Featured Article

Two-period linear mixed effects models to analyze clinical trials with run-in data when the primary outcome is continuous: Applications to Alzheimer's disease

Guoqiao Wang[a,*], Andrew J. Aschenbrenner[b], Yan Li[b], Eric McDade[b], Lei Liu[a], Tammie L. S. Benzinger[c], Randall J. Bateman[b], John C. Morris[b], Jason J. Hassenstab[b,d], Chengjie Xiong[a]

[a]Division of Biostatistics, Washington University School of Medicine, St. Louis, MO, USA
[b]Department of Neurology, Washington University School of Medicine, St. Louis, MO, USA
[c]Department of Radiology, Washington University School of Medicine, St. Louis, MO, USA
[d]Department of Psychological and Brain Sciences, Washington University School of Medicine, St. Louis, MO, USA

Abstract

Introduction: Study outcomes can be measured repeatedly based on the clinical trial protocol before randomization during what is known as the "run-in" period. However, it has not been established how best to incorporate run-in data into the primary analysis of the trial.

Methods: We proposed two-period (run-in period and randomization period) linear mixed effects models to simultaneously model the run-in data and the postrandomization data.

Results: Compared with the traditional models, the two-period linear mixed effects models can increase the power up to 15% and yield similar power for both unequal randomization and equal randomization.

Discussion: Given that analysis of run-in data using the two-period linear mixed effects models allows more participants (unequal randomization) to be on the active treatment with similar power to that of the equal-randomization trials, it may reduce the dropout by assigning more participants to the active treatment and thus improve the efficiency of AD clinical trials.

© 2019 The Authors. Published by Elsevier Inc. on behalf of the Alzheimer's Association. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

Keywords: Alzheimer's disease; Linear mixed effects model; Run-in clinical trials; Unequal randomization; Two-period models

1. Introduction

To facilitate the development of disease-modifying therapies for Alzheimer's disease (AD), trial-ready cohorts have been established where participants provide longitudinal measurements on clinical, cognitive, or other measures while investigational drugs are being identified [1,2]. In this prerandomization period, the primary end points for the future clinical trials, such as clinical or cognitive tests, are assessed based on the master protocol of the platform trials allowing for easy incorporation of the prerandomization data into the primary analysis. This longitudinal period before randomization is historically referred to as the run-in period during which potential participants who have met all entry criteria for a randomized clinical trial are assigned no regiment or the same regimen (e.g., placebo) [3]. Planning a run-in period before randomization has been extensively implemented in many landmark clinical trials [4–7] including trials for AD [7], and it is expected to continue to be an essential design element [8]. The run-in design has been implemented in the dominantly

inherited Alzheimer network (DIAN) trial unit platform trial [1] and the European Prevention of Alzheimer's Dementia Proof of Concept Platform [2]. In these settings, each participant's duration and the number of primary end point assessments in the run-in period may vary and depend on the timing of enrollment.

The assessments of the primary outcome collected during run-in can potentially be used in the primary efficacy analysis at the end of the clinical trials. However, it has not been fully established how best to incorporate run-in data into final analyses. When only a single assessment is collected in the run-in period, the run-in data are often used as a covariate in the primary analysis model [9], whereas when multiple assessments are available, the rate of change (slope) in the run-in period can be used as a covariate [10] within linear mixed effects (LMEs) models or mixed effects models for repeated measures frameworks. Although these methods are helpful, they did not fully take advantage of the run-in data especially when multiple run-in assessments are present. In addition, when the run-in duration varies by individual, the variability of the run-in data over time is not fully accounted for. In AD clinical trials, the primary end points are continuous and the primary efficacy inference is based on the slowing of the rate of decline in cognition. For these types of end points, we propose a two-period (run-in period and randomization period) LME model to simultaneously model the run-in data and the randomization data. We investigated the behavior of the two-period LME by simulating clinical trials using parameters estimated from the DIAN study and evaluated the gain in power compared with the LME models using run-in data (baseline or rate of change) as a covariate.

The remainder of this article is as follows. Section 2 presents the model formulations of the LME with a covariate and the two-period LME. Section 3 evaluates model behavior through simulated hypothetical clinical trials. Section 4 presents the power formulas, and Section 5 presents the discussion.

2. Methods

2.1. Using information from run-in period as a covariate

As mentioned, the traditional model to analyze clinical trials with run-in data is LME model. The baseline assessment or the rate of change estimated using the run-in assessments will be included in the LME as a covariate. This traditional model can be expressed as follows.

Let y_{ijk} denote the longitudinal assessments for subject i at time t_{ij} for treatment group k, and it can be modeled as

$$y_{ijk} = \mu_0 + u_{0i} + \beta_1 * X_{1i} + \beta_2 * X_{1i} * t_{ij} + (\mu_{1k} + u_{1i}) * t_{ij} + \varepsilon_{ij} \tag{1}$$

where u_{0i}, u_{1i} are the random effects for the intercept and the slope and follow a bivariate normal distribution

$$\begin{pmatrix} u_{0i} \\ u_{1i} \end{pmatrix} \sim N \left(0, \begin{bmatrix} \sigma_{u_{0i}}^2 & \sigma_{u_{0i}u_{1i}} \\ \sigma_{u_{0i}u_{1i}} & \sigma_{u_{1i}}^2 \end{bmatrix} \right);$$

the residual follows normal distributions $\varepsilon_{ij} \sim N(0, \sigma_e^2)$, β's are the coefficients associated with the corresponding covariate X_{1i}, μ_0 is the baseline group mean and is assumed to be the same for the treatment group and the placebo group because of randomization, μ_{1k} represents the rate of change, $i = 1, 2, \ldots, n, j = 0, 1, \ldots, n_i$, and $k = 1, 2$ represents the placebo group and the treatment group. The primary efficacy test is to compare the rate of change of the treatment group (μ_{12}) to that of the placebo group (μ_{11}) during the randomization period.

2.2. Two-period LME

We propose the two-period LME to model the run-in period and the randomization period simultaneously. We investigate two scenarios: the slope of the placebo group in the run-in period is the same as (scenario 1) or is different from (scenario 2) that in the postrandomization period of the placebo group.

2.2.1. Scenario 1

When the slopes are the same, the two-period LME model can be presented as

$$y_{ijk} = \begin{cases} \mu_0 + u_{0i} + (\mu_1 + u_{1i}) * t_{ij} + \varepsilon_{ij}, & t_{ij} \leq t_{ibl} \\ \mu_0 + u_{0i} + (\mu_1 + u_{1i}) * t_{ibl} + (\mu_1 + \Delta\mu_k + u_{1i}) * (t_{ij} - t_{ibl})_+ + \varepsilon_{ij}, & t_{ij} > t_{ibl} \end{cases}, \tag{2}$$

where $\Delta\mu_k$ represents the treatment effect and equals to 0 for the placebo group; t_{ibl} represents the baseline time of the randomization period; $(t_{ij} - t_{ibl})_+ = \max(t_{ij} - t_{ibl}, 0)$; $j = 0, 1, \ldots, bl, bl + 1, bl + 2, bl + 3, \ldots; t_{i0} = 0$ represents the baseline of the run-in period; μ_0, u_{0i}, u_{1i}, and ε_{ij} are defined in the same way as in Section 2.1; μ_1 is the slope of the placebo group in the run-in period and the randomization period.

2.2.2. Scenario 2

Similarly, when the slopes are different, the two-period LME model can be presented as

$$\begin{bmatrix} \sigma^2_{u_{0i}} & \sigma_{u_{0i}u_{1i}} \\ \sigma_{u_{0i}u_{1i}} & \sigma^2_{u_{1i}} \end{bmatrix}$$

$$y_{ijk} = \begin{cases} \mu_0 + u_{0i} + (\mu_1 + u_{1i}) * t_{ij} + \varepsilon_{ij}, & t_{ij} \le t_{ibl} \\ \mu_0 + u_{0i} + (\mu_1 + u_{1i}) * t_{ibl} + (\mu_2 + \Delta\mu_k + u_{2i}) * (t_{ij} - t_{ibl})_+ + \varepsilon_{ij}, & t_{ij} > t_{ibl} \end{cases}, \quad (3)$$

where μ_1 and μ_2 are the slopes of the placebo arm during the run-in period and the randomization period; $\Delta\mu_k$, $(t_{ij} - t_{ibl})_+$, and t_{ibl} are defined as in equation (2); μ_0 and ε_{ij} are defined in the same way as in Section 2.1, whereas u_{0i}, u_{1i}, and u_{2i} follow a multivariate normal distribution:

$$\begin{pmatrix} u_{0i} \\ u_{1i} \\ u_{2i} \end{pmatrix} \sim N \left(\begin{array}{c} 0 \\ 0 \\ 0 \end{array}, \begin{bmatrix} \sigma^2_{u_{0i}} & \sigma_{u_{0i}u_{1i}} & \sigma_{u_{0i}u_{2i}} \\ \sigma_{u_{0i}u_{1i}} & \sigma^2_{u_{1i}} & \sigma_{u_{1i}u_{2i}} \\ \sigma_{u_{0i}u_{2i}} & \sigma_{u_{1i}u_{2i}} & \sigma^2_{u_{2i}} \end{bmatrix} \right).$$

The duration of the run-in period could be different for each individual, and there can be multiple assessments during the run-in period.

3. Evaluation of the behavior of various LMEs

3.1. Participants from DIAN study

The DIAN study is an international, longitudinal observational study established in 2008. As of June 2018 it has enrolled 529 participants from families with confirmation of a causal autosomal dominant Alzheimer's disease mutation and a 50% chance of inheriting the mutation. The details of participants' demographics, clinical, cognitive, imaging, and biochemical measures have been reported in previous publications [11,12]. For this study, only mutation carriers were included because mutation noncarriers are healthy control subjects and are not allowed to be given any treatment. The data include DIAN quality-controlled data from July 2008 to June 2018 consisting of 310 mutation carriers. As many clinical trials use a cognitive composite score as the primary outcome [1,13], we formed a cognitive composite consisting of a digit symbol substitution task test from the Wechsler Adult Intelligence Scale-Revised [14], the Mini-Mental State Examination [15], the DIAN word list delayed recall test [16], and the Wechsler Memory Scale-Revised logical memory delayed recall test [17]. The cognitive composite is an average of the z-score of these four tests [11,12].

3.2. Power comparison

We first estimated the baseline mean (μ_0), the annual slope (μ_1), and the variance-covariance for the random intercept and the random slope

and the residual σ^2_e. Furthermore, we assume $\mu_2 = 0.9*\mu_1$, $\sigma^2_{u_{2i}} = 0.9^2\sigma^2_{u_{1i}}$, the correlation between u_{0i} and u_{2i} is 0.4, and between u_{1i} and u_{2i} is 0.8. The values of these variables are presented in Table 1.

To evaluate the advantage of the two-period model relative to the traditional LME with/without run-in data as a covariate, we simulated clinical trials based on data of the DIAN study to closely mimic AD trials. This creates four models for comparison: (1) traditional LME without run-in, (2) traditional LME with the first run-in assessment as a covariate, (3) traditional LME with the slope of change across all run-in visits included as a covariate, and (4) the two-period model with run-in. Simulation SAS codes are provided in the Supplementary Material. We simulated trials with 1:1 and 3:1 treatment to placebo randomization ratio for a total 400 patients. Overall, we make the following assumptions for our simulated trials:

- Four-year trial after randomization without/with run-in period (Fig. 1).
- Individual duration of the run-in period: uniform distribution (0.3, 1.2) (Fig. 1).
- Primary outcome measured every 0.5 year in the run-in period until the individual was randomized to the treatment, and then every 1 year in the randomization period.
- The last measurement in the run-in period is also the first one in the randomization period, and it was measured at the time of randomization regardless how far this measurement was from the last measurement in the run-in period (Fig. 1).
- The slopes of the placebo group in the run-in period and the randomization period were the same and the primary outcome was simulated based on formula (2).

Table 1
Estimated simulation parameters for the cognitive composite using the DIAN observational study

Parameter	Variance-covariance matrix			Mean
	u_{0i}	u_{1i}	u_{2i}	
u_{0i}	1.0656	0.09253	0.05674	−0.6289
u_{1i}	0.09253	0.02331	0.01678	−0.09506
u_{2i}	0.05674	0.01678	0.01888	−0.08555
σ^2_e	0.05160			

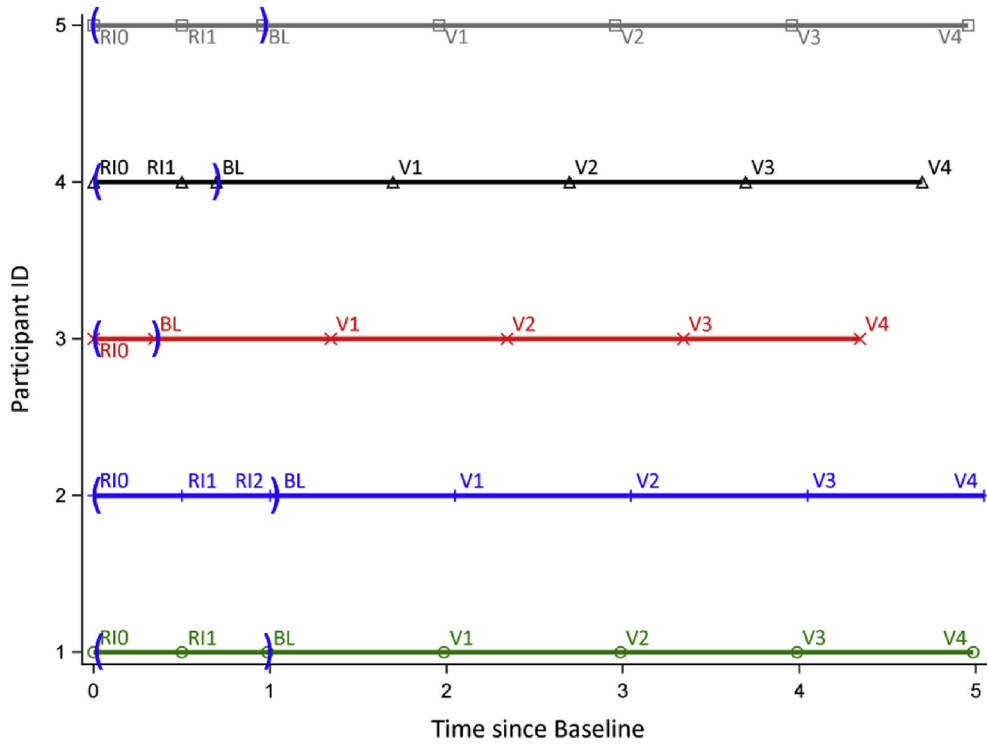Abbreviation: DIAN, dominantly inherited Alzheimer network.

Fig. 1. The run-in period and the randomization period. The run-in period was simulated using a uniform distribution (0.3, 1.2). The "BL" assessments of the randomization period were measured at the time of randomization and could be very close to the last run-in assessments (participant 2). The run-in period had at least one (participant 3) and up to three (participant 2) assessments. Abbreviation: BL, baseline.

- The slopes of the placebo group in the run-in period and the randomization period were different and the primary outcome was simulated based on formula (3).
- Effect size (% reduction in the slope): 0%, 30%, 40%, 50%, and 60%.

For each of the models mentioned previously, we simulated 1000 clinical trials, and calculated type I error and power as the proportion of 1000 simulated trials per scenario with P values less than .05. The 4-year trials without run-in were used as the anchor point to demonstrate the power improvement of run-in trials. The power/type I error comparison is presented in Figs. 2 and 3. Each figure includes the comparison among the four types of design/models with 1:1 randomization (left panel) and the comparison between the 1:1 randomization and the 3:1 randomization (right panel). Fig. 2 represents the scenario where the slope of the placebo group in the run-in period is the same as that in the randomization period, whereas Fig. 3 displays the case where the two slopes are different. For both scenarios, the type I error is well controlled for all models. The two-period LME leads up to 15% increase
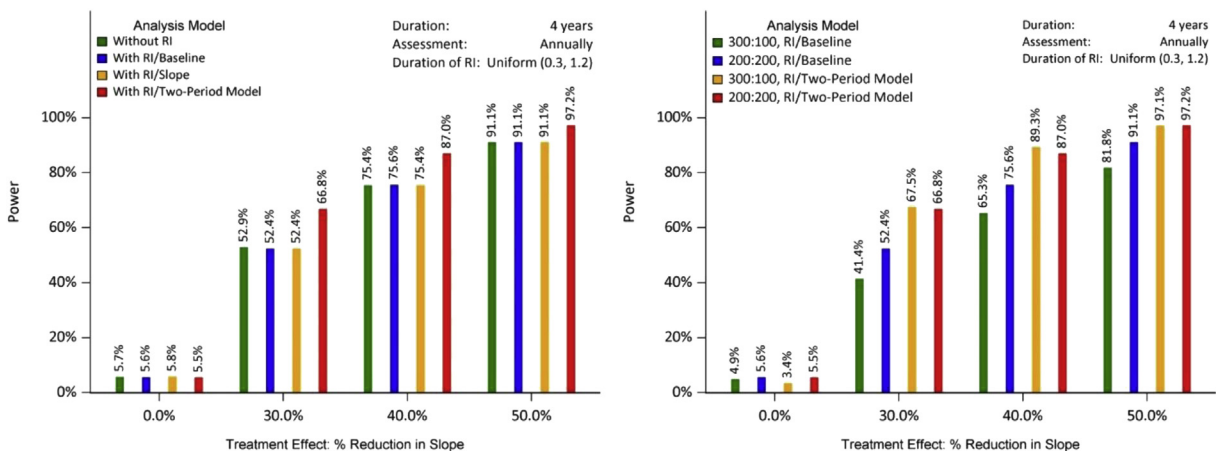

Fig. 2. Power/type I error for each design (with/without RI), different analysis models, and different randomization ratios assuming the same rate of change in the RI period and the randomization period. Sample size for the left panel: 200/arm. With RI/slope: with RI, LME with individual slope as a covariate; with RI/baseline: with RI, LME with individual baseline value as a covariate. 300:100 RI/baseline: 300 on treatment and 100 on placebo; 200:200:200 on treatment and 200 on placebo. Abbreviations: LME, linear mixed effect; RI, run-in.
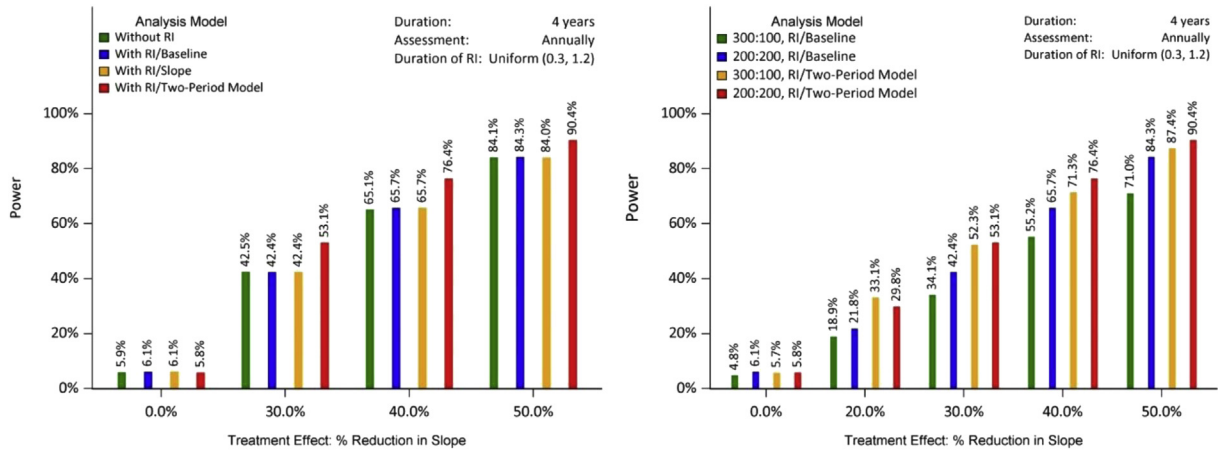
Fig. 3. Power/type I error for each design (with/without RI), different analysis models, and different randomization ratios assuming the different rates of change in the RI period and the randomization period. Sample size for the left panel: 200/arm. With RI/slope: with RI, LME with individual slope as a covariate; with RI/baseline: with RI, LME with individual baseline value as a covariate. 300:100 RI/baseline: 300 on treatment and 100 on placebo; 200:200:200 on treatment and 200 on placebo. Abbreviations: LME, linear mixed effect; RI, run-in.

in power for the same slope scenario with 1:1 randomization. When comparing the 3:1 with the 1:1 randomization, the two-period LME yields almost identical power, whereas the traditional LME yields more power for the equal randomization. For the two-slope scenario, the power improvement for the two-period model is up to 11% compared with the LME with a covariate. The 3:1 randomization has slightly less power than 1:1, but the discrepancy for two-period LME is much smaller than that for the traditional LME.

4. Power estimation of the two-period LME

Under the framework of LME, we first presented the power estimation formulas for the two-period model assuming no dropout and no intermittent missing data, then proposed the algorithm to account for the dropout.

4.1. The same slope for the placebo group in the run-in period and the randomization period

To get a closed formula, we rewrote the treatment group of equation (2) as

$$y_{ijk} = \begin{cases} \mu_0 + u_{0i} + (\mu_1 + u_{1i}) * t_{ij} + \varepsilon_{ij}, & t_{ij} \leq t_{ibl} \\ \mu_0 + u_{0i} + (\mu_1 + u_{1i}) * t_{ij} + \Delta\mu_k * (t_{ij} - t_{ibl})_+ + \varepsilon_{ij}, & t_{ij} > t_{ibl} \end{cases}$$

Further simplification yielded

$$y_{ij} = \mu_0 + u_{0i} + (\mu_1 + u_{1i}) * t_{ij} + \Delta\mu_k * (t_{ij} - t_{ibl})_+ + \varepsilon_{ij}.$$

The null hypothesis is H_0: $\Delta\mu_2 = 0$ and the alternative is H_1: $\Delta\mu_2 \neq 0$. For the fixed effects, the design matrix (X) of the treatment group is

$$X = \begin{pmatrix} 1 & t_{i0} & 0 \\ 1 & t_{i1} & 0 \\ \vdots & \vdots & \vdots \\ 1 & t_{ibl} & 0 \\ 1 & t_{ij} & t_{ij} - t_{ibl} \\ 1 & t_{ij+1} & t_{ij+1} - t_{ibl} \\ \vdots & \vdots & \vdots \\ 1 & t_{in_i} & t_{in_i} - t_{ibl} \end{pmatrix},$$

although it only includes the first two columns for the placebo group. The design matrix for the random effect also includes only the first two columns. Thus $E(Y_i|U_i) = X\beta + ZU_i$, where $\beta = \begin{pmatrix} \mu_0 \\ \mu_1 \\ \Delta\mu_2 \end{pmatrix}$ represents fixed effects, $U_i = \begin{pmatrix} u_{0i} \\ u_{1i} \end{pmatrix}$ represents the random effects, $U_i \sim N(0, G)$. The fixed effect can be estimated by: $\hat{\beta} = (X^T\Sigma^{-1}X)^{-1} X^T\Sigma^{-1}Y$, and $V(\hat{\beta}) = (X^T\Sigma^{-1}X)^{-1}$, where $\Sigma = R + ZGZ'$, R is the diagonal residual matrix. To determine the power for a complex run-in design, we adopted the same strategy as in a previous study [10]. This is to calculate the variance/standard deviation (s) for a single subject and then estimate the standard error for a given sample size. Briefly, first, using pilot data or published results, we estimated the residual variance R and the covariance of the random

intercepts and random slopes. Then plugging the design matrix X and Z for a single subject into Σ and $V(\hat{\beta})$ sequentially to estimate s for $\Delta\mu$. Next, the power for a trial with N^T subjects in the treatment group and N_P subjects in the placebo group can be determined from

$$1-\gamma=\Pr\left(\left|\frac{\Delta\mu_2}{s/\sqrt{N_T}}\right|\geq z_{\alpha/2}\middle|H_1:\quad\mu_2=\delta\right)$$
$$=\Pr\left(Z\geq z_{\alpha/2}-\frac{\delta}{s/\sqrt{N_T}}\right)$$
$$+\Pr\left(Z\leq-z_{\alpha/2}-\frac{\delta}{s/\sqrt{N_T}}\right),$$

where α is the type I error and is often set to be 5% and γ is the type II error and is often set to be 20%; z_α is upper αth quantile of the standard normal distribution.

It is noted that the variance of $\Delta\mu_2$ is estimated using all the data from the N_T+N_P subjects, but the standard error ($s\sqrt{N_T}$) is only related to N_T. Thus, theoretically, given the total sample size, the larger the N_T, the more power the run-in design has, leading to more power for the unequal randomization than the equal randomization. This benefit is attributed to two facts: (1) the same slope for the placebo group in both periods; and (2) the run-in data help estimate the slope of the placebo group and the variances of the random effects and the residuals.

4.2. Different slopes for the placebo group in the run-in period and the randomization period

In this scenario, we rewrote equation (3) as

$$y_{ijk}=\begin{cases}\mu_0+u_{0i}+(\mu_1+u_{1i})*t_{ij}+\varepsilon_{ij}, & t_{ij}\leq t_{ibl}\\\mu_0+u_{0i}+(\mu_1+u_{1i})*t_{ibl}+(\mu_{2k}+u_{2i})*(t_{ij}-t_{ibl})_++\varepsilon_{ij}, & t_{ij}>t_{ibl}\end{cases},$$

where $\mu_{2k}=\mu_2+\Delta\mu_k$, $k=1,2$ represent the placebo group and the treatment group. The null hypothesis is H_0: $\mu_{21}-\mu_{22}=0$ and the alternative is H_1: $\mu_{21}-\mu_{22}\neq0$. Then the design matrices for the fixed effects and the random effects for formula (3) are the same, and they are also the same for both groups:

$$X=\begin{pmatrix}1 & t_{i0} & 0\\1 & t_{i1} & 0\\\vdots & \vdots & \vdots\\1 & t_{ibl} & 0\\1 & t_{ibl} & t_{ij}-t_{ibl}\\1 & t_{ibl} & t_{ij+1}-t_{ibl}\\\vdots & \vdots & \vdots\\1 & t_{ibl} & t_{in_i}-t_{ibl}\end{pmatrix};$$

$$\beta_k=\begin{pmatrix}\mu_0\\\mu_1\\\mu_{2k}\end{pmatrix},\quad\text{and}\quad U_i=\begin{pmatrix}u_{0i}\\u_{1i}\\u_{2i}\end{pmatrix}.$$

Like Section 4.1, V ($\hat{\beta}$k) can be obtained for a single subject using the aforementioned formulas for Σ and for V (β). The power for a total sample size of $NT+N_P$ can be estimated from

$$1-\gamma=\Pr\left(\left|\frac{\mu_{21}-\mu_{22}}{s\sqrt{(1/N_P)+(1/N_T)}}\right|\geq z_{\alpha/2}\middle|H_1:\quad\mu_{21}-\mu_{22}=\delta\right)$$
$$=\Pr\left(Z\geq z_{\alpha/2}-\frac{\delta}{s\sqrt{(1/N_P)+(1/N_T)}}\right)$$
$$+\Pr\left(Z\leq-z_{\alpha/2}-\frac{\delta}{s\sqrt{(1/N_P)+(1/N_T)}}\right),$$

where α, γ, and z_α are defined as in Section 4.2.

4.3. Algorithm to account for dropout

For scenarios with dropout, the sample size in the power formulas can be approximated by $N_{dropout}=(N_{no-dropout}/(1-m)^n)$, where m is the annual dropout rate, n is the total duration in years, $N_{dropout}$ and $N_{no-dropout}$ are the sample sizes for each treatment group with/without dropout. This method assumes that participants who drop out before the end of study do not contribute to the estimate of the treatment effect and its variance at all, and thus will underestimate the power and overestimate the sample size. An alternative method that accounts for the contribution of the early dropout participants has been proposed in previous research [10,18]. Briefly, assuming the proportion and the sample size for each dropout pattern are p_i and n_i for a given treatment group, then the total sample size for that treatment group is approximated by [10,18]

$$N=\frac{1}{(p_1/n_1+\cdots+p_k/n_k)},$$

where k is the total number of dropout patterns for this given treatment group. This method, however, assumes no intermittent missing data within each dropout pattern, or data after the intermittent missing data do not contribute.

5. Discussion

In this article, we proposed the two-period LME model to analyze clinical trials with run-in design when the efficacy

inference is based on the rate of change. This two-period LME model offers two important benefits when compared with a traditional LME that uses measures from run-in as co-variates: (1) model the run-in data directly instead of using them as covariates; and (2) assign more participants to the active treatment without losing power compared with the traditional equal randomization clinical trials because of the fact that the run-in data serve as placebos. The first advantage allows the luxury to fully account for the run-in information in terms of the number and the frequency of assessments, and yields more accurate estimation of the variance-covariance matrix of the random effects and the within-subject error. The latter may greatly appeal to participants to enroll and remain the trials and maintain drug compliance (as they are more likely to be assigned to the treatment arm), which is especially important for diseases without any effective treatments such as AD. Furthermore, we also provided concise power estimation formulas for the two-period LME model by manipulating the design matrices of the fixed effects and the random effects. Similar manipulation of the design matrices will generalize the two-period model to other variation of run-in designs such as all participants in the run-in period are given the active treatment.

The proposed two-period model is very flexible, in that it allows the fixed effects (slopes), the random effects, and even the ancillary parameters to be different in the two periods. The flexibility can alleviate various concerns about the run-in design. For example, assuming the slope in the run-in period to be different from that in the randomization period takes care of the concern that participants may behave differently before and after randomization. Using the parameters estimated from the DIAN study, we conducted extensive simulations to evaluate the model behavior mimicking real AD clinical trials. Also we showed that the two-period LME model yielded accurate estimations of the treatment effect, controlled type I error, and led to large increases in power compared with models that used the run-in data as covariates. An additional advantage of the two-period LME is that it can be implemented using the well-established SAS procedures such as PROC NLMIXED (see Supplementary Material for details), which makes these models easier to use.

It is important to note that our focus is to propose an optimal model for analysis of run-in clinical trials, it was not our intent to compare trials with and without run-in design although we anchored the comparison based on the trials without run-in. For such comparison, extensive research has been done by Frost et al. [10]. Under the framework of LME and using three data points (one run-in assessment, baseline assessment, and one postrandomization assessment), Frost et al. demonstrated that given the same follow-up duration the run-in designs can be more efficient (requiring smaller sample size) than designs without run-in provided that true between-subject variability in the rate of change (slopes) is large relative to within-subject error [10]. Our study was inspired by theirs, but different in that the two-period LME is more general, and its power calculation formula can handle any number of assessments and any assessment schedule both in the run-in period and the randomization period. Because both studies are under the same framework, the conclusions of Frost et al. also apply to the two-period LME model. For AD clinical trials, the primary outcome is usually a cognitive test [19–21] or a composite of multiple cognitive tests [1,13]. For these cognitive outcomes, the between-subject variability in the rate of change (slopes) is typically smaller relative to within-subject error, thus given the same follow-up duration and the same sample size, trials without the run-in design should have larger power than those with run-in because the former put participants on the treatment from the beginning and the latter after the run-in period. Of course, it is always optimal to start participants on a treatment as soon as possible. In other words, a 4-year AD trial with 1 year of run-in (in which treatment only begins after the first year) is always less powerful/optimal than a 5-year AD trial without run-in (in which treatment begins from the baseline). However, our results show that if run-in data are available (e.g., from a prior observational study) or if some cognitive data can be collected when other aspects of the clinical trial are still being developed (e.g., when a drug is being finalized) then the two-period model provides an optimal way to combine run-in data with trial data to maximize the probability of detecting a significant treatment effect.

Our study has some limitations. First, the two-period LME assumes the rate of change during the follow-up is linear. Although multiple studies have shown that the decline in cognition was linear, especially within a relatively short period like 2 years [22,23], it is not clear if this linearity assumption is still true over a longer course of follow-up or under the influence of disease-modifying treatments. Second, although some clinical trials with run-in designs have been conducted, we were not able to obtain these real clinical trial data to validate the two-period LME model. Instead, we simulated clinical trials using parameters estimated from a longitudinal observational AD study to mimic real clinical trials as closely as possible.

In summary, the two-period LME model optimizes the use of run-in data, is flexible to account for design variations, can increase the power of clinical trials, and allows more participants (unequal randomization) to be on the active treatment without losing power compared with the equal-randomization trials. It may serve as a superior primary analysis model for platform clinical trials where "trial-ready" populations are enrolled in longitudinal observational studies waiting for randomization to clinical trials such as DIAN and European Prevention of Alzheimer's Dementia.

Acknowledgments

Supplementary Data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.trci.2019.07.007.

RESEARCH IN CONTEXT

1. Systematic review: We reviewed the existing literature about statistical models that can be used to analyze clinical trials with run-in design. Most methods use the run-in data as a covariate, leading to inefficient use of the run-in data.

2. Interpretation: The proposed two-period linear mixed effects models jointly model the run-in data and the double-blinded randomized data, can lead up to 15% power increase, and allow unequal randomization without losing significant power compared with equal randomization.

3. Future directions: The generalization of the two-period models to other mixed effects model such as the mixed effects model for repeat measures using time as categorical is of great interest as mixed effects model for repeat measures does not have the linearity assumption.

References

[1] Bateman RJ, Benzinger TL, Berry S, Clifford DB, Duggan C, Fagan AM, et al. The DIAN-TU next generation Alzheimer's prevention trial: adaptive design and disease progression model. Alzheimers Dement 2017;13:8–19.

[2] Solomon A, Kivipelto M, Molinuevo JL, Tom B, Ritchie CW. European Prevention of Alzheimer's Dementia Longitudinal Cohort Study (EPAD LCS): study protocol. BMJ Open 2018;8.

[3] Berger VW, Rezvani A, Makarewicz VA. Direct effect on validity of response run-in selection in clinical trials. Control Clin Trials 2003; 24:156–66.

[4] Steering Committee of the Physicians' Health Study Research Group. Final report on the aspirin component of the ongoing Physicians' Health Study. N Engl J Med 1989;321:129–35.

[5] SOLVD Investigators, Yusuf S, Pitt B, Davis CE, Hood WB, Cohn JN. Effect of enalapril on survival in patients with reduced left ventricular ejection fractions and congestive heart failure. N Engl J Med 1991; 325:293–302.

[6] SOLVD Investigators, Yusuf S, Pitt B, Davis CE, Hood WB, Cohn JN. Effect of enalapril on mortality and the development of heart failure in asymptomatic patients with reduced left ventricular ejection fractions. N Engl J Med 1992;327:685–91.

[7] Davis KL, Thal LJ, Gamzu ER, Davis CS, Woolson RF, Gracon SI, et al. A double-blind, placebo-controlled multicenter study of tacrine for Alzheimer's disease. N Engl J Med 1992;327:1253–9.

[8] Packer M. Why has a run-in period been a design element in most landmark clinical trials? Analysis of the critical role of run-in periods in drug development. J Card Fail 2017;23:697–9.

[9] Frison L, Pocock SJ. Repeated measures in clinical trials: analysis using mean summary statistics and its implications for design. Stat Med 1992;11:1685–704.

[10] Frost C, Kenward MG, Fox NC. Optimizing the design of clinical trials where the outcome is a rate. Can estimating a baseline rate in a run-in period increase efficiency? Stat Med 2008;27:3717–31.

[11] Wang G, Berry S, Xiong C, Hassenstab J, Quintana M, McDade EM, et al. A novel cognitive disease progression model for clinical trials in autosomal-dominant Alzheimer's disease. Stat Med 2018; 37:3047–55.

[12] McDade E, Wang G, Gordon BA, Hassenstab J, Benzinger TLS, Buckles V, et al. Longitudinal cognitive and biomarker changes in dominantly inherited Alzheimer disease. Neurology 2018; 91:e1295–306.

[13] Sperling RA, Rentz DM, Johnson KA, Karlawish J, Donohue M, Salmon DP, et al. The A4 study: stopping AD before symptoms begin? Sci Transl Med 2014;6:228fs213.

[14] Wechsler D. WAIS-R Manual: Wechsler Adult Intelligence Scale-Revised. New York, NY: Psychological Corp; 1981.

[15] Folstein MF, Folstein SE, McHugh PR. "Mini-mental state": a practical method for grading the cognitive state of patients for the clinician. J Psychiatr Res 1975;12:189–98.

[16] Storandt M, Balota DA, Aschenbrenner AJ, Morris JC. Clinical and psychological characteristics of the initial cohort of the Dominantly Inherited Alzheimer Network (DIAN). Neuropsychology 2014; 28:19–29.

[17] Naveh-Benjamin M. Adult age differences in memory performance: tests of an associative deficit hypothesis. J Exp Psychol Learn Mem Cogn 2000;26:1170.

[18] Dawson JD, Lagakos SW. Size and power of two-sample tests of repeated measures data. Biometrics 1993;49:1022–32.

[19] Doody RS, Thomas RG, Farlow M, Iwatsubo T, Vellas B, Joffe S, et al. Phase 3 trials of solanezumab for mild-to-moderate Alzheimer's disease. N Engl J Med 2014;370:311–21.

[20] Kennedy RE, Cutter GR, Wang G, Schneider LS. Using baseline cognitive severity for enriching Alzheimer's disease clinical trials: how does Mini-Mental State Examination predict rate of change? Alzheimers Dement 2015;1:46–52.

[21] Schneider LS, Mangialasche F, Andreasen N, Feldman H, Giacobini E, Jones R, et al. Clinical trials and late-stage drug development for Alzheimer's disease: an appraisal from 1984 to 2014. J Intern Med 2014; 275:251–83.

[22] Thomas RG, Albert M, Petersen RC, Aisen PS. Longitudinal decline in mild-to-moderate Alzheimer's disease: analyses of placebo data from clinical trials. Alzheimers Dement 2016;12:598–603.

[23] Donohue M, Aisen P. Mixed model of repeated measures versus slope models in Alzheimer's disease clinical trials. J Nutr Health Aging 2012;16:360–4.