Published in final edited form as:

Biostat Epidemiol. 2017 ; 1(1): 59–77. doi:10.1080/24709360.2017.1331822.

Validity and power of minimization algorithm in longitudinal analysis of clinical trials

Hua Weng, PhD¹, Randall Bateman, MD², John C. Morris, MD^{2,3,4}, and Chengjie Xiong, PhD^{1,4}

¹Division of Biostatistics, Washington University in St. Louis, St. Louis, MO

²Department of Neurology, Washington University in St. Louis, St. Louis, MO

³Departments of Pathology and Immunology, Washington University in St. Louis, St. Louis, MO

⁴Knight Alzheimer's Disease Research Center, Washington University in St. Louis, St. Louis, MO

Abstract

We studied the validity of longitudinal statistical inferences of clinical trials using minimization, a dynamic randomization algorithm designed to minimize treatment imbalance for prognostic factors. Repeated measures analysis of covariance and the random intercept and slope models, were used to simulate longitudinal clinical trials randomized by minimization or simple randomization. The simulations represented a wide range of analyses in real-world trials, including missing data caused by dropouts, unequal allocation of treatment arms, and efficacy analyses on either the original outcome or its change from baseline. We also analyzed the database from the Dominantly Inherited Alzheimer Network (DIAN), and used the estimated parameters to simulate the ongoing DIAN trial. Our analyses demonstrated minimization had conservative type I errors when the prognostic factor used in the minimization algorithm had a relatively strong correlation with the outcome and was not adjusted for in analyses. In contrast, adjusted tests for the prognostic factor as a covariate resulted in type I errors close to the nominal significance level. In many simulation scenarios, the adjusted tests using minimization had slightly greater statistical power than those using simple randomization, whereas in the other scenarios, the power of adjusted tests using these two randomization methods are almost indistinguishable.

Keywords

Clinical trial; Randomization; Minimization; Validity; Dominantly Inherited Alzheimer Network Trial Unit (DIAN TU)

1. Introduction

In clinical trials, simple randomization, also known as complete randomization, assigns participants independently into treatment arms with equal probability. In large trials, simple randomization could make it highly likely to balance on both known and unknown

prognostic factors or concomitant observations so that it helps prevent selection bias and maximize precision for statistical inference [1]. However, simple randomization may fail if it creates treatment arms that are unbalanced for critical prognostic factors. Particularly, in small to moderate sized trials, the problem of imbalance for factors in treatment assignments may exist to an extent that is not easily ignorable. For instance, Kernan et al. [2] showed that in a trial with 30 total patients, a binary factor, which is present in 15% of the study subjects, will result in chance imbalances of more than 10% between two treatment arms with probability 0.33. If such a factor is known or suspected to have a large effect on prognosis, the validity of statistical conclusions drawn from these imbalanced trials would be seriously challenged.

Stratified randomization, another widely-used method in clinical trials to assign subjects into treatment arms, is intended to prevent imbalance for prognostic factors. Stratified randomization is achieved by specifying stratification factors, such as gender, clinical center (for multicenter trials) and disease severity, before a trial is started. Each factor has two or more levels and a stratum is a combination of factor levels. Thus, the total number of strata is simply the product of the number of levels of each factor. As study subjects are enrolled, they are first assigned to the stratum to which they belong. Randomization is then carried out in each stratum. Unfortunately, as the number of strata increases, the imbalance of the number of complete blocks within strata will increase concomitantly, which will nullify the benefits of stratification [3]. Trials with more than 4 strata, however, have become the trend, because of the need to conduct multi-center trials as well as the increasing numbers of biomarkers linked to numerous types of diseases [4]. Thus, alternative methods to prevent imbalance for prognostic factors is in great need in modern clinical trials.

Minimization is an algorithm that dynamically allocates subjects into treatment arms [5–6]. For a new subject to be allocated, an imbalance score is computed for each treatment arm, based on levels of prognostic factor(s) of previously allocated subjects and the treatment each of them is assigned to. The new subject then will be assigned to the treatment arm with the lowest imbalance score with a pre-chosen allocation probability ranging from 0.5 to 1 (for a simple example for illustration, see [7]). Since the middle of 1980s, utilization of minimization in clinical trials has been gradually increasing. The proportion of randomized trials using randomization from 1999 to 2008 is over twice as much as that from 1989 to 1998 [8].

Even though minimization is gaining popularity in the community of clinical trials, people have been challenging the validity of standard tests routinely employed for statistical inferences in completely randomized studies to analyzing data derived from minimization, because such tests assume complete random treatment assignments [9]. For the first three decades after the introduction of minimization, very little work has been done to reveal theoretical properties of statistical inference associated with minimization [10]. However, there are a few simulation-based studies comparing the validity and power of tests between simple randomization and minimization. Forsythe used a simple simulation model, $y_{ij} = T_j + bx_j + e_{ij}$ (where j is the two treatment groups, T is the mean for each treatment and x is the subject-level covariate), with aims to illustrate the validity of minimization in cross sectional analysis of clinical studies [11]. It was shown if the covariate used for minimization

algorithm is adjusted for in the outcome analyses, the minimization technique could yield valid test of significance. Besides, the analysis of covariance provides slightly more statistical power with minimization than with simple randomization [11]. More recently, similar findings were reported with other simulation models in the context of clinical trials for acute stroke and rectal cancer, respectively [12–13]. In an attempt to provide more solid theoretical ground for minimization, Shao and Yu focused on two-sample t-test and proved the test was valid if a correctly specified model between outcomes and covariates was used for adjustment in the test [14]. It was also shown that with the treatment assignment being generated through minimization, the simple two-sample t-test, without adjusting any covariate, was conservative in terms of its type I error [14]. Shao and Yu then extended their work to prove that in generalized linear model with even unknown link functions; those conclusions drawn in simple two-sample t-tests still held true [15].

Longitudinal data arise frequently in clinical research, in both observational studies and controlled clinical trials. In a longitudinal clinical trial on chronic diseases such as Alzheimer's disease (AD), efficacy outcomes are usually measured repeatedly over the course of the study. Therefore, a key characteristic of a longitudinal study is that observations within the same subject are usually correlated, which stimulates the development of many unique statistical methods for the analysis of longitudinal data. In addition, since there are multiple measurements over time from the same subjects in a clinical trial, the efficacy of the treatment can be determined based on the effects at the endpoint, or through some summary measure over time such as the rate of change of an outcome measure. Another feature of longitudinal studies lies in the fact that missing data and dropouts are very common so that such issues need to be taken into consideration for reaching reliable statistical conclusions from the data. However, the afore-mentioned questions on the validity and power of utilizing minimization in clinical trials had only been addressed through cross-sectional analyses, i.e., on the last-observation-carried-forward (LOCF) endpoint. Cross-sectional analysis of clinical trials on LOCF has been widely known to be less than optimal, and sometimes even biased [16]. Hence, given that all clinical trials on chronic diseases are longitudinal in nature, the most appropriate analyses for such trials should be based on adequate longitudinal statistical models. For example, the efficacy outcome can be analyzed in two different ways: using the repeated measures ANOVA, or using the random intercept and slope model. Further, limited work in the literature thus far on the theoretical foundation for the validity of statistical inferences in the context of minimization has only been established primarily on cross-sectional analyses [14]. It is therefore important to examine how statistical inferences based on various longitudinal statistical models as well as different options of treating outcome measures (repeated raw measures, vs. repeated changes from baseline) are affected by the method of subject allocations: the minimization algorithm vs. simple randomization, especially in small to mid-sized clinical trials.

The goal of the present study is to investigate the validity and power of some of the well-used longitudinal statistical models using different options of outcome measures (repeated raw measures, vs. repeated changes from baseline) when minimization is used for subject allocation in simulated clinical trials. This paper is organized as follows. In Section 2, the motivating study, The Dominantly Inherited Alzheimer Network (DIAN) trial, is described.

The longitudinal statistical models and simulation scenarios are laid out in Section 3. In Section 4, data sets simulated based on these models are analyzed, followed by more simulation studies using parameters estimated from DIAN database in Section 5.

2. The DIAN clinical trial

Alzheimer's disease (AD) is an age-related brain-damaging disorder that results in progressive cognitive and functional impairment and death. An estimated 5.4 million Americans are now living with AD, and that number will rise to as many as 16 million by 2050 [Alzheimer's Association: http://www.alz.org/downloads/Facts_Figures_2014.pdf]. The Dominantly Inherited Alzheimer Network (DIAN; U19 AG032438) was launched in 2008 to establish an international, multicenter registry of individuals at risk or with a known causative mutation of AD in the amyloid precursor protein (*APP*), presenilin 1 (*PSEN1*), or presenilin 2 (*PSEN2*) genes [17]. The DIAN is an observational study, which evaluates participants at entry and longitudinally thereafter with clinical and cognitive batteries, structural, functional, metabolic, and amyloid imaging protocols, and biological fluid (blood and cerebrospinal fluid) collection with the goal of determining the sequence of changes in pre-symptomatic mutation carriers who are destined to develop AD. The DIAN is now well established, having enrolled approximately 400 participants. Interim cross-sectional analyses indicate a cascade of AD biomarker changes that begin at least 20 years before symptomatic onset of disease [18]. Although people having mutations on any of the three genes, *APP*, *PSEN1* and *PSEN2*, are only about 1 percent of AD cases worldwide, the scientific findings from the unique cohort will have important implications for understanding the disease progression and designing clinical trials to test the efficacy of novel therapies for treating other forms of AD.

One major objective of DIAN, as mandated by the NIH when it funded the study, is to implement clinical trials to test a wide range of potential therapeutic agents for their efficacy in terms of both biomarker endpoints as well as cognitive and functional endpoints. For this purpose, DIAN Trial Unit (TU) was established [19]. In December 2012, the first person was enrolled in the first clinical trial of DIAN. Building on the data from the DIAN, the ongoing DIAN trial will measure the effects of drugs on a comprehensive set of AD biomarkers and imaging markers (e.g., amyloid deposition, cerebrospinal fluid (CSF) A β and tau, magnetic resonance imaging (MRI) brain atrophy, and positron emission tomography (PET) imaging with 2-[¹⁸F] fluoro-2-deoxy-D-glucose (FDG)) to determine if a drug is likely to have a cognitive benefit in a subsequent trial with cognitive as the endpoint. To this end, an adaptive design was used for the DIAN trial in which candidate drugs will be first tested for their targeted biomarker efficacy based on the intended mechanism of action for the drugs during the initial biomarker phase. The drugs that show a good biomarker engagement will then be tested for their cognitive efficacy during the following and much larger cognitive phase. Two different active drugs are currently tested in the biomarker phase of the DIAN trial in comparison to a shared placebo. The primary efficacy outcomes of the biomarker phase are CSF biomarkers and PET amyloid imaging markers. The current cross sectional and longitudinal data from DIAN play a crucial role in designing the DIAN trial. For example, data from the DIAN have demonstrated that the estimated years to onset (EYO), computed as the difference in age between the subjects' age

and their affected parents' age of onset [18], is a major covariate affecting changes in biomarker and cognitive outcomes. Due to the limited sample size of individuals carrying these rare mutations, the DIAN trial has to enroll mutation carriers with a wide interval of EYO, e.g., from -15 y to 10 y, so that a reasonable statistical power can be achieved to test the biomarker efficacy during the initial stage. Because of these, it is important to balance the treatment allocations across different intervals of EYO. Hence, the current DIAN trial protocol for the biomarker phase implements a minimization algorithm that assigns DIAN participants into one of the 3 treatments arms with a 1:1:1 ratio, and at the same time balances the treatment allocations across different levels of EYO as well as other covariates.

Given the uncertainty of the biomarker changes as a function of time, the biomarker phase of the DIAN trial plans to examine biomarker changes over time through standard longitudinal statistical models. Given the relatively small sample size in the biomarker phase of the DIAN trial, one major concern in proposing the statistical analysis plan for the DIAN trial is whether the statistical inferences from standard longitudinal models (i.e., those based on general linear mixed effects models) are still valid when a dynamic randomization such as minimization algorithm is used for subject allocation. In addition, it is important to know whether the repeated raw efficacy measures or the repeated change of efficacy measures from the baseline provides better statistical power for efficacy comparison in longitudinal analyses. Further, the observational study of DIAN has already witnessed different degrees of attrition over time across biomarker modalities, similar attrition rates may be expected in the DIAN trial. It is therefore also important to assess how missing data might further affect the inferences of the longitudinal analyses on the DIAN trial, especially given that a dynamic randomization such as minimization algorithm has been implemented.

Hence, we design the current study to address these questions with simulated clinical trial data from longitudinal statistical models with a wide spectrum of hypothetical parameters as well as parameters directly estimated from the existing database of the DIAN.

3. Statistical models and Simulation Scenarios

3.1. Statistical models

We consider two longitudinal models that can be commonly used to analyze repeated measures in a clinical trial, such as the ongoing trial of the DIAN TU.

Model (1) (Repeated measures analysis of covariance: RM ANCOVA):

$$y_{ijt} = \mu_{it} + \gamma x_{ij} + p_{ij} + e_{ijt} \quad (1)$$

y_{ijt} is the response for subject j in treatment group i at time t . x is the prognostic variable that is used in the minimization algorithm. γ is a fixed parameter that measures the change in y when there is a change in x . $i = 0$ or 1 , is the indicator for treatment arms, j is subject indicator within each treatment arm, and $t = 0, 1, 2, 3$ is indicating the time points ($t = 0, 1, 2$ for DIAN trial simulation). μ is the mean for all subjects in treatment arm i at time t when $x = 0$. p represents random effect between subjects, simulated from standard normal distribution. e stands for within-subject

random error or residual, simulated as normally distributed with zero mean and a standard deviation of 0.5. In addition, a covariance matrix

$$\begin{bmatrix} 0.5 & 0.15 & 0.045 & 0.0135 \\ 0.15 & 0.5 & 0.15 & 0.045 \\ 0.045 & 0.15 & 0.5 & 0.15 \\ 0.0135 & 0.045 & 0.3 & 0.5 \end{bmatrix}$$

was introduced to the within-subject random error e in order to simulate data with autoregressive covariance structure with order 1 (AR(1)). Such autoregressive model is referred to as Model (1.1) hereafter.

Model (2) (Random intercept and random slope):

$$y_{ijt} = (\beta_{0i} + \gamma_0 x_{ij} + e_{0ij}) + (\beta_{1i} + \gamma_1 x_{ij} + e_{1ij})t + e_{ijt}. \quad (2)$$

Indicators i, j, t are the same as in Model (1). Prognostic factor x serves as a fixed effect in both the intercept and slope. β_0 and β_1 are fixed parameters in the intercept and slope, respectively, while e_0 and e_1 are random effects in the intercept and slope, respectively, simulated from standard normal distribution. e is the within-subject residual or error, simulated as normally distributed with zero mean and a standard deviation of 0.5.

3.2. Simulation of prognostic factor and treatment allocation

Two simulation studies are independently conducted using both Model (1) and (2): one simulated longitudinal data from a hypothetical clinical trial with a treatment and a placebo arm, and the other simulated data for the ongoing DIAN trial using the available data from the DIAN observational study.

3.2.1. Simulation of hypothetical clinical trials—5000 clinical trials each with a total sample size of 40 were simulated. In each trial, subjects were assigned to two treatment arms. A continuous variable x was generated as a prognostic factor, whose values were simulated with a standard normal distribution. x was further categorized into eight categories: $(-\infty, -1.5]$, $(-1.5, -1]$, $(-1, -0.5]$, $(-0.5, 0]$, $(0, 0.5]$, $(0.5, 1]$, $(1, 1.5]$, $(1.5, \infty)$. The categorized x , designated as x_{cat} hereafter, was used as the factor for minimization in the R package, Subject Randomization System (SRS). In SRS, a series of options could be chosen for calculating the imbalance score of treatment allocations, and the probability to assign a subject to treatment arms could also be customized. In this study, range was used as the imbalance score and 2/3 was used as the assigning probability. For treatment allocation by simple randomization, in order to keep the number of subjects balanced across treatment arms, block permutations, as a form of simple randomization, were implemented by SAS® Proc Plan [20] with four as the fixed block size. The above-mentioned two longitudinal models (1) and (2) were used to simulate outcomes of trials. The residual errors in both models were simulated as normally distributed with mean zero and a standard deviation of 0.5, and were independent of the random effects in these models.

3.2.2. Simulation of the ongoing DIAN trial with parameters estimated from the current database of DIAN—

From the data freeze 8 of DIAN as of June 30, 2014, individuals with mutation in at least one of the three target genes were selected for further analyses as these are the likely participants of the ongoing clinical trial of DIAN TU. Because the ongoing DIAN trial only enrolls subjects whose baseline clinical dementia rating (CDR) is less or equal to 1 and whose baseline age is from 15 years younger to 10 years older than the age of onset of their affected parent (i.e. Estimated Years to Onset (EYO) is between $[-15, 10]$), our analysis was also restricted to these inclusion criteria. The mean of outcomes of 6 brain regions from a typical brain PET scan using a compound known as Pittsburgh Compound-B (PIB) was used as the outcome in the simulations [21]. There are 154 participants who meet the DIAN trial inclusion criteria. One hundred and sixteen participants have baseline PIB assessment. In addition, 44 and 21 of these subjects have had longitudinal assessments of PIB for follow-up year 1 and 2, respectively.

RM ANCOVA was first used to analyze the participants with longitudinal PIB, using PIB and EYO as the outcome variable and the covariate being adjusted for, respectively. The estimates for adjusted mean of PIB at baseline and year 1 and 2 were: 2.443, 2.510, and 2.586. And the estimate on the fixed effect (i.e. γ in Model (1)) of EYO was 0.028. The between and within subject variances were 1.103 and 0.042, respectively.

The random intercept and slope model was next fitted to the same dataset with the same outcome (PIB) and covariate (EYO). Using random interception and slope model, the estimates for the mean intercept (i.e. β_{0i} of Model (2)) and slope on time (i.e. β_{1i}) were 2.453 and 0.059, the gamma values of EYO (γ_0) and EYO $\times$ time (γ_1) were 0.040 and -0.018 , and the variances of residual, intercept and slope were 0.020, 0.998, and 0.0191, with covariance between intercept and slope being 0.080.

The estimates obtained from the DIAN data freeze 8 were then used to simulate for the placebo arm of clinical trials on DIAN, and we assumed that the active treatment arm would share the estimated variance components as well as the effect of the prognostic factor (EYO), while allowing an additional fixed effect for the active treatment. To simulate prognostic factor (i.e., EYO) in 5000 trials with 40 total participants in each trial, EYO values of the 154 participants in the DIAN dataset were repeatedly selected with SAS® PROC SURVEYSELECT procedure (with equal probability for selecting each subject). Then EYO was categorized into 4 groups for the implementation of minimization algorithm (Table 1). For the simulation of repeated measures of PIB, the estimated variance components from Model (1) and Model (2) were used to first simulate random effects and errors. The estimated fixed effects along with the simulated EYO were then used to simulate PIB values for the placebo arm. A similar process was used to simulate the outcome for the active treatment arm, while allowing a set of specified treatment effects.

3.2.3. Simulation of missing data—

There are several mechanisms of missing data, such as missing completely at random (MCAR), missing at random (MAR), and missing not at random or non-ignorable (NI). MCAR was assumed in the present study. For the simulated hypothetical trials, missing data were simulated with a dropout rate of 10% for each of the three time points after baseline. If a subject was set as missing at a given time

point, observations at subsequent time points were also set as missing. Thus, the accumulative missing rate in the simulated hypothetical trial was about 20% and 30% for time point three and four, respectively. For the simulated DIAN trials, the dropout rate of 20% was used. Thus, the two follow-up visits had 20% and 40% missing rate, respectively.

4. Analyses of simulated hypothetical trials

Minimization was used to avoid serious covariate imbalance between the treatment groups in our simulated hypothetical trials. We first compared covariate imbalance of simulated datasets across treatment arms with minimization and simple randomization as methods of treatment assignment, respectively. In each of our simulated samples ($N=40$ for each sample, and 5000 samples in total), we compared the categorized prognostic factor (x_{cat}) between the two treatment arms using Fisher's exact test. P-values are indicating the chance of x_{cat} being different in the two arms. Table 2 shows the degree of imbalance in x_{cat} based on 5000 simulations, using Fisher's exact test on the equality of covariate distribution across two treatment arms. With simple randomization, the p-value from 5000 Fisher's exact tests is expected to be uniformly distributed between 0 and 1. Our results showed that the 75 percentile of p-values for x_{cat} was around 0.75, while the mean (50 percentile) of p-values was around 0.5 (Table 2). Compared to simple randomization, however, minimization had much higher p-values at each percentile presented in Table 2, suggesting that it is less likely to have imbalance in terms of x_{cat} across treatment arms.

To compare inferences with and without the covariate in the longitudinal models, the repeated measure (RM) ANOVA without covariate x and ANCOVA with covariate x were used on each of the 5000 simulated trials to approximate the type I error and statistical power. All these analyses were done in SAS®, using PROC MIXED procedure. Time points are treated as repeated measures in the REPEATED statement and compound symmetry was used as the covariance type. Figure 1 and Table 3 show the percentages of simulated trials rejecting the null hypothesis ($\mu_{03} = \mu_{13}$) with a range of true mean difference at the first time point of the trial (i.e., the time point for the efficacy comparison between two treatment arms). When there is no treatment effect, i.e. $\mu_{03} = \mu_{13}$ in simulation, the test is considered valid if it achieves a rejection rate around 5% at the designated nominal significance level of 0.05. Various values of γ which determines the association between outcome y and covariate x in clinical trials simulated by Model (1), were selected to test the validity of statistical tests when subjects are allocated by a minimization algorithm. As shown in Figure 1, not surprisingly, if simple randomization was used in simulation, the type I error was always around 5% with all values of γ , regardless of whether x was adjusted for or not. However, if minimization was used and x was not adjusted in analyses, i.e. RM ANOVA was used, type I error decreased and became too conservative as γ increased. Such conservativeness in type I error was not present if RM ANCOVA was used, adjusting for x in the analyses.

When the null hypothesis ($\mu_{03} = \mu_{13}$) is not true, i.e., when differences between treatment arms were introduced in simulation, percentages of 5000 simulated trials rejecting the null hypothesis were computed as empirical power. The power of the four efficacy tests from the RM ANOVA and RM ANCOVA with either simple randomization or minimization as the method of subject allocation was presented in Table 3 with a wide range of effect size. The

results indicated that RM ANCOVA had higher power than RM ANOVA for efficacy comparison regardless of the method of subject allocation. These differences were tested when the effect size equals 0.35 (the smallest size to achieve at least 80% power, as highlighted in Table 3). A chi-squared test was used to compare the power from two different scenarios and it was found the power for RM ANCOVA with minimization was not significantly greater than that of RM ANOVA with minimization (82.5% vs. 81.0%, $p = 0.055$). The difference between RM ANCOVA and RM ANOVA with simple randomization was not significant, either (81.6% vs. 80.2%, $p = 0.079$). However, the p values of both these tests were at the border line of 0.05. A larger size of simulation may result with statistical significance. It is also noteworthy that minimization had slightly greater power than randomization under both RM ANCOVA and RM ANOVA, especially when the effect size was greater than 0.1. However, the differences between minimization and randomization were not significant with either RM ANCOVA or RM ANOVA ($p = 0.252$ and 0.324 , for RM ANCOVA and RM ANOVA, respectively). In order to find out if the above findings are applicable to other simulation scenarios, two modifications were made and tested, 1) 0.01, instead of 0.05, was chosen as the designated nominal significance level, 2) sample size for each of those 5000 simulations was increased to 60. Under such modifications, similar results of type I error could be obtained, as shown in supplement figure 1 and 2, respectively. Meanwhile, similar patterns of power between minimization and simple randomization could also be observed, as shown in supplement table 1 and 2.

Many clinical trials utilize the difference of measurements between follow-up time points and baseline as outcomes. In order to simulate such trials, a new variable ($y\_diff_{ijt} = y_{ijt} - y_{j0}$) was generated in our simulated dataset for each subject, where $i = 0$ or 1 , is the treatment arms, j is the subject indicator, and $t = 1, 2, 3$ denotes follow-up time points. Then, RM ANOVA and RM ANCOVA were implemented on y_diff after baseline, testing the efficacy hypotheses on whether the mean of y_diff at the last time is the same between the two treatment arms. As expected, because the impact of x on y was subtracted out when y_diff was used as the outcome, power calculated from models with or without adjusting for x were almost identical (supplement table 3). In addition, because y_diff had lower variance than the original y , when the same effect size was assumed, statistical tests on y_diff generated much higher power than the tests on the original y (Figure 2).

Simulated data from Model (2) were also analyzed similarly with SAS® PROC MIXED procedure. Intercept and time are included in the RADOM statement and unstructured was used as the variance type. If minimization was used, but the covariate used in the minimization algorithm was not adjusted in analyses, type I error decreased and became too conservative as γ_1 of Model (2) is increased (supplement figure 3). Such conservativeness in type I error was not present if x was adjusted for in the analyses (supplement figure 3). In addition, the following two conclusions reached by Model (1) still held true in analyses resulted from Model (2): 1) adjusting for x increased power, regardless of whether the treatment allocation is based on minimization or simple randomization (supplement Table 4); 2) when x was adjusted for in the analyses, minimization performed slightly better than simple randomization with respect to the statistical power (supplement Table 4). However, when y_diff was used as the outcome, it appeared to have slightly lower power than original y in Model (2) (supplement figure 6).

As we did for model 1, two modifications regarding the designated nominal significance level (from 0.05 to 0.01) and sample size (from 40 to 60) for each of those 5000 simulations were made. Similar results of type I error could be obtained, as shown in supplement figure 4 and 5, respectively. Besides, results of power at various effect size are also shown in supplement table 5 and 6, which are largely the same as those in supplement table 4.

Two other situations were further taken into consideration in our simulation study, 1) missing data caused by dropouts in longitudinal clinical trials, and 2) possible unequal allocation of participants to treatment arms either by design (i.e., when more subjects are randomized to active drug to encourage enrollment) or by chance. In the further simulation study, 10% dropouts were simulated assuming MCAR. In addition, 2:1 allocation (treatment : control) was introduced to each simulated clinical trial. Datasets with dropouts or unequal allocation were then analyzed for type I error rate and power, and the results were compared with those of complete datasets with equal allocation, as shown in Figure 3 and 4, respectively. Similar to the results as shown above, with $\gamma = 0.3$, trials with missing data and trials with unequal allocation both maintain 5% type I error rate. Besides, RM ANCOVA had greater power than RM ANOVA regardless of the methods of treatment assignment. Minimization had slightly greater power than randomization under both RM ANCOVA and RM ANOVA (supplement Table 7 and 8). As compared to complete dataset with the same simulation parameters, data sets with missing or unbalanced treatment allocation resulted in lower power across all the effect size tested, especially when effect size was greater than 15% (Figure 3 and 4, and supplement table 7 and 8).

In the end, another covariance structure besides compound symmetry, Autoregressive with order 1 (AR(1)), was also used for simulations of Model (1). Similar to results shown in Figure 1, if minimization was used but not adjusted for, type I error decreased and became too conservative as γ in Model (1) increased in the AR(1) model (supplement figure 7). Besides, statistical tests on y_diff had higher power than the tests on the original y (supplement figure 8). In addition, ANCOVA still generated slightly more power than ANOVA, regardless of which allocation method was used (supplement Table 9). However, contrary to the results of compound symmetry model (as shown in Table 3), the power of analysis on minimization-associated data is slightly lower than that of analysis on data generated by simple randomization (supplement Table 9).

5. Analyses of simulated DIAN trial

Among a variety of baseline characteristics of participants enrolled in the DIAN trial, the Estimated Years to Onset (EYO) for participant is of particular importance. It is critical to have EYO well balanced between treatment arms. Table 1 shows the distributions of EYO, grouped into 4 groups to be used in the minimization process. In addition to EYO, another important variable for the trial is PET imaging with PIB as the tracer, which provides quantitative information on amyloid deposits, a pathological hallmark in Alzheimer's disease [22].

The ongoing DIAN TU trial is designed for 2 annual follow-up assessments after baseline. Baseline EYO and PIB were used as the prognostic factor and the outcome variable,

respectively. To simulate 5000 DIAN trials with Model (1), the DIAN longitudinal database was analyzed to estimated parameters in Model (1). 40 was assumed to be the total sample size per simulated trial. EYO and PIB were designated as x and y in Model (1), respectively. RM ANOVA/ANCOVA were used to analyze each simulated DIAN trial, with the null hypothesis, $\mu_{02} = \mu_{12}$. With either minimization or simple randomization, the type I error rates are all around 5% in both RM ANOVA/ANCOVA analyses (Table 4). The power of RM ANCOVA was similar to that of RM ANOVA for both minimization and simple randomization (Table 4). With effect size at 40%, the power for ANCOVA with minimization is slightly greater than that of ANOVA with minimization, however, the difference is not significant (85.0% vs. 84.0%, $p = 0.167$). For RM ANOVA, the power between minimization and simple randomization were not significantly different (85.4% vs. 84.0%, $p = 0.052$). When changes from baseline on PIB were used as the outcome in Model (1), the power was much higher than that of using original PIB at each follow-up visit (Figure 5). In addition, DIAN trials were also simulated with missing values. As compared to complete dataset with the same simulation parameters, data sets with 20% dropout rate (at MCAR mechanism) resulted in slightly lower power (supplement figure 9).

Although the random intercept and slope model, e.g., Model (2), does not fit the DIAN longitudinal data as well as Model (1), as indicated by AICs for the two models, for demonstration purpose, we also analyzed DIAN longitudinal database on the rate of change of PIB across follow-up visits and used the parameter estimates to simulate another 5000 DIAN trials using Model (2). However, the power was too small even when the effect size of 100% was assumed, regardless of which allocation method was used or whether EYO was adjusted in the model. In order to be still able to compare the impact of minimization on DIAN trial with this model, a serial of large effect sizes ranging from 100% to 2000% were assumed. Under such extreme conditions, it could be shown that the difference between minimization and simple randomization is marginal and adjusting for covariate x did not increase power (Table 5).

6. Discussion

If participants of a longitudinal clinical trial are allocated into treatment arms by a minimization algorithm on a prognostic factor which is strongly correlated with the outcome, that factor must be included as a covariate to achieve a valid test of significance in the longitudinal analyses using the repeated measures (RM) ANCOVA or random intercept and slope models. This conclusion in the present report is consistent with those reached from previous studies focusing on cross-sectional analyses (i.e., using LOCF as the efficacy endpoint by ignoring intermediate longitudinal data) of clinical trial data [11–13]. When the association between the prognostic factor and the outcome is not so strong (i.e., γ in Model (1) is small), both RM ANOVA and RM ANCOVA are valid in the sense the nominal significance level of 5% is reasonably preserved (Figure 1). In the situations when analyses with or without including the covariate are both preserving the nominal type I error rate, including the covariate in the longitudinal model always resulted in higher statistical power. When the minimization algorithm is compared to the simple randomization in subject allocation, in some of our simulations, the former is slightly more powerful than the latter; however, such difference is not to a large degree. In the other situations, statistical power in

clinical trials implementing a minimization algorithm is almost indistinguishable from those using simple randomization, whether the covariate is adjusted for or not.

We implemented both RM ANCOVA (Model (1)) and random intercept and slope model (Model (2)) to simulate both hypothetical clinical trials and the ongoing DIAN TU trial. Whereas Model (1) allows any longitudinal pattern over time, Model (2) requires a linear growth pattern for each individual. Our analysis using RM ANCOVA on DIAN longitudinal data showed largely consistent findings to those from analyses on simulated hypothetical trials, as regard to the validity and power. We also considered two options of outcome measures in the longitudinal analyses that can be done on clinical trials data: using the repeated measures over time, and using the repeated changes from baseline. The finding that repeated changes from baseline yielded higher power than the original repeated measures is not surprising. In RM ANCOVA, a subject level random effect (between-subject effect) was typically included in the model on the repeated measures, in addition to the random errors (within-subject effect) that captures the time to time variation within the same subjects. When changes from baseline are computed for each subject, the subject-level random effect is subtracted, leaving only the within-subject random errors in the analyses. Because the between-subject variance is usually higher than the within-subject variance, better statistical power will be achieved with longitudinal analysis on changes from baseline. Whereas the assumption of MCAR is in general not reasonable in studies of late onset AD, the DIAN clinical trials are prevention trials on young and middle aged individuals who are cognitively normal at baseline, and the assumption of MCAR may not be severely violated. For example, many of these individuals remain employed and have a very busy and fully functioning life. As a matter of fact, our DIAN observational study tracked the reasons of dropouts, and the data indicated that the absolute majority of missing data were not cognition-related, but due to the reasons related to subjects' employment, family and/or other health issues, or relocation. While we expect that our findings remain valid when the assumption of MCAR is relaxed to MAR, further work is needed to confirm this, and also to examine the effect of non-ignorable missing data.

Using non-extreme allocation probabilities to assign subjects to treatment arms to minimize imbalance of the covariate(s) maintains unpredictability. Weir and Lees tested a variety of allocation probabilities increasing from 0.85 to 1 and showed progressively greater balances [12]. As a tradeoff, the predictability of the next subject to a treatment increased. In this study, a relatively small allocation probability of 2/3 was used throughout all simulations. In the future, we plan to explore how the validity of longitudinal statistical inferences may be affected with more extreme allocation probability in the minimization algorithm. Further, as compared to stratified randomization, minimization is advantageous as it could take multiple prognostic factors into consideration simultaneously. In the present study on DIAN TU trial, only EYO was chosen as the only prognostic factor for minimization. In fact, there are several other potential or even more well-established risk factors such as age, the presence of the Apolipoprotein E-epsilon4 allele and education are available in the dataset [23–24]. In the future, we will also extend the study by including a relatively large number of factors as covariates in the minimization algorithm. Besides, for longitudinal clinical trials, there may be risk factors whose impact on the efficacy outcomes may change over time. Special attention is needed in the longitudinal analysis of such clinical trials when a minimization

algorithm is implemented. Finally, although minimization achieves marginal balance for each stratification variable, it may suffer from drawback of large within-stratum imbalances. Particularly, when multiple factors are considered for minimization, a specific stratum may be severely unbalanced, which will make the subgroup analysis impossible [25]. To overcome such a potential drawback of minimization, other randomization procedures were proposed as alternatives to minimization [25–26]. How these alternative algorithms will affect the validity and power of longitudinal analyses of clinical trials remains another direction in our future effort.

In conclusion, results presented here represent one of the first studies focusing on the validity and power of longitudinal statistical models on longitudinal clinical trials utilizing minimization. The findings indicate the validity maintains for longitudinal analyses of clinical trials with minimization without losing power, as long as the factor used in minimization is adjusted for in the longitudinal models. Although the ultimate proof to validate longitudinal statistical models with minimization awaits more complex theoretical development, results of the present study may serve as an initial evidence for assuring the application of minimization in the longitudinal analyses of small to mid-sized clinical trials, especially for those (e.g., DIAN trial) that are currently ongoing and have already implemented a minimization algorithm for subject allocations.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

References

1. Vickers AJ. How to randomize. *J Soc Integr Oncol*. 2006; 4:194–8. [PubMed: 17022927]
2. Kernan WN, Viscoli CM, Makuch RW, Brass LM, Horwitz RI. Stratified randomization for clinical trials. *J Clin Epidemiol*. 1999; 52:19–26. [PubMed: 9973070]
3. Tu D, Shalay K, Pater J. Adjustment of treatment effect for covariates in clinical trials: statistical and regulatory issues. *Drug Inf J*. 2000; 34:511–523.
4. Kiddle SJ, Sattlecker M, Proitsi P, Simmons A, Westman E, Bazenet C, Nelson SK, Williams S, Hodges A, Johnston C, Soininen H, Kloszewska I, Mecocci P, Tsolaki M, Vellas B, Newhouse S, Lovestone S, Dobson RJ. Candidate blood proteome markers of Alzheimer's disease onset and progression: a systematic review and replication study. *J Alzheimers Dis*. 2014; 38:515–31. [PubMed: 24121966]
5. Taves DR. Minimization: a new method of assigning patients to treatment and control groups. *Clin Pharmacol Therap*. 1974; 15:443–453. [PubMed: 4597226]
6. Pocock SJ, Simon R. Sequential Treatment Assignment with Balancing for Prognostic Factors in the Controlled Clinical Trial. *Biometrics*. 1975; 31:103–115. [PubMed: 1100130]
7. Levy DE, Emily A. Minimization: Testing A Dynamic Randomization Algorithm. *Northeast SAS Users Group*. 2005; 17:1–15.
8. Taves DR. The use of minimization in clinical trials. *Contemp Clin Trials*. 2010; 31:180–4. [PubMed: 20060500]
9. Kahan BC, Morris TP. Improper analysis of trials randomized using stratified blocks or minimisation. *Stat Med*. 2012; 31:328–40. [PubMed: 22139891]
10. Rosenberger WF, Sverdlov O. Handling Covariates in the Design of Clinical Trials. *Statist. Sci*. 2008;404–419.
11. Forsythe AB. Validity and power of tests when groups have been balanced for prognostic factors. *Computational Statistics & Data Analysis*. 1987; 5:193–200.

12. Weir CJ, Lees KR. Comparison of stratification and adaptive methods for treatment allocation in an acute stroke clinical trial. *Stat Med*. 2003; 15:705–26.
13. Hagino A, Hamada C, Yoshimura I, Ohashi Y, Sakamoto J, Nakazato H. Statistical comparison of random allocation methods in cancer clinical trials. *Control Clin Trials*. 2004; 25:572–84. [PubMed: 15588744]
14. Shao J, Yu X. A theory for testing hypotheses under covariate-adaptive randomization. *Biometrika*. 2010; 97:347–360.
15. Shao J, Yu X. Validity of tests under covariate-adaptive biased coin randomization and generalized linear models. *Biometrics*. 2013; 69(4):960–9. DOI: 10.1111/biom.12062 [PubMed: 23848580]
16. Mallinckrodt CH, Sanger TM, Dubé S, DeBrotta DJ, Molenberghs G, Carroll RJ, Potter WZ, Tollefson GD. Assessing and interpreting treatment effects in longitudinal clinical trials with missing data. *Biol Psychiatry*. 2003; 15:754–60.
17. Morris JC, Aisen PS, Bateman RJ, Benzinger TLS, Cairns NJ, Fagan AM, Ghetti B, Goate AM, Holtzman DM, Klunk WE, McDade E, Marcus DS, Martins RN, Masters CL, Mayeux R, Oliver A, Quaid K, Ringman JM, Rossor MN, Salloway S, Schofield PR, Selsor N, Sperling RA, Weiner MW, Xiong C, Moulder KL, Buckles VD. Developing an international network for Alzheimer's research: the Dominantly Inherited Alzheimer Network. *Clin. Invest*. 2012; 2:975–984.
18. Bateman RJ, Xiong C, Benzinger TL, Fagan AM, Goate A, Fox NC, Marcus DS, Cairns NJ, Xie X, Blazey TM, Holtzman DM, Santacruz A, Buckles V, Oliver A, Moulder K, Aisen PS, Ghetti B, Klunk WE, McDade E, Martins RN, Masters CL, Mayeux R, Ringman JM, Rossor MN, Schofield PR, Sperling RA, Salloway S, Morris JC. Clinical and biomarker changes in dominantly inherited Alzheimer's disease. *N Engl J Med*. 2012; 367:795–804. [PubMed: 22784036]
19. Mills SM, Mallmann J, Santacruz AM, Fuqua A, Carril M, Aisen PS, Althage MC, Belyew S, Benzinger TL, Brooks WS, Buckles VD, Cairns NJ, Clifford D, Danek A, Fagan AM, Farlow M, Fox N, Ghetti B, Goate AM, Heinrichs D, Hornbeck R, Jack C, Jucker M, Klunk WE, Marcus DS, Martins RN, Masters CM, Mayeux R, McDade E, Morris JC, Oliver A, Ringman JM, Rossor MN, Salloway S, Schofield PR, Snider J, Snyder P, Sperling RA, Stewart C, Thomas RG, Xiong C, Bateman RJ. Preclinical trials in autosomal dominant AD: implementation of the DIAN-TU trial. *Rev Neurol (Paris)*. 2013; 169(10):737–43. [PubMed: 24016464]
20. Deng C, Graz J. Generating Randomization Schedules Using SAS® Programming. SAS user group international. 2002; 27:267.
21. Benzinger TL, Blazey T, Jack CR Jr, Koeppe RA, Su Y, Xiong C, Raichle ME, Snyder AZ, Ances BM, Bateman RJ, Cairns NJ, Fagan AM, Goate A, Marcus DS, Aisen PS, Christensen JJ, Ercole L, Hornbeck RC, Farrar AM, Aldea P, Jasielec MS, Owen CJ, Xie X, Mayeux R, Brickman A, McDade E, Klunk W, Mathis CA, Ringman J, Thompson PM, Ghetti B, Saykin AJ, Sperling RA, Johnson KA, Salloway S, Correia S, Schofield PR, Masters CL, Rowe C, Villemagne VL, Martins R, Ourselin S, Rossor MN, Fox NC, Cash DM, Weiner MW, Holtzman DM, Buckles VD, Moulder K, Morris JC. Regional variability of imaging biomarkers in autosomal dominant Alzheimer's disease. *Proc Natl Acad Sci U S A*. 2013; 110(47):E4502–9. DOI: 10.1073/pnas.1317918110 [PubMed: 24194552]
22. Hardy J. Has the amyloid cascade hypothesis for Alzheimer's disease been proved? *Curr Alzheimer Res*. 2006; 3:71–73. [PubMed: 16472206]
23. Tilvis RS, Kähönen-Väre MH, Jolkkonen J, Valvanne J, Pitkala KH, Strandberg TE. Predictors of cognitive decline and mortality of aged people over a 10-year period. *J Gerontol A Biol Sci Med Sci*. 2004; 59:268–74. [PubMed: 15031312]
24. Flicker LJ. Modifiable lifestyle risk factors for Alzheimer's disease. *Alzheimers Dis*. 2010; 20:803–11.
25. Signorini DF, Leung O, Simes RJ, Beller E, Gebiski VJ, Gallagher T. Dynamic balanced randomization for clinical trials. *Stat Med*. 1993:2343–2350. [PubMed: 8134737]
26. Hu Y, Hu F. Asymptotic properties of covariate-adaptive randomization. *The Annals of Statistics*. 2012; 40:1794–1815.

Figure. 1.

Empirical type I error of efficacy tests over 5000 simulated trials from repeated measure Analyses of Variance (RM ANOVA) and Covariance (RM ANCOVA) with simple randomization and minimization algorithm as a function of gamma (γ) of Model (1), with 0.05 as the designated nominal significance level..

Figure. 2.

Comparison of type I error and power achieved by Repeated Measures Analyses of Covariance (RM ANCOVA) between two different outcomes (original y vs. y_diff). Data simulated from Model (1) ($\gamma = 0.3$).

Figure. 3.

Comparison of type I error and power achieved by Repeated Measures Analyses of Covariance (RM ANCOVA) between trials without missing data and with 10% dropout rate (Missing Completely at Random mechanism, MCAR). Data simulated from Model (1) ($\gamma = 0.3$).

Figure. 4.

Comparison of type I error and power achieved by Repeated Measures Analyses of Covariance (RM ANCOVA) between trials with equal and unequal treatment allocation (2:1, treatment:control). Data simulated from Model (1) ($\gamma = 0.3$).

Figure. 5.

Comparison of type I error and power achieved by Repeated Measures Analyses of Covariance (RM ANCOVA) between two different outcomes (original PIB or PIB change from baseline). Data simulated from Model (1) using estimated parameters on DIAN cohort.

Table 1

Baseline distribution of estimated years to onset (EYO) from the DIAN database. Figures represent numbers of participants and percentage of participants.

variable	Total N = 154	
EYO (in years)	-15 to -11	31 (20.1)
	-10 to -5	34 (22.1)
	-4 to 1	61 (39.6)
	2 to 10	28 (18.2)

Table 2

Balance in the marginal distribution of x_cat of simulated trials. Figures in table show 5th, 25th, 50th, and 75 percentiles of p-values of Fisher's exact test from 5000 simulation data sets (sample size $N = 40$ per trial) for two different methods of subject allocation.

	minimization	randomization
75 percentile	0.928	0.770
50 percentile (mean)	0.785	0.523
25 percentile	0.554	0.270
5 percentile	0.217	0.058

Table 3

Repeated measures Analyses of Variance and Covariance of simulated theoretical trials with 40 patient per trial, with 0.05 as the designated nominal significance level. Figures represent empirical type I error and power of 5000 simulations. (with $\gamma = 0.3$ for Model (1)).

Effect size (%)	RM ANCOVA		RM ANOVA	
	randomization	minimization	randomization	minimization
0	5.9%	5.1%	5.5%	4.5%
5	7.6%	7.9%	7.4%	7.1%
10	13.7%	14.2%	13.4%	14.0%
15	24.1%	24.9%	23.0%	23.6%
20	38.7%	39.2%	36.9%	37.9%
25	54.8%	55.9%	52.3%	53.8%
30	70.2%	70.4%	68.3%	68.6%
35	81.6%	82.5%	80.2%	81.0%
40	90.6%	91.1%	89.5%	90.0%
45	95.9%	96.5%	95.2%	95.8%

Table 4

Repeated measures Analyses of Variance and Covariance of simulated DIAN trials with 40 patient per trial (Model (1)). Figures represent type I error and power of 5000 simulations.

effect size (%)	RM ANCOVA		RM ANOVA	
	randomization	minimization	randomization	minimization
0	5.5%	4.8%	5.3%	5.0%
5	6.7%	7.5%	6.9%	7.4%
10	11.2%	13.1%	11.6%	13.1%
15	19.9%	21.6%	20.1%	21.6%
20	32.0%	33.2%	32.1%	33.4%
25	46.0%	46.1%	46.4%	46.9%
30	59.5%	60.6%	60.1%	60.9%
35	73.0%	74.0%	74.0%	74.7%
40	84.0%	85.0%	84.0%	85.4%
45	91.1%	91.9%	91.3%	92.0%
50	96.0%	96.2%	96.2%	96.3%

Table 5

Repeated measures Analyses of Variance and Covariance of simulated DIAN trials with 40 patient per trial (Model (2)). Figures represent type I error and power of 5000 simulations.

effect size (%)	adjusting for EYO		without adjustment	
	randomization	minimization	randomization	minimization
100	4.5%	4.5%	4.6%	4.5%
200	6.1%	6.0%	6.3%	6.1%
300	8.7%	8.1%	8.8%	8.2%
400	12.7%	11.7%	12.8%	11.8%
500	17.4%	17.2%	18.1%	17.3%
600	23.1%	23.0%	24.0%	23.7%
700	29.8%	30.4%	30.7%	30.9%
800	37.5%	37.6%	38.6%	38.3%
900	45.9%	46.4%	46.7%	47.2%
1000	54.6%	55.2%	56.2%	56.1%
1100	63.0%	63.7%	64.7%	64.4%
1200	71.0%	71.8%	72.3%	72.6%
1300	78.1%	79.2%	79.2%	79.8%
1400	84.1%	84.8%	85.0%	85.4%
1500	88.7%	89.6%	89.5%	90.2%
1600	92.2%	93.2%	92.9%	93.7%
1700	95.0%	96.0%	95.5%	96.2%
1800	96.8%	97.5%	97.2%	97.8%
1900	98.1%	98.5%	98.2%	98.7%