained unknown. It is therefore important to carefully assess the linearity assumption before applying our proposed optimum designs. Second, we have focused on the simplest longitudinal design in which only three repeated efficacy assessments are planned for each subject: one at the baseline, one at the end of the trial, and the other at the time when the treatments are switched. More frequent observations per subject allow statistical assessment of the assumption of a linear response within each of the three groups, as well as appropriate analyses of missing data. Further, with reasonable statistical models, more frequent observations will also imply better statistical power for testing the DM efficacy. In addition, our proposed methodology has focused on one-sided tests to reflect the current definition of DM efficacy, i.e., the requirement of both the symptomatic efficacy before the treatments are switched and the additional efficacy demonstrating that the delayed treatment arm will never ‘catch up’ the continuously treated arm. The generalization to the efficacy tests with two-sided alternatives through the IUT is needed in real world applications when there is no guarantee that a novel treatment is always better than the placebo. Further, our proposed methodology only applies when the treatment will not have any serious side effect so that treatment can be given without stop and there is no alternative treatment besides the treatment in the trial. Finally, FDA typically requires two primary efficacy endpoints in clinical trials on AD (cognition and function). Although our proposed method can be applied to each efficacy endpoint in the DM trials of AD, the optimal design parameters for cognition is not necessarily optimal for function. Future research is needed to address these crucial questions.

Acknowledgments

The authors would like to thank the editor and two reviewers for their constructive critiques that have helped improve the manuscript considerably. This study was supported by National Institute on Aging (NIA) grant R01 AG029672 and R01 AG034119 for Chengjie Xiong. This study was also partly supported by the NIA grant P50 AG05681, P01 AG03991, P01AG26276, and U01 AG032438 for Chengjie Xiong and John Morris.

References

Aisen PS. Commentary on “Challenges to demonstrating disease-modifying effects in Alzheimer’s disease clinical trials”. Alzheimer’s & Dementia. 2006; 6:272–274.

Andrieu S, Rascol O, Lang T, et al. Methodological issues and statistical analyses. J Nutr Health Aging. 2006; 10:116–117. [PubMed: 16554944]

Berger RL. Uniformly more powerful tests for hypotheses concerning linear inequalities and normal means. Journal of the American Statistical Association. 1989; 84:192–199.

Berger RL, Sinclair DF. Testing hypotheses concerning unions of linear subspaces. Journal of the American Statistical Association. 1984; 79:158–163.

Bonnans, J-F.; Gilbert, JC.; Lemarechal, C.; Sagastizábal, CA. Numerical optimization: Theoretical and practical aspects. Berlin: Springer-Verlag; 2006.

Chi EM. Analysis of cross-over trials when within-subject errors follow an AR(1) process. Biom. J. 1992; 34:359–365.

Chow S-C, Chang M. Adaptive design methods in clinical trials—a review. Orphanet Journal of Rare Diseases. 2008; 3:11. [PubMed: 18454853]

Citron M. Strategies for disease modification in Alzheimer's disease. Net Rev Neurosci. 2004; 5:677–685.

Cummings JL. Challenges to demonstrating disease-modifying effects in Alzheimer's disease clinical trials. Alzheimer's & Dementia. 2006; 6:263–271.

Cummings JL. Optimizing phase II of drug development for disease-modifying compounds. Alzheimers Dement. 2008; 4:S15–S20. [PubMed: 18631992]

Cummings JL, Doody R, Clark C. Disease-modifying therapies for Alzheimer disease: challenges to early intervention. Neurology. 2007; 69:1622–1634. [PubMed: 17938373]

Diggle, PJ.; Heagerty, P.; Liang, K-Y.; Zeger, SL. Analysis of Longitudinal Data. 2nd ed.. New York: Oxford University Press; 2002.

Fagan AM, Mintun MA, Mach RH, et al. Inverse relation between in vivo amyloid imaging load and cerebrospinal fluid $A\beta_{42}$ in humans. Ann Neurol. 2006; 59:512–519. [PubMed: 16372280]

Head D, Buckner RL, Shimony JS, Williams LE, Akbudak E, Conturo TE, McAvoy M, Morris JC, Snyder AZ. Differential vulnerability of anterior white matter in nondemented aging with minimal acceleration in dementia of the Alzheimer type: evidence from diffusion tensor imaging. Cereb Cortex. 2004; 14(4):10–23.

Heritier SR, Gebski VJ, Keech AC. Inclusion of patients in clinical trial analysis: the intention-to-treat principle. Med J Aust. 2003; 179:438–440. [PubMed: 14558871]

Jarjoura D. Crossing controls to treatment in repeated-measures trials. Controlled Clinical Trials. 2003; 24:306–323. [PubMed: 12757996]

Johnson DK, Storandt M, Morris JC, Galvin JE. Longitudinal study of the transition from healthy aging to Alzheimer's disease. Arch Neurol. 2009; 66:1254–1259. [PubMed: 19822781]

Kaduszkiewicz H, Zimmermann T, Beck-Bornholdt H-P, van den Bussche H. Cholinesterase inhibitors for patients with Alzheimer's disease: systematic review of randomised clinical trials. BMJ. 2005; 331:321–327. [PubMed: 16081444]

Kryscio RJ, Mendiondo MS, Schmitt FA, Markesbery WR. Designing a large prevention trial: statistical issues. Stat Med. 2004; 23:285–296. [PubMed: 14716729]

Laird NM, Ware JH. Random-effects models for longitudinal data. Biometrics. 1982; 38:963–974. [PubMed: 7168798]

Leber P. Slowing the progression of Alzheimer's disease: methodologic issues. Alzheimer Dis Assoc Disord. 1997; (Suppl 5):S10–S20. [PubMed: 9348423]

Liu H, Berger RL. Uniformly more powerful, one-sided tests for hypotheses about linear inequalities. Annals of Statistics. 1995; 23:55–72.

Mani RB. The evaluation of disease modifying therapies in Alzheimer's disease: a regulatory viewpoint. Stat Med. 2004; 23:305–314. [PubMed: 14716731]

Mintun MA, LaRossa GN, Sheline YI, et al. [^{11}C] PIB in a nondemented population: Potential antecedent marker of Alzheimer disease. Neurology. 2006; 67:446–452. [PubMed: 16894106]

Montori VM, Guyatt GH. Intention-to-treat principle. CMAJ. 2001; 165:1339–1341. [PubMed: 11760981]

Noble, B.; Daniel, JW. Applied Linear Algebra. Englewood Cliffs, NJ: Prentice-Hall Inc.; 1977.

Qizilbash N, Whitehead A, Higgins J, Wilcock G, Schneider L, Farlow M. for the Dementia Trialists' Collaboration. Cholinesterase inhibition for Alzheimer disease: A meta-analysis of the Tacrine trials. JAMA. 1998; 280:1777–1782. [PubMed: 9842955]

Rafii MS, Walsh S, Little JT, Behan K, Reynolds B, Ward C, Jin S, Thomas R, Aisen PS. A phase II trial of huperzine A in mild to moderate Alzheimer disease. Neurology. 2011; 76:1389–1394. [PubMed: 21502597]

Ringman JM, Grill J, Rodriguez-Agudelo Y, Chavez M, Xiong C. Prevention Trials in Persons At-Risk for Dominantly-Inherited Alzheimer's Disease: Opportunities and Challenges. Alzheimer's & Dementia. 2009; 5:166–171.

Rogers SL, Farlow MR, Doody RS, Mohs R, Friedhoff LT. Donepezil Study Group. A 24-week, double-blind, placebo-controlled trial of donepezil in patients with Alzheimer's disease. Neurology. 1998; 50:136–145. [PubMed: 9443470]

Rosen WG, et al. A new rating scale for the Alzheimer’s disease. Am J Psychiatry. 1984; 141:1356–1364. [PubMed: 6496779]

Salloway S, Mintzer J, Weiner MF, Cummings JL. Disease-modifying therapies in Alzheimer's disease. Alzheimer's Dement. 2008; 4:65–79. [PubMed: 18631951]

Sampaio C. Alzheimer disease: disease modifying trials. Where are we? Where do we need to go? A reflective paper. J Nutr Health Aging. 2006; 10:113–115. [PubMed: 16554943]

Sano M, Bell KL, Galasko D, Galvin JE, Thomas RG, van Dyck CH, Aisen PS. A randomized, double-blind, placebo-controlled trial of simvastatin to treat Alzheimer disease. Neurology. 2011; 77:556–563. [PubMed: 21795660]

SAS Institute Inc.. SAS Language: Reference, Version 6. First Edition. Cary, NC: SAS Institute Inc.; 1990.

Storandt M, Grant EA, Miller JP, Morris JC. Longitudinal course and neuropathological outcomes in original versus revised MCI and in PreMCI. Neurology. 2006; 67:467–473. [PubMed: 16894109]

Storandt M, Mintun MA, Head D, Morris JC. Cognitive decline and brain volume loss as signatures of cerebral amyloid-β peptide deposition identified with Pittsburgh compound B. Arch Neurol. 2009; 66:1476–1481. [PubMed: 20008651]

Whitehouse PJ, Kittner B, Roessner M, et al. Clinical trial designs for demonstrating disease-course-altering effects in dementia. Alzheimer Dis Assoc Disord. 1998; 12:281–294. [PubMed: 9876956]

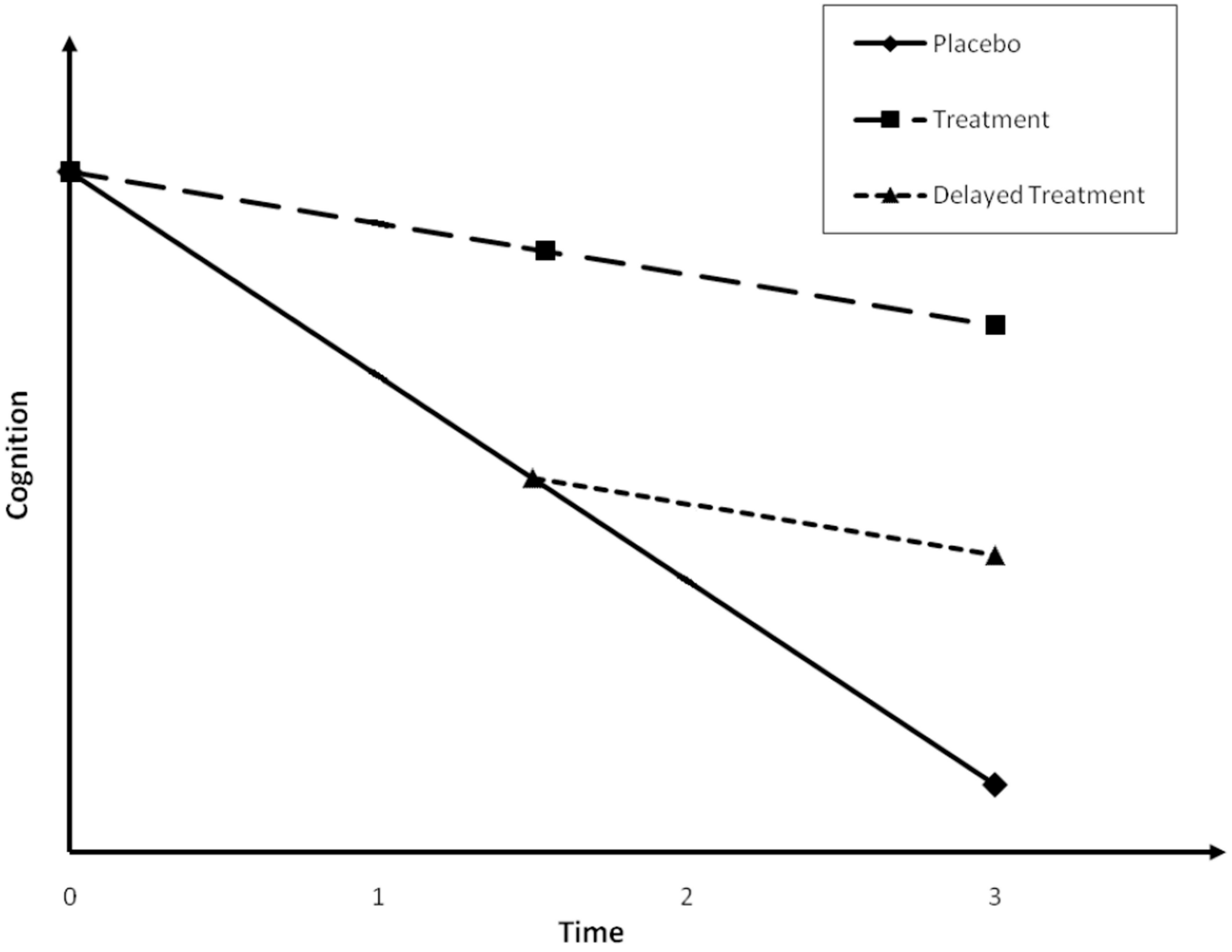

Fig. 1.
Expected cognitive progression in a simple disease modifying trial on AD

Table 1

Variance and covariance of estimated parameters for testing DM efficacy (Notice that the covariance becomes variance when two estimates are the same)

Estimate 1	Estimate 2	Covariance between Estimate 1 and Estimator 2
$\hat{\beta}^{tt}$	$\hat{\beta}^{tt}$	$\frac{c^2}{d_1^2}\left(\frac{\sigma_2^2}{n_{tt}}+\frac{s_1}{n}\right)+\frac{s_4(1-c)^2}{d_2^2 n_{tt}}+\frac{2c(1-c)}{d_1 d_2}\left(\frac{\sigma_{23}-\sigma_2^2}{n_{tt}}-\frac{s_3}{n}\right)$
$\hat{\beta}^{pp}$	$\hat{\beta}^{pp}$	$\frac{f^2}{d_1^2}\left(\frac{\sigma_2^2}{n-n_{tt}}+\frac{s_1}{n}\right)+\frac{(1-f)^2}{d_2^2}\left(\frac{\sigma_3^2}{n_{pp}}+\frac{s_2}{n-n_{tt}}\right)+\frac{2f(1-f)}{d_1 d_2}\left(\frac{\sigma_{23}-\sigma_2^2}{n_{pp}+n_{pt}}-\frac{s_3}{n}\right)$
$\hat{\beta}^{tt}$	$\hat{\beta}^{pp}$	$\frac{cfs_1}{nd_1^2}-\frac{(c+f-2cf)s_3}{nd_1 d_2}$
$\hat{\beta}^{pt}$	$\hat{\beta}^{pt}$	$\frac{1}{d_2^2}\left(\frac{\sigma_3^2}{n_{pt}}+\frac{s_2}{n_{pp}+n_{pt}}\right)$
$\hat{\beta}^{tt}$	$\hat{\beta}^{pt}$	$\frac{-cs_3}{nd_1 d_2}$
$\hat{\beta}^{pt}$	$\hat{\beta}^{pp}$	$\frac{f}{d_1 d_2}\left(\frac{\sigma_{23}-\sigma_2^2}{n_{pp}+n_{pt}}-\frac{s_3}{n}\right)+\frac{(1-f)s_2}{(n_{pp}+n_{pt})d_2^2}$
$\hat{\delta}$	$\hat{\delta}$	$\sigma^2_{\hat{\beta}^{tt}}+\sigma^2_{\hat{\beta}^{pt}}-2\mathrm{cov}(\hat{\beta}^{tt},\hat{\beta}^{pt})$
$\hat{\Delta}$	$\hat{\Delta}$	$\sigma^2_{\hat{\beta}^{pt}}+\sigma^2_{\hat{\beta}^{pp}}-2\mathrm{cov}(\hat{\beta}^{pt},\hat{\beta}^{pp})$
$\hat{\delta}$	$\hat{\Delta}$	$\mathrm{cov}(\hat{\beta}^{tt},\hat{\beta}^{pt})+\mathrm{cov}(\hat{\beta}^{pt},\hat{\beta}^{pp})-\sigma^2_{\hat{\beta}^{pt}}-\mathrm{cov}(\hat{\beta}^{tt},\hat{\beta}^{pp})$

Table 2

Optimum time of treatment switch (t_2) and sample size allocations (λ_{tt}, λ_{pt}) as a function of the correlation (γ) between the latent intercept and the slope and the correlation (ρ) in the autoregressive covariance structure for $\varepsilon_j^u(t)$ in a hypothetical DM trial. A total trial duration of unit one assumed (i.e., $t_1 = 0, t_3 = 1$). 10% subjects assumed to receive the placebo throughout the trial.

γ	ρ	t_2	λ_{tt}	λ_{pt}
0.1	0.1	0.25	0.17	0.73
0.1	0.3	0.30	0.18	0.72
0.1	0.5	0.33	0.20	0.70
0.1	0.7	0.34	0.20	0.70
0.1	0.9	0.35	0.19	0.71
0.3	0.1	0.26	0.17	0.73
0.3	0.3	0.30	0.19	0.71
0.3	0.5	0.33	0.20	0.70
0.3	0.7	0.34	0.20	0.70
0.3	0.9	0.35	0.19	0.71
0.5	0.1	0.27	0.18	0.72
0.5	0.3	0.30	0.19	0.71
0.5	0.5	0.32	0.20	0.70
0.5	0.7	0.33	0.20	0.70
0.5	0.9	0.34	0.19	0.71
0.7	0.1	0.27	0.18	0.72
0.7	0.3	0.31	0.20	0.70
0.7	0.5	0.31	0.20	0.70
0.7	0.7	0.32	0.20	0.70
0.7	0.9	0.33	0.18	0.72
0.9	0.1	0.27	0.18	0.72
0.9	0.3	0.30	0.20	0.70
0.9	0.5	0.31	0.20	0.70
0.9	0.7	0.32	0.20	0.70
0.9	0.9	0.32	0.17	0.73

The optimum allocation to group *pt* is from 59% to 67% in Table 2 if 20% subjects are assumed to receive the placebo throughout the trial.

Table 3

Reported statistics from three recently published symptomatic trials on ADAS-Cog for the placebo arm

Study: author (year)	Sample size	Years of follow-up	Mean (SD) at baseline	Mean (SD) for the change from baseline
Rogers *et al.* 1998	162	0.46	27.28 (11.87)	1.82 (6.06)
Rafii *et al.* 2011	73	0.31	27.10 (10.58)	−0.34 (5.17)
Sano *et al.* 2011	202	1.50	23.90 (10.50)	8.18 (8.70)

Table 4

Optimum time of treatment switch ($t2$) and sample size allocations (λ_{tt}, λ_{pt})in a DM trial on AD with three repeated outcome assessments. γ = correlation between the latent intercept and the slope. 10% subjects assumed to the placebo throughout the trial

Trial duration (years)	γ	λ_{tt}	λ_{pt}	t_2 (years)
1.5	0.1	0.15	0.75	0.45
1.5	0.3	0.15	0.75	0.45
1.5	0.5	0.15	0.75	0.45
1.5	0.7	0.15	0.75	0.45
1.5	0.9	0.15	0.75	0.45
2.0	0.1	0.14	0.76	0.54
2.0	0.3	0.15	0.75	0.57
2.0	0.5	0.15	0.75	0.57
2.0	0.7	0.15	0.75	0.57
2.0	0.9	0.15	0.75	0.57
2.5	0.1	0.14	0.76	0.65
2.5	0.3	0.15	0.75	0.68
2.5	0.5	0.15	0.75	0.69
2.5	0.7	0.15	0.75	0.70
2.5	0.9	0.16	0.74	0.73

The optimum allocation to group *pt* is from 62% to 64% in Table 4 if 20% subjects are assumed to receive the placebo throughout the trial.

Table 5

Sample sizes (*N*) for individual treatment arms (*u=tt* or *pt*) in a future simple randomized start design required to detect a selected set of effect sizes (i.e., differences on the annual rate of decline in ADAS-cog $(\delta,\Delta) = (\beta^{tt} - \beta^{pt} - \beta^{pt} - \beta^{pp})$) with a statistical power of 80% from the IUT. 10% subjects are randomized into placebo arm throughout the trial. A correlation of 0.5 (i.e., γ) between the latent intercept and the slope is assumed.

Trial duration (years)	$\delta = \beta^{tt} - \beta^{pt}$	$\Delta = \beta^{pt} - \beta^{pp}$	*N* for *u=tt*	*N* for *u=pt*	Total *N*
1.5	1.0	1.0	460	2301	3067
1.5	1.0	1.5	337	1689	2251
1.5	1.0	2.0	318	1593	2123
1.5	1.0	2.5	317	1589	2117
1.5	1.5	1.0	361	1805	2406
1.5	1.5	1.5	204	1023	1363
1.5	1.5	2.0	159	799	1064
1.5	1.5	2.5	145	725	966
1.5	2.0	1.0	351	1757	2342
1.5	2.0	1.5	167	836	1114
1.5	2.0	2.0	115	576	767
1.5	2.0	2.5	94	470	626
1.5	2.5	1.0	351	1756	2341
1.5	2.5	1.5	157	787	1049
1.5	2.5	2.0	97	486	647
1.5	2.5	2.5	73	368	490
2.0	1.0	1.0	294	1472	1962
2.0	1.0	1.5	213	1065	1420
2.0	1.0	2.0	199	996	1327
2.0	1.0	2.5	198	992	1322
2.0	1.5	1.0	234	1171	1561
2.0	1.5	1.5	130	654	871
2.0	1.5	2.0	101	506	674
2.0	1.5	2.5	91	455	606

Trial duration (years)	$\delta = \beta^{tt} - \beta^{at}$	$\Delta = \beta^{tt} - \beta^{pp}$	N for $u=tt$	N for $u=pt$	Total N
2.0	2.0	1.0	229	1146	1527
2.0	2.0	1.5	108	540	720
2.0	2.0	2.0	73	368	490
2.0	2.0	2.5	59	298	396
2.0	2.5	1.0	229	1146	1527
2.0	2.5	1.5	102	513	683
2.0	2.5	2.0	62	313	416
2.0	2.5	2.5	47	235	313
2.5	1.0	1.0	207	1035	1380
2.5	1.0	1.5	148	744	991
2.5	1.0	2.0	138	693	923
2.5	1.0	2.5	138	690	920
2.5	1.5	1.0	165	828	1103
2.5	1.5	1.5	92	460	613
2.5	1.5	2.0	70	354	471
2.5	1.5	2.5	63	318	423
2.5	2.0	1.0	162	812	1082
2.5	2.0	1.5	76	381	507
2.5	2.0	2.0	51	258	343
2.5	2.0	2.5	41	209	277
2.5	2.5	1.0	162	812	1082
2.5	2.5	1.5	72	363	483
2.5	2.5	2.0	44	221	294
2.5	2.5	2.5	33	165	220

(some rows do not add precisely due to the rounding up in each sample size)