



Published in final edited form as:

Stat Med. 2017 February 10; 36(3): 403–415. doi:10.1002/sim.7136.

Comparison of Adaptive Treatment Strategies Based on Longitudinal Outcomes in Sequential Multiple Assignment Randomized Trials

Zhiguo Li

Department of Biostatistics and Bioinformatics, Duke University, Durham, NC, USA

Abstract

In sequential multiple assignment randomized trials, longitudinal outcomes may be the most important outcomes of interest since this type of trials are usually conducted in areas of chronic diseases or conditions. We propose to use a weighted generalized estimating equation (GEE) approach to analyzing data from such type of trials for comparing two adaptive treatment strategies based on generalized linear models. Although the randomization probabilities are known, we consider estimated weights in which the randomization probabilities are replaced by their empirical estimates, and prove that the resulting weighted GEE estimator is more efficient than the estimators with true weights. The variance of the weighted GEE estimator is estimated by an empirical sandwich estimator. The time variable in the model can be linear, piece-wise linear, or more complicated forms. This provides more flexibility which is important because in the adaptive treatment setting the treatment changes over time and hence a single linear trend over the whole period of study may not be practical. Simulation results show that the weighted GEE estimators of regression coefficients are consistent regardless of the specification of the correlation structure of the longitudinal outcomes. The weighted GEE method is then applied in analyzing data from the Clinical Antipsychotic Trials of Intervention Effectiveness (CATIE).

Keywords

Adaptive treatment strategy; generalized estimating equation; generalized linear model; longitudinal data analysis; piece-wise linear model; sequential multiple assignment randomized trial

1 Introduction

The sequential multiple assignment randomized trial (SMART) [1-2] has been developed to obtain data to make inference about adaptive treatment strategies. An adaptive treatment strategy [2-4], also called a treatment strategy or simply a strategy, is a sequence of decision rules that specifies the treatment a patient receives at each stage based on patient baseline characteristics and performance on previous treatments. It arises often in treating chronic diseases or conditions since in these situations the treatment is an ongoing process involving multiple decisions over time. These decisions may include, but are not restricted to, change of treatment type or dosage, according to patient response to previous treatments in terms of efficacy, tolerance, burden and so on. For example, in treating attention deficit hyperactivity disorder (ADHD) in children, one possible adaptive treatment strategy is to start with low

dose medication, and if the child responds well then stay on low dose medication, but if there is inadequate response then the treatment is switched to low dose behavioral therapy. In a SMART, subjects may be randomized multiple times during the trial. One example of SMART is a trial conducted to assess different treatment options for ADHD [5]. In this trial, children with ADHD were initially randomly assigned to low dose behavioral therapy or medication. The response was monitored after the initiation of the treatments. If at some time during the study, which lasted 36 weeks, the predefined nonresponse criterion was met, then the child was rerandomized to either an intensification of the current treatment or the combination of the two types of treatment. In this trial, 4 embedded adaptive treatment strategies can be studied, one of which is the example adaptive treatment strategy described above. There are at most two decision points in this trial and thus two randomizations. A graphical illustration of the randomization scheme is provided, for example, in Li and Murphy [6]. Trials with more than two randomizations, an example of which is the STAR*D trial for treatment of depression [7], exist but are uncommon. In this article, we focus on trials with at most two randomizations, which are also called two-stage randomized trials. Generalization to trials with more than two stages is straightforward in principle.

Almost all of the previous work on the design and analysis of two-stage randomized trials has been focused on cases in which the primary outcome is either a single continuous outcome or a survival type outcome. For example, the work on analysis include [2] and [8-15]. The work focusing on the design of such trials include [2], [6], and [16-18]. Orellana et al. [19] proposed an estimating equation that can be used to estimate the mean outcomes under different DTRs using observational data for the single outcome case. Recently, Ertefaie et al. [20] adopted the methodology in [19] to estimate the best embedded treatment strategy using SMART data. In [21], Moodie et al. focus on inference about adaptive treatment strategies with discrete outcomes but under the Q-learning setting. Earlier work before the name SMART appeared exists, examples of which include [22-23]. However, in the area of chronic diseases or conditions in which the focus is mainly on symptoms rather than survival, longitudinal outcomes arise very often and may be more important than survival outcomes or single continuous outcomes. For example, in the ADHD trial mentioned above, a behavioral score is measured on each child weekly for 36 weeks after the initial treatment. It is an indicator of the severity of the disorder and is the basis on which the efficacy of a treatment is assessed. Another example is the Clinical Antipsychotic Trials in Intervention Effectiveness (CATIE) [24], which also involved two randomizations. In this trial, a first-generation antipsychotic, perphenazine, was compared to several atypical antipsychotics. The patients were initially randomized to perphenazine or one of four atypical antipsychotics. The initial drug may be switched due to either low efficacy or bad tolerance. Those who decided to switch treatment were further randomized to one of several antipsychotics different from the initial drug. The positive and negative syndrome scale (PANSS), a primary assessment instrument for psychopathology, was measured at month 0 (baseline), 1, and 3, and then every three months afterwards. Change of the PANSS scores signifies the change in severity of the symptoms. The drug is deemed efficacious if the PANSS scores drop significantly after taking the drug; otherwise there is no reason to believe the drug is efficacious. Hence, the change in PANSS scores after taking the drug is a direct way of measuring the efficacy of the drug. Lieberman et al. [24] used the time to all

cause discontinuation as the outcome when they compared the drugs. However, longer time to discontinuation may only be due to better tolerance which does not necessarily imply higher efficacy. If the focus is more on efficacy, the trend of change in longitudinal PANSS scores would be more relevant. Analysis of longitudinal data in SMART trials has not been well studied in the literature, except that authors in [25-26] proposed methods for estimating mean responses of treatment strategies based on longitudinal data in observational studies.

In this paper we focus on the analysis of two-stage randomized trials for comparing two adaptive treatment strategies when the outcome is longitudinal. We propose to use a weighted generalized estimating equation (GEE) approach for the inference of unknown parameters in generalized linear models for the counterfactual outcomes under the two strategies. The weights are used to take into account the fact that different subjects may be consistent with a strategy with different probabilities and the same subject may be consistent with more than one strategy (see, e.g., [6] and [8]). As the standard GEE, the weighted GEE can handle both continuous and discrete longitudinal data. Also, the weighted GEE yields consistent estimators of unknown parameters regardless of whether the correlation structure among the longitudinal outcomes is correctly specified or not. The asymptotic variance of the weighted GEE estimator is consistently estimated by an empirical sandwich estimator. Although the randomization probabilities are known in SMART trials, we consider estimated weights in which the randomization probabilities are estimated by observed proportions, resulting in more efficient estimators. The inverse probability weighting method has been used in most of the literature in SMART trials, while estimation with estimated weights has never been considered before. Moreover, in our model, the time variable can appear as linear, piece-wise linear, or more complicated forms such as polynomial. This provides more flexibility which is important because in the adaptive treatment setting the treatment changes over time and hence the rate of change of mean outcomes may also change over time. Consequently a linear model over the whole period of study may not be adequate to describe the trend of change over time. This work is novel in several aspects. At first, it is the first to consider the generalized estimating equation approach for analyzing longitudinal data in SMART studies, an area where existing work is very scarce. Second, in order to adapt the methodology to our special setting, we propose both linear models and piece-wise linear models for the longitudinal trajectories, and propose different methods for comparing treatment strategies, e.g., comparing rates of change and more importantly, the areas under curve for the two strategies compared.

The following content is arranged as follows. Section 2 describes the weighted GEE method for comparing two treatment strategies using generalized linear models. Section 3 presents results of a simulation study and in Section 4 we apply the proposed method in analyzing data from the CATIE. We conclude with a discussion in Section 5. A sketch of the proof of asymptotic results is put in the Appendix.

2 The weighted generalized estimating equation

2.1 The general method

In the typical two-stage randomized trials we consider, subjects are initially randomized to one of two treatments, denoted by $A_1 = 1$ and $A_1 = 2$. After the initial treatment, some

subjects are rerandomized to one of two second-stage treatments and some subjects are not rerandomized. Suppose the subjects who are rerandomized are randomly assigned one of two second-stage treatments, denoted by $A_2 = 1$ and $A_2 = 2$. Denote $p_1 = P(A_1 = 1)$, $p_{21} = P(A_2 = 1/A_1 = 1)$ and $p_{22} = P(A_2 = 1/A_1 = 2)$. Here the rerandomization criterion can be different in different situations. For example, in the ADHD trial mentioned above, nonresponders to the initial treatments are rerandomized to either intensification of the initial treatment or the addition of another type of treatment. In some cases, responders to the initial treatment are further randomized to one of several maintenance treatments, which is common in cancer trials. Moreover, those who are not rerandomized may stay on the initial treatment or are assigned a different treatment. For example, in some cancer trials, nonresponders to the induction therapy are not rerandomized and are all put on salvage treatment. For definiteness, we assume that only nonresponders to the initial treatment are rerandomized, and denote R to be the rerandomization indicator, i.e., $R = 1$ if the subject is rerandomized and $R = 0$ otherwise. In this type of trials, there are four embedded adaptive treatment strategies, denoted by strategy “11”, “12”, “21” and “22”, respectively. Here strategy “ jk ” means a subject starts with $A_1 = j$, and if the subject does not respond well then he or she switches to $A_2 = k$ as the second-stage treatment, for $j, k = 1, 2$.

The goal of two-stage randomized trials is to assess different treatment strategies. Specifically, we assume the purpose is to compare two strategies “ st ” and “ $s't'$ ”, where s, t, s', t' can take values 1 or 2. Suppose we observe i.i.d. data for n subjects, and the longitudinal outcome of the i th subject is observed at M_i time points. Denote $y_i = (y_{i1}, \dots, y_{iM_i})^T$ as the vector of longitudinal outcomes for subject i , for $1 \leq i \leq n$. As usual, we adopt the counterfactual outcome framework [27]. Denote $y_i^{st} = (y_{i1}^{st}, \dots, y_{iM_i}^{st})^T$ as the vector of counterfactual longitudinal outcomes for subject i if the subject, possibly contrary to the fact, followed strategy “ st ”, for $1 \leq i \leq n$ and $s, t = 1, 2$. We make the consistency assumption [3] that, if subject i actually followed strategy “ st ”, then $y_i = y_i^{st}$. Denote $\mu_{im}^{st} = E(y_{im}^{st} | d_{im}^{st})$. Denote x_i to be a vector of covariates and t_{im} to be the m th measurement time, both for subject i , for $1 \leq m \leq M_i$ and $1 \leq i \leq n$. Let d_{im}^{st} and $d_{im}^{s't'}$ be d -dimensional design vectors which may include x_i and t_{im} as components. In order to compare the two strategies, we assume a generalized linear model

$$h(\mu_{im}^{st}) = (d_{im}^{st})^T \beta, \quad \text{and} \quad h(\mu_{im}^{s't'}) = (d_{im}^{s't'})^T \beta \quad (1)$$

for $1 \leq m \leq M_i$, where $h(\cdot)$ is a known link function. Here different choices for d_{im}^{st} and $d_{im}^{s't'}$ correspond to different models and different tests for the equivalence of the two strategies. Specific choices are discussed below.

Use of weights to adjust for the unique design of two-stage randomized trials is standard (see, e.g., [6], [8] and [10]). As in [6], we define weight

$$W_{jk} = \frac{I(A_1=j)}{p_1^{2-j}(1-p_1)^{j-1}} \left\{ 1 - R + \frac{I(A_2=k)}{p_2^{2-k}(1-p_2)^{k-1}} R \right\}$$

for strategy “ jk ”, where $j, k = 1, 2$. These weights are essentially the inverse of the proportions of subjects being consistent with the strategies. Analogous to the usual GEE [29], we use the following estimating equation to estimate the parameter β in model (1) based on the observed data:

$$\sum_{(j,k) \in \{(s,t), (s',t')\}} \sum_{i=1}^n W_{jk,i} \frac{\partial (\mu_i^{jk})^T}{\partial \beta} \left(V_i^{jk} [\hat{\alpha}_{jk} \{ \beta, \hat{\phi}(\beta) \}] \right)^{-1} (y_i - \mu_i^{jk}) = 0, \quad (2)$$

where $\mu_i^{jk} = (\mu_{i1}^{jk}, \dots, \mu_{iK_i}^{jk})^T$. In the above estimating equation, the “working” covariance matrix $V_i^{jk}(\alpha_{jk})$ is defined as

$$V_i^{jk}(\alpha_{jk}) = (A_i^{jk})^{1/2} R_{jk}(\alpha_{jk}) (A_i^{jk})^{1/2} / \phi$$

where $R_{jk}(\alpha_{jk})$ is a “working” correlation matrix for $W_{jk,i}(y_i - \mu_i^{jk})$, and the l th diagonal element of the diagonal matrix A_i^{jk} / ϕ equals the variance of $W_{jk,i}(y_{il}^{jk} - \mu_{il}^{jk})$ which is a function of μ_{il}^{jk} and ϕ . If $R_{jk}(\alpha_{jk})$ is the true correlation matrix for $W_{jk,i}(y_i - \mu_i^{jk})$, then $V_i^{jk}(\alpha_{jk}) = \text{var} \{ W_{jk,i}(y_i - \mu_i^{jk}) \}$. In the above estimating equation, the parameter α_{jk} in the matrix $V_i^{jk}(\alpha_{jk})$ is estimated first with $\hat{\alpha}_{jk} \{ \beta, \hat{\phi}(\beta) \}$ for fixed β , where $\hat{\phi}(\beta)$ is an estimator for ϕ for fixed β . As in the standard GEE, different “working” correlation structures can be used. In more general situations, the consistency of the resulting GEE estimator depends on the specification of the correlation matrix (see, e.g., [28] and references cited therein). However, in our case it is true that the above estimator for β does not depend on the specification of the correlation matrix, and the estimator is the most efficient when the correlation structure is correctly specified. The equation given above is in its most general form. In many cases, it may be reasonable to assume that

$V_i^{jk}(\alpha_{jk}) = V_i^{j'k'}(\alpha_{j'k'}) = V_i(\alpha)$. For computation, the iterative algorithm of [29] which iterates between β and the nuisance parameters can be used to find the solution to (2). Finally, it is worth mentioning that in the above estimating equation, the two treatment strategies “ st ” and “ $s't'$ ” can be any two strategies embedded in the trial, including strategies with different initial treatments such as “11” and “22” and strategies with the same initial treatment such as “11” and “12”.

Although the randomization probabilities are known, using estimated weights in which p_1 , p_{21} and p_{22} are replaced by their estimates, can improve the efficiency of the weighted GEE estimator, if the estimator for $p = (p_1, p_{21}, p_{22})^T$ is efficient (see, e.g., [30]). Let \hat{p}_1 be the proportion of subjects randomized to $A_1 = 1$ and let \hat{p}_{2j} be the proportion of subjects rerandomized to $A_2 = 1$ among all subjects who receive $A_1 = j$ and who are rerandomized, for $j = 1, 2$. Denote $\hat{p} = (\hat{p}_1, \hat{p}_{21}, \hat{p}_{22})^T$. Replace p in the definition of W_{st} by \hat{p} , and denote the resulting weight as $W_{st}(\hat{p})$, for $s, t = 1, 2$. The weighted GEE with estimated weights is

$$\sum_{(j,k) \in \{(s,t), (s',t')\}} \sum_{i=1}^n W_{jk,i}(\hat{p}) \frac{\partial (\mu_i^{jk})^T}{\partial \beta} \left(V_i^{jk} \left[\hat{\alpha}_{jk} \left\{ \beta, \hat{\phi}(\beta) \right\} \right] \right)^{-1} (y_i - \mu_i^{jk}) = 0. \quad (3)$$

Denote the solutions to (2) and (3) as $(\hat{\beta}, \hat{\alpha}_1, \hat{\alpha}_2)$ and $(\hat{\beta}^*, \hat{\alpha}_1^*, \hat{\alpha}_2^*)$, respectively. Let $\varepsilon_i^{st} = y_i - \mu_i^{st}$. Let $\dot{W}_{st,i}(p) = \partial W_{st,i}(p) / \partial p$. In addition, denote $p_{tj} = P(R = 1/A_1 = j)$, which is the probability of being rerandomized given that the initial treatment is $A_1 = j$, for $j = 1, 2$. The following theorem states the asymptotic properties of the estimators $\hat{\beta}$ and $\hat{\beta}^*$ and shows that the latter is more efficient. The necessary regularity conditions are stated in the Appendix.

Proposition 1—Under some regularity conditions, we have

$$\sqrt{n} (\hat{\beta} - \beta) \rightarrow_d N \left(0, U^{-1} \Sigma U^{-1} \right),$$

and

$$\sqrt{n} (\hat{\beta}^* - \beta) \rightarrow_d N \left(0, U^{-1} \Sigma U^{-1} - U^{-1} B P B^T U^{-1} \right),$$

as $n \rightarrow \infty$, where

$$U = E \sum_{(j,k) \in \{(s,t), (s',t')\}} \frac{\partial \mu_i^{jk}}{\partial \beta} \left\{ V_i^{jk}(\alpha_{jk}) \right\}^{-1} \frac{\partial (\mu_i^{jk})^T}{\partial \beta},$$

$$\Sigma = \text{cov} \left[\sum_{(j,k) \in \{(s,t), (s',t')\}} W_{jk,i} \frac{\partial (\mu_i^{jk})^T}{\partial \beta} \left\{ V_i^{jk}(\alpha_{jk}) \right\}^{-1} \varepsilon_i^{jk} \right].$$

$$B=E \sum_{(j,k) \in \{(s,t), (s',t')\}} \frac{\partial \mu_i^{jk}}{\partial \beta} \{V_i^{jk}(\alpha_{jk})\}^{-1} \varepsilon_i^{jk} W_{jk,i}^T(p),$$

and $P = \text{diag}(p_1(1-p_1), p_{21}(1-p_{21})/p_1 p_1, p_{22}(1-p_{22})/(1-p_1)p_{22})$.

Denote $\hat{D}_i^{jk} = \partial \mu_i^{jk} / \partial \beta|_{\beta=\hat{\beta}}$, $\hat{D}_i^{*jk} = \partial \mu_i^{jk} / \partial \beta|_{\beta=\hat{\beta}^*}$, $\hat{\varepsilon}_i^{jk} = y_i - \hat{\mu}_i^{jk}$, and $\hat{\varepsilon}_i^{*jk} = y_i - \hat{\mu}_i^{*jk}$, where $\hat{\mu}_i^{jk}$ and $\hat{\mu}_i^{*jk}$ are calculated by model (1) with the unknown parameters replaced by $(\hat{\beta}, \hat{\alpha}_1, \hat{\alpha}_2)$ and $(\hat{\beta}^*, \hat{\alpha}_1^*, \hat{\alpha}_2^*)$, respectively. Denote \hat{V}_i^{jk} and \hat{V}_i^{*jk} to be $V_i[\hat{\alpha}_{jk} \{\beta, \hat{\phi}(\beta)\}]$ with β replaced by $\hat{\beta}$ and $\hat{\beta}^*$, respectively. The asymptotic covariance matrix for $\hat{\beta}$ can be consistently estimated by the empirical sandwich estimator $\hat{U}_n^{-1} \hat{\Sigma}_n \hat{U}_n^{-1}$, where

$$\hat{U}_n = \frac{1}{n} \sum_{i=1}^n \sum_{(j,k) \in \{(s,t), (s',t')\}} W_{jk,i} (\hat{D}_i^{jk})^T (\hat{V}_i^{jk})^{-1} \hat{D}_i^{jk}$$

and

$$\hat{\Sigma}_n = \frac{1}{n} \sum_{i=1}^n \left[\sum_{(j,k) \in \{(s,t), (s',t')\}} W_{jk,i} \frac{\partial (\mu_i^{jk})^T}{\partial \beta} \{V_i^{jk}(\alpha_{jk})\}^{-1} \varepsilon_i^{jk} \right]^{\otimes 2},$$

where $A^{\otimes 2} = A^T A$ for a matrix A . To estimate the asymptotic covariance matrix for $\hat{\beta}^*$, we estimate U and Σ similarly as above but with $(\hat{\beta}, \hat{\alpha}_1, \hat{\alpha}_2)$ replaced by $(\hat{\beta}^*, \hat{\alpha}_1^*, \hat{\alpha}_2^*)$. We denote the estimators by \hat{U}_n^* and $\hat{\Sigma}_n^*$, respectively. Then estimate B by

$$\hat{B}_n = \frac{1}{n} \sum_{i=1}^n \sum_{(j,k) \in \{(s,t), (s',t')\}} \hat{D}_i^{*jk} (\hat{V}_i^{*jk})^{-1} \hat{\varepsilon}_i^{*jk} W_{jk,i}(p).$$

Denote the number of subjects who receive $A_1 = j$ and who are further rerandomized to be n_{Tj} for $j = 1, 2$. The asymptotic variance of $\hat{\beta}^*$ is consistently estimated by

$$\hat{U}_n^{*-1} \hat{\Sigma}_n^* \hat{U}_n^{*-1} - \hat{U}_n^{*-1} \hat{B}_n \hat{P} \hat{B}_n^T \hat{U}_n^{*-1}, \text{ where } \hat{P} = \text{diag}(p_1(1-p_1), p_{21}(1-p_{21})/(n_{T1}/n), p_{22}(1-p_{22})/(n_{T2}/n)).$$

2.2 Specific models

The model we consider is linear or generalized linear model. Here “linear” means the transformed mean outcome is a linear function in the unknown parameters. Under this framework, however, the trajectory of the longitudinal outcomes over time can be modeled as either a linear or a nonlinear function, as described in detail below. Without loss of

generality, we assume that the two strategies compared are either “11” and “22” or “11” and “12”.

2.2.1 Linear model—At first, the simplest model for the trajectory is a linear model. This is a plausible model for comparing two strategies with different initial treatments. To be specific, assume that the interest is in comparing strategies “11” and “22”. This corresponds to

$$d_{im}^{gg} = \left(1, g - 1, t_{im}, (g - 1)t_{im}, x_i^T\right)^T \quad (4)$$

for $g = 1, 2$, and $\beta = (\beta_1, \dots, \beta_4, \beta_5^T)^T$ in the general formulation $h(\mu_{im}^{gg}) = (d_{im}^{gg})^T \beta$.

Although in randomized studies the effect of baseline covariate x_i is not of primary interest, there is the possibility and advantage to estimate it in these studies [31]. In this model, since β_4 is the interaction between the group indicator and time, i.e., the difference in the rates of change of the longitudinal outcomes over time under the two strategies, the comparison of the two strategies reduces to the test of the hypothesis of $H_0 : \beta_4 = 0$. Denote the (4,4)th

elements of $\hat{U}_n^{-1} \hat{\Sigma}_n \hat{U}_n^{-1}$ and $\hat{U}_n^{*-1} (\hat{\Sigma}_n^* - \hat{B}_n \hat{P} \hat{B}_n^T) \hat{U}_n^{*-1}$ by $\hat{\sigma}_{4,4}^2$ and $\hat{\sigma}_{4,4}^{*2}$, respectively. In order to test for $H_0 : \beta_4 = 0$, we use the test statistics

$$T_n = \frac{\sqrt{n} \hat{\beta}_4}{\hat{\sigma}_{4,4}}, \quad \text{or} \quad T_n^* = \frac{\sqrt{n} \hat{\beta}_4^*}{\hat{\sigma}_{4,4}^*},$$

for the test with true weights and estimated weights, respectively. By Theorem 1 and the consistency of $\hat{\sigma}_{4,4}$ and $\hat{\sigma}_{4,4}^*$, T_n and T_n^* both have a standard normal asymptotic distribution under H_0 . Hence, we reject H_0 when $|T_n| > z_{1-\alpha/2}$ (or when $|T_n^*| > z_{1-\alpha/2}$ if estimated weights are used) at a two-sided significance level α , where z_q denotes the upper 100 q % percentile of the standard normal distribution.

2.2.2 Piece-wise linear model—In adaptive treatment strategies, the treatment may change over time, and the change of treatment is usually triggered by patient response to previous treatments. It may be expected that the rates of change of longitudinal outcomes may be different when the patient is on different treatments. This motivates a piece-wise linear model for the trajectory of the longitudinal outcomes. Specifically, suppose that for all subjects who are rerandomized to a second-stage treatment, the second-stage treatment starts at the same time, which is denoted by t^* . Suppose that the mean longitudinal outcome changes linearly before time t^* at a rate r_1 for those who start with $A_1 = j$ and who are rerandomized to $A_2 = k$, and after time t^* , the rate of change becomes r_2 . Also, suppose that for those who start with $A_1 = j$ but are not rerandomized, the rate of change is r_3 over the whole study period. Under these assumptions, the mean counterfactual outcome for strategy “ jk ” is still a piece-wise linear function of time with a change point t^* . For comparing strategies “11” and “22”, this piece-wise linear model corresponds to

$$d_{im}^{gg} = (1, g-1, a(t_{im}), (g-1)a(t_{im}), b(t_{im}), (g-1)b(t_{im}), x_{im}^T)^T, \quad \text{for } g=1, 2, \quad (5)$$

In (1) where $a(t) = t^*I(t < t^*) + tI(t \geq t^*)$, $b(t) = (t - t^*)I(t < t^*)$, and $\beta = (\beta_1, \dots, \beta_6, \beta_7^T)^T$. Here β_2 is the difference in intercepts, β_4 is the difference in slopes before time t^* , and β_6 is the difference in slopes after time t^* , between the two strategies compared.

The above model is of particular interest in comparing two strategies with the same initial treatment, specifically, strategies “11” and “12”, when the start time for the second-stage treatments is the same for all subjects, since in that case, the linear model is impossible if there is any difference between the two strategies, while the piece-wise linear model remains a plausible model. In this case, we have to set

$$d_{im}^{1g} = (1, g-1, a(t_{im}), b(t_{im}), (g-1)b(t_{im}), x_{im}^T)^T, \quad \text{for } g=1, 2, \quad (6)$$

in (1), where $\beta = (\beta_1, \dots, \beta_5, \beta_6^T)^T$. Here β_2 is the difference in intercepts at time 0, and β_5 is the difference in slopes after time t^* , which is the parameter of interest. Note that in this model there is no interaction term $(g-1)a(t_{im})$ because the slopes must be the same before time t^* when the treatments before time t^* coincide for the two treatment strategies “11” and “12”.

For comparison of the two strategies under model (5), one possibility is to compare the mean outcomes at a fixed time point, e.g., at the end of study. Another possibility is to compare the areas under the curve (from time 0 to, say, the end of study) for the two trajectories of the mean longitudinal outcomes. First, suppose we want to test for the difference in the mean outcomes at time T which is usually the end of study period. By simple algebra, the difference in the means of the counterfactual outcomes at T between the two strategies is $\beta_2 + t^* \beta_4 + (T - t^*)\beta_6$. Ignoring β_2 , the difference resulting from difference at baseline, we test for the hypothesis $H_0^{(1)}: \lambda_1^T \beta = 0$, where $\lambda_1 = (0, 0, 0, t^*, 0, T - t^*)^T$. This amounts to comparing the differences in mean outcomes from time 0 to time T between the two strategies. We use the test statistic

$$T_{1n} = \frac{\sqrt{n} \lambda_1^T \hat{\beta}}{\sqrt{\lambda_1^T \hat{U}_n^{-1} \hat{\Sigma}_n \hat{U}_n^{-1} \lambda_1}} \quad \text{or} \quad T_{1n}^* = \frac{\sqrt{n} \lambda_1^T \hat{\beta}^*}{\sqrt{\lambda_1^T \hat{U}_n^{*-1} (\hat{\Sigma}_n^* - \hat{B}_n \hat{P} \hat{B}_n^T) \hat{U}_n^{*-1} \lambda_1}}$$

for the test with true weights and estimated weights, respectively, and reject $H_0^{(1)}$ when $|T_{1n}| > z_{1-\alpha/2}$ (or when $|T_{1n}^*| > z_{1-\alpha/2}$ if estimated weights are used) at a two-sided significance level α . Second, suppose we want to test for the difference in the areas under curve from

time 0 to T between the two strategies. The difference in the areas under curve can be calculated as

$$T\beta_2 + \frac{1}{2}t^*(2T - t^*)\beta_4 + \frac{1}{2}(T - t^*)^2\beta_6.$$

We also ignore the difference resulting from the difference at baseline and test for the hypothesis $H_0^{(2)}: \lambda_2^T \beta = 0$, where $\lambda_2 = (0, 0, 0, t^*(2T - t^*), 0, (T - t^*)^2)^T$. We use the test statistic

$$T_{2n} = \frac{\sqrt{n}\lambda_2^T \hat{\beta}}{\sqrt{\lambda_2^T \hat{U}_n^{-1} \hat{\Sigma}_n \hat{U}_n^{-1} \lambda_2}} \quad \text{or} \quad T_{2n}^* = \frac{\sqrt{n}\lambda_2^T \hat{\beta}^*}{\sqrt{\lambda_2^T \hat{U}_n^{*-1} (\hat{\Sigma}_n^* - \hat{B}_n \hat{P} \hat{B}_n^T) \hat{U}_n^{*-1} \lambda_2}}$$

for the tests with true weights and estimated weights, respectively, and reject $H_0^{(2)}$ when $|T_{2n}| > z_{1-\alpha/2}$ (or when $|T_{2n}^*| > z_{1-\alpha/2}$ if estimated weights are used) at a two-sided significance level α .

Since in model (6), the only parameter of interest is β_5 , to compare strategies “11” and “12” under this model, we test for the hypothesis $H_0: \beta_5 = 0$. This is similar to testing for $H_0: \beta_4 = 0$ in the linear model (4) thus the details are omitted.

2.2.3 Polynomial model—Another alternative model that can be considered is the polynomial model. As a starting point, the quadratic polynomial can be applied, and if higher flexibility is necessary, a cubic model may be considered. As an example, for comparing strategies “11” and “22”, to formulate a quadratic model, we let

$$d_{im}^{gg} = \left(1, g - 1, t_{im}, (g - 1)t_{im}, t_{im}^2, (g - 1)t_{im}^2, x_{im}^T\right)^T, \quad \text{for } g=1, 2,$$

where $\beta = (\beta_1, \dots, \beta_6, \beta_7^T)^T$. This type of models is of particular interest in cases where the start time of the second-stage treatment is random, which makes the piece-wise linear model less applicable because there is no logically meaningful change point. To compare two strategies under this type of models, we can compare the mean outcomes at a fixed time point (T) or compare the areas under the curve in the interval $(0, T)$, with the corresponding null hypothesis being $H_0^{(1)}: \lambda_1^T \beta$ and $H_0^{(2)}: \lambda_2^T \beta$, respectively, where $\lambda_1 = (0, 0, 0, T, 0, T^2, 0)^T$ and $\lambda_2 = (0, 0, 0, T^2/2, T^3/3, 0)^T$. The test statistics and the rejection regions are constructed in the same way as those in Section 2.2.2 and thus the details are omitted.

3 Simulation

We conduct a simulation study to assess the performance of the weighted GEE method for data analysis. The simulation is performed to compare strategies “11” and “22”. At first, we generate data from model (1), where $s = t = 1$ and $s' = t' = 2$. The link function $h(\cdot)$ is

chosen to be the identity function and the measurement error is assumed to be normally distributed. The choice for d_{im}^{gg} is (4) in one simulation and (5) in another simulation, corresponding to a linear model and a piece-wise linear model, respectively. Suppose the outcome of interest is measured at 6 time points $t_{im} = m$, for $0 \leq m \leq 5$. The change point in the piece-wise linear model is supposed to be $t^* = 2$. First, A_1 is generated from a Bernoulli distribution with probability 0.5. In the piece-wise linear model, we first generate the longitudinal outcomes up to time 2 using a linear trajectory model with a positive slope. If the measurement at $t = 2$ is greater than twice of the measurement at time $t = 0$, then R is set to be 0 and the measurements after time $t = 2$ are still generated using the same linear model as for $t \leq 2$. Otherwise, R is set to be 1 and we generate A_2 from a Bernoulli distribution with probability 0.5, and the measurements after $t = 2$ are generated using a linear model with a slope that is different from the slope before time $t = 2$, where the intercepts of the linear models are chosen to make the trajectories continuous at time $t = 2$. The one-dimensional covariate X in the model is generated from a standard normal distribution, independently from the other variables. The parameters are chosen such that the resulting β is $\beta = (2.5, -0.25, 2, 1.5, 1)^T$ for the linear model and $\beta = (2, 0, 0.5, 0.2, 0.47, 0.2, 0.5)^T$ for the piece-wise linear model. The vector of measurement errors ε_i is generated from a multivariate normal distribution where the correlation coefficient between ε_{il}^g and ε_{im}^g follows an AR(1) correlation structure: $0.6^{|t_{il} - t_{im}|}$, where t_{il} and t_{im} are the l th and m th measurement times for subject i , respectively, for $0 \leq l, m \leq 5$ and $1 \leq i \leq n$. The common variance of the components of ε_i^g is set to be 5. For the working correlation structure in the weighed GEE, we try both the “independence” correlation structure and the “true” correlation structure. Here the “independence” and “true” correlation structure mean that we replace $R_{gg}(\alpha_g)$ in the expression for $V_i^{gg}(\alpha_{gg})$ with the identity matrix and the AR(1) correlation structure, respectively. They are not really the “independence” and the “true” correlation structures because here the true correlation structure refers to that of $W_{gg,i}(y_i - \mu_i^g)$ instead of $y_i - \mu_i^g$. The correct specification of the correlation of $W_{gg,i}(y_i - \mu_i^g)$ is difficult because in simulation we generate the y_i s using a specific correlation structure but the correlation structure of $W_{gg,i}(y_i - \mu_i^g)$ is different and complicated. We run the simulation for sample sizes 50, 100 and 300, and in each simulation scenario, 1000 replications are run.

Tables 1 and 2 below present the results of this simulation. We list the biases, mean estimated variances, empirical variances, and the empirical coverage rates of the 95% confidence intervals of selected parameters of the most interest. These results show that the weighted GEE method works well, yielding low biases and consistent estimates of variances and confidence intervals with expected coverage rates. The consistency of the estimators is achieved under both of the correlation structures. Moreover, the estimators based on estimated weights are slightly more efficient than those with true weights in this simulation.

4 An Application

In this section we apply the weighted GEE method in analyzing data from the CATIE. The CATIE was conducted to determine the long-term effects and usefulness of antipsychotic medications in persons with schizophrenia [24]. In this trial, 1500 schizophrenia patients

were initially randomized to receive one of 5 antipsychotic drugs including olanzapine, perphenazine, quetiapine, risperidone, and ziprasidone. Patients may discontinue or switch treatment at some time after the initial treatment. Patients who chose to switch treatment can choose one of two further randomization pathways. In one of the pathways, patients were randomized to receive either clozapine or one of olanzapine, quetiapine or risperidone; and in the other pathway patients were randomized to receive either ziprasidone or one of olanzapine, quetiapine, and risperidone. In both pathways no one is assigned to the same drug as the initial drug. It was pointed out by [24] that, usually those who chose to switch treatment due to low efficacy were likely to choose the first randomization pathway and those who switched due to bad tolerance were likely to choose the second randomization pathway.

Results of data analysis [24] showed that olanzapine is the best treatment with respect to patient adherence among the 5 drugs. A natural further question to ask is that which treatment should a patient switch to if he or she starts with olanzapine and chooses to switch treatment. This is a problem of assessing adaptive treatment strategies. For this purpose we consider only patients who started with olanzapine. Denote the two randomization pathways as “E” and “T” (corresponding to efficacy and tolerance, respectively). For those who chose the “E” randomization pathway, denote the two options they were randomized to as $A_1 = 1$ and $A_1 = 2$, where $A_1 = 1$ means clozapine and $A_1 = 2$ means choosing one from olanzapine, quetiapine or risperidone. For those who chose the “T” randomization pathway, denote the two options as $A_2 = 1$ and $A_2 = 2$, where $A_2 = 1$ means ziprasidone and $A_2 = 2$ is the same as $A_1 = 2$. For our purpose we assume that the outcome of interest is the longitudinal PANSS scores, which were measured at months 0, 1, and 3, and then every three months afterwards up to month 18. This trial can be analyzed similarly as the typical two-stage randomization trials in the previous section. However, the weights will be different because of the difference in the trial design. Denote strategy “ jk ” to be the strategy in which a subject starts with olanzapine, and then switches to $A_1 = j$ if he or she chooses to switch treatment because of “E” and switches to $A_2 = k$ if it is because of “T”, for $j, k = 1, 2$. In the CATIE trial, 333 patients started with olanzapine, among whom 214 patients did not switch treatment during the following 18 month period and thus are consistent with all 4 treatment strategies. The remaining 119 patients switched treatment in that time period and each of them is consistent with two of the 4 treatment strategies. Denote $p_1 = P(A_1 = 1|E = 1)$ and $p_2 = P(A_2 = 1|T = 1)$ to be the randomization probabilities. Denote R to be the indicator for switching treatment in the study period. Let E to be the indicator for switching treatment due to “E” and T to be the indicator for switching treatment due to “T”. Then the weight function for strategy “ jk ” is:

$$\tilde{W}_{jk=1} = R + \frac{EI(A_1=j)}{p_1^{2-j}(1-p_1)^{j-1}} + \frac{TI(A_2=k)}{p_2^{2-k}(1-p_1)^{k-1}}, \quad \text{for } j, k=1, 2.$$

Figure 1 shows the trajectories of the mean PANSS score for a number of randomly selected patients in CATIE. There seems to be little difference in the rates of change among patients. The same is expected for different treatment strategies but formal statistical analyses need to be carried out to confirm this. Since the start time of the second-stage treatment can be

different for different patients in this trial, we use the quadratic polynomial model instead of the linear or piece-wise linear model. Based on the data from all the subjects who started with olanzapine, and using the method in Section 2.2.3, we compare the mean outcomes at month 18 as well as the areas under the curve from time 0 to month 18 for each pair of the 4 adaptive treatment strategies. The results of this analysis show that the differences between strategies are very small, with the absolute values of the estimated coefficients in the quadratic model being close to 0.01 and the p values ranging from 0.29 to 0.75 when the true weights or estimated weights are used. Obviously these differences are too small to be interesting.

5 Discussion

In the above, we mainly focused on a typical two-stage randomized trials in which only responders or nonresponders to the first-stage treatments are rerandomized to second-stage treatments. The proposed method generalizes easily to more general SMART trials, for example, trials with more than two stages, trials in which only responders to one of the two initial treatments are rerandomized, trials in which both responders and nonresponders to the initial treatments are rerandomized, and so on. The only difference is the formulae for the weights. Specific weights for some examples are illustrated in [6].

When the second-stage treatment initiates at different (random) times for different subjects, the comparison of two treatment strategies is more challenging than the other cases. In this case, the linear model is still a possible approximation of the trajectories but it is likely to fail. The piece-wise linear model may be used, but the change point is not known priori and can be different for different strategies. Hence, the piece-wise linear assumption as well as the change point (if the assumption is supported by preliminary data) need to be ascertained by preliminary data.

Although it is not the focus of this article, the proposed method can be used to estimate the best treatment strategy among all embedded strategies, for which the method in [21] can be applied as well. To use our method, we first impose an appropriate model for the longitudinal outcomes under each strategy. Then use the proposed methods and the models to estimate a summary quantity for each strategy, and the estimate of the best strategy is based on this quantity. For example, this summary quantity can be the mean outcome at a specific time point or the area under the curve, as described in Section 2.2.2.

Finally, missing data may be a problem in longitudinal studies, and for any GEE approach to yield unbiased results the missing mechanism need to be missing completely at random [32]. Some modifications (e.g., [32]), have been proposed to guarantee unbiasedness estimation under the more general missing at random mechanism. Extension of our proposed method for this purpose is a potential topic for future research.

Acknowledgements

This research was supported in part by a grant in cancer clinical trials with grant number 1P01CA142538-01 from the National Institutes of Health. The content is solely the responsibility of the author and does not necessarily represent the official views of the National Institutes of Health. The author thank an anonymous referee for the helpful comments that helped in improving this article.

Appendix: proofs of asymptotic results

For simplicity, assume $\alpha_1 = \alpha_2 = \alpha$ and $R_1(\alpha) = R_2(\alpha) = R(\alpha)$. The following are the assumptions that are used in the proof:

- A1** The parameter space for β , \mathcal{B} , is a compact set in R^p where p is the dimension of β .
- A2** The parameter space for α , \mathcal{A} , is a compact set in R^q where q is the dimension for α .
- A4** $\hat{\alpha}(\hat{\beta}, \hat{\phi}(\beta))$ is uniformly consistent for α .
- A5** The derivative of the components of $R^{-1}(\alpha)$ is continuous in \mathcal{A} , and the first and second derivative of μ_{ig} with respect to β is continuous in \mathcal{B} .
- A6** For any $\beta_1, \beta_2 \in \mathcal{B}$, $E\mu_{ig}(\cdot, d_{ig}, \beta_1) = E\mu_{ig}(\cdot, d_{ig}, \beta_2)$ if and only if $\beta_1 = \beta_2$.

At first, we assume that the true weights are used. Denote $\alpha^*(\beta) = \hat{\alpha}(\beta, \hat{\phi}(\beta))$, and

$$S_i(\beta, \alpha) = \sum_{g=1}^2 W_{gg,i} \frac{\partial \mu_{ig}^T}{\partial \beta} V_{ig}^{-1}(\alpha) (y_i - \mu_{ig}).$$

Then $\hat{\beta}$ is the solution to $\sum_{i=1}^n S_i\{\beta, \alpha^*(\beta)\} = 0$. Taylor expansion and rearranging terms yield

$$\sqrt{n}(\hat{\beta} - \beta) = U_{1n}^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^n S_i\{\beta, \alpha^*(\beta)\} + o_p(1) = 0,$$

where

$$\begin{aligned} U_{1n} &= \frac{1}{n} \sum_{i=1}^n \frac{dS_i\{\beta, \alpha^*(\beta)\}}{d\beta} \\ &= \frac{1}{n} \sum_{i=1}^n \left[\frac{\partial S_i\{\beta, \alpha^*(\beta)\}}{\partial \beta} + \frac{\partial S_i\{\beta, \alpha^*(\beta)\}}{\partial \alpha^*} \frac{\partial \alpha^*(\beta)}{\partial \beta} \right] \\ &= E \sum_{g=1}^2 \frac{\partial \mu_{ig}^T}{\partial \beta} V^{-1}(\alpha) \frac{\partial \mu_{ig}}{\partial \beta} + o_p(1) \\ &= U_1 + o_p(1). \end{aligned}$$

Furthermore,

$$\begin{aligned} \frac{1}{\sqrt{n}} \sum_{i=1}^n S_i\{\beta, \alpha^*(\beta)\} &= \frac{1}{\sqrt{n}} \sum_{i=1}^n S_i(\beta, \alpha) + \frac{1}{n} \sum_{i=1}^n \frac{\partial S_i(\beta, \alpha)}{\partial \alpha} \sqrt{n}\{\alpha^*(\beta) - \alpha\} \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^n S_i(\beta, \alpha) + o_p(1), \end{aligned}$$

where we used the facts that $\sum_{i=1}^n \partial S_i(\beta, \alpha) / \partial \alpha / n = o_p(1)$ by the law of large numbers and that $\sqrt{n} \{\alpha^*(\beta) - \alpha\} = O_p(1)$ [29]. It follows that

$$\sqrt{n}(\hat{\beta} - \beta) = U_1^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^n S_i(\beta, \alpha) + o_p(1) = 0,$$

which yields the desired result when true weights are used.

When p is estimated in the weights, Taylor expansion with arguments similar as above yields that

$$\sqrt{n}(\hat{\beta}^* - \beta) = U_1^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^n S_i(\beta, \alpha) + U_1^{-1} B \sqrt{n}(\hat{p} - p) + o_p(1) = 0.$$

Denote the number of subjects who are randomized to $A_1 = j$ and who are further rerandomized to be n_{2j} , for $j = 1, 2$. Straightforwardly, we have

$$\sqrt{n}(\hat{p} - p) = \begin{pmatrix} \sqrt{n}(\hat{p}_1 - p_1) \\ \sqrt{n/n_{21}} \sqrt{n_{21}}(\hat{p}_{21} - p_{21}) \\ \sqrt{n/n_{22}} \sqrt{n_{22}}(\hat{p}_{22} - p_{22}) \end{pmatrix} \rightarrow_d N(0, P).$$

Since \hat{p} is an efficient estimator for p , the three conditions in [33] are satisfied and the conclusion follows from (1.3) therein.

References

1. Lavori PW, Dawson R. A design for testing clinical strategies: biased individually tailored within-subject randomization. *Journal of the Royal Statistical Society A*. 2000; 163:29–38. DOI: 10.1111/1467-985X.00154.
2. Murphy SA. An experimental design for the development of adaptive treatment strategies. *Statistics in Medicine*. 2005; 24:1455–1481. DOI: 10.1002/sim.2022. [PubMed: 15586395]
3. Robins JM. A new approach to causal inference in mortality studies with sustained exposure periods-application to control of the healthy worker survivor effect. *Computers and Mathematics with Applications*. 1986; 14:1393–1512.
4. Lavori PW, Dawson R, Roth AJ. Flexible treatment strategies in chronic disease: clinical and research implications. *Biological Psychiatry*. 2000; 48:605–614. DOI: 10.1016/S0006-3223(00)00946-x. [PubMed: 11018231]
5. Nahum-Shani I, Qian M, Almira D, Pelham W, Gnagy B, Fabiano G, Waxmonsky J, Yu J, Murphy SA. Experimental design and primary data analysis methods for comparing adaptive interventions. *Psychological Methods*. 2012; 17(4):457–477. DOI: 10.1037/a0029372. [PubMed: 23025433]
6. Li Z, Murphy SA. Sample size formulae for two-stage randomized trials with survival outcomes. *Biometrika*. 2011; 98(3):503–518. DOI: 10.1093/biomet/asr019. [PubMed: 22363091]
7. Rush AJ, Fava M, Wisniewski SR, Lavori PW, Trivedi MH, Sackeim HA, Thase ME, Nierenberg AA, Quitkin FM, Kashner TM. Sequenced treatment alternatives to relieve depression (STAR*D): rationale and design. *Controlled Clinical Trials*. 2004; 25:119–142. DOI: 10.1016/S0197-2456(03)00112-0. [PubMed: 15061154]

8. Lunceford JK, Davidian M, Tsiatis AA. Estimation of survival distributions of treatment strategies in two-stage randomization designs in clinical trials. *Biometrics*. 2002; 58:48–57. DOI: 10.1111/j.0006-341X.2002.00048.x. [PubMed: 11890326]
9. Guo X, Tsiatis AA. A weighted risk set estimator for survival distributions in two-stage randomization designs with censored survival data. *The International Journal of Biostatistics*. 2005; 1(1):1–15. DOI: 10.2202/1557-4679.1000.
10. Wahed SA, Tsiatis AA. Semiparametric efficient estimation of survival distribution in two-stage randomization designs in clinical trials with censored data. *Biometrika*. 2006; 93:163–177. DOI: 10.1093/biomet/93.1.163.
11. Lokhnygina Y, Helderbrand JD. Cox regression methods for two-stage randomization designs. *Biometrics*. 2007; 63:422–428. DOI: 10.1111/j.1541-0420.2006.00707.x. [PubMed: 17425633]
12. Miyahara S, Wahed SA. Weighted Kaplan-Meier estimators for two-stage treatment regimes. *Statistics in Medicine*. 2010; 29:2581–2591. DOI: 10.1002/sim.4020. [PubMed: 20799259]
13. Ko JH, Wahed SA. Up-front vs. sequential randomizations for inference on adaptive treatment strategies. *Statistics in Medicine*. 2012; 31(1):812–830. DOI: 10.2202/1557-4679.1000. [PubMed: 22362642]
14. Wang L, Rotnitzky A, Lin X, Millikan R, Thall P. Evaluation of Viable Dynamic Treatment Regimes in a Sequentially Randomized Trial of Advanced Prostate Cancer. *Journal of the American Statistical Association*. 2012; 107(498):493–508. DOI: 10.1080/01621459.2011.641416. [PubMed: 22956855]
15. Kidwell KM, Wahed SA. Weighted log-rank statistic to compare shared-path adaptive treatment strategies. *Biostatistics*. 2013; 14(2):299–312. DOI: 10.1093/biostatistics/kxs042. [PubMed: 23178734]
16. Oetting, AI.; Levy, JA.; Weiss, RD.; Murphy, SA. Statistical methodology for a SMART design in the development of adaptive treatment strategies.. In: Shrout, PE., editor. *Causality and Psychopathology: Finding the Determinants of Disorders and their Cures*. American Psychiatric Publishing, Inc.; Arlington VA: 2007.
17. Feng W, Wahed SA. A supremum log rank test for comparing adaptive treatment strategies and corresponding sample size formula. *Biometrika*. 2008; 95(3):695–707. DOI: 10.1093/biomet/asn025.
18. Feng W, Wahed SA. Sample size for two-stage studies with maintenance therapy. *Statistics in Medicine*. 2009; 28:2028–2041. DOI: 10.1002/sim.3593. [PubMed: 19382105]
19. Orellena L, Rotnitzky A, Robins JM. Dynamic regime marginal structural mean models for estimation of optimal dynamic treatment regimes, Part I: main content. *The International Journal of Biostatistics*. 2010; 6(2) Article 8. DOI: 10.2202/1557-4679.1200.
20. Ertefaie A, Wu T, Lynch KG, Nahum-Shani I. Identifying a set that contains the best dynamic treatment regimes. *Biostatistics*. 2015 Published Online. DOI: 10.1093/biostatistics/kxv025.
21. Moodie EEM, Dean N, Sun YR. Q-learning: Flexible learning about useful utilities. *Statistics in Biosciences*. 2014; 6:223–243.
22. Thall PF, Millikan RE, Sung HG. Evaluating multiple treatment courses in clinical trials. *Statistics in Medicine*. 2000; 19:1011–1028. [PubMed: 10790677]
23. Thall PF, Sung HG, Estey EH. Selecting therapeutic strategies based on efficacy and death in multicourse clinical trials. *Journal of the American Statistical Association*. 2002; 97(457):29–39. DOI: 10.1198/016214502753479202.
24. Lieberman JA, Stroup TS, McEvoy JP, Swartz MS, Rosenheck RA, Perkins DO, Keefe RS, Davis SM, Davis CE, Lebowitz BD, Severe J, Hsiao JK. Clinical Antipsychotic Trials of Intervention Effectiveness (CATIE) Investigators. Effectiveness of antipsychotic drugs in patients with chronic schizophrenia. *New England Journal of Medicine*. 2005; 53(12):1209–1223. DOI: 10.1056/NEJMoa051688.
25. Murphy SA, van der Laan MJ, Robins JM, CPPRG. Marginal mean models for dynamic regimes. *Journal of the American Statistical Association*. 2001; 96:1410–1423. [PubMed: 20019887]
26. Murphy SA. Optimal dynamic treatment regimes (with Discussion). *Journal of the Royal Statistical Society, Series B*. 2003; 65:331–366.

27. Holland PW. Statistics and causal inference. *Journal of the American Statistical Association*. 1986; 81:945–960.
28. Vansteelandt S. On confounding, prediction and efficiency in the analysis of longitudinal data and cross-sectional clustered data. *Scandinavian Journal of Statistics*. 2007; 34:478–498.
29. Liang KY, Zeger SL. Longitudinal data analysis using generalized linear models. *Biometrika*. 1986; 81:945–960.
30. Robins JM, Rotnitzky A, Zhao LP. Estimation of regression coefficients when some regressors are not always observed. *Journal of the American Statistical Association*. 1994; 89:846–866.
31. Tsiatis AA, Davidian M, Zhang M, Lu X. Covariate adjustment for two-sample treatment comparisons in randomized clinical trials: A principled yet flexible approach. *Statistics in Medicine*. 2008; 27:4658–4677. DOI: 10.1002/sim.3113. [PubMed: 17960577]
32. Paik, Myunghee Cho. The generalized estimating equation approach when data are not missing completely at random. *Journal of the American Statistical Association*. 1997; 92(440):1320–1329. DOI: 10.2307/2965402.
33. Pierce DA. The asymptotic effect of substituting estimators for parameters in certain types of statistics. *The Annals of Statistics*. 1982; 10(2):475–478.

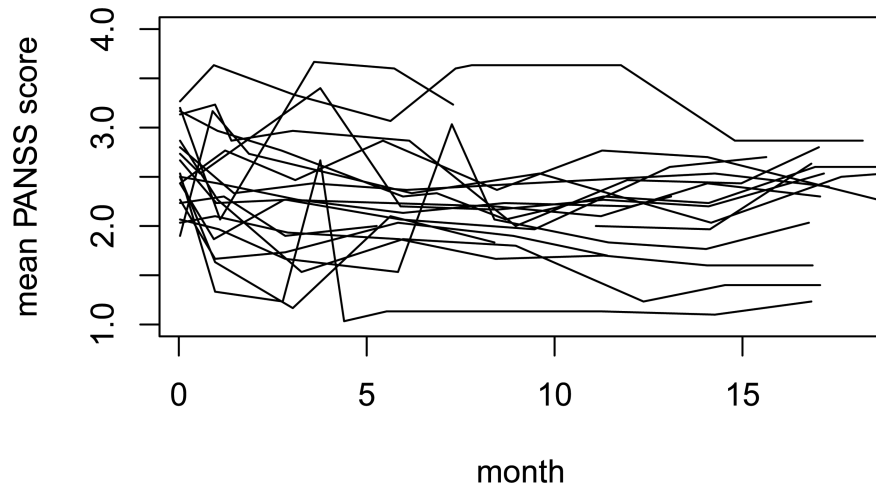


Figure 1. Plot of longitudinal mean PANSS scores for a number of randomly selected patients in the CATIE.

Table 1

Simulation results for selected parameters in the linear trajectory model.

	<u>true weights</u>				<u>estimated weights</u>			
	bias	V1	V2	coverage	bias	V1	V2	coverage
<u>“independence” working correlation structure</u>								
<i>n</i> = 50								
β_2	-0.02	0.92	0.99	0.93	-0.01	0.93	0.94	0.93
β_4	0.01	0.14	0.15	0.92	0.01	0.12	0.13	0.92
β_5	0.01	0.20	0.23	0.91	0.00	0.22	0.24	0.91
<i>n</i> = 100								
β_2	-0.03	0.47	0.47	0.95	-0.03	0.48	0.48	0.95
β_4	0.01	0.08	0.08	0.95	0.01	0.06	0.06	0.95
β_5	0.02	0.11	0.12	0.92	0.02	0.12	0.12	0.92
<i>n</i> = 300								
β_2	-0.02	0.16	0.16	0.95	-0.02	0.16	0.16	0.95
β_4	0.00	0.023	0.023	0.95	-0.00	0.018	0.019	0.95
β_5	0.01	0.04	0.04	0.94	0.01	0.03	0.03	0.94
<u>“true” working correlation structure</u>								
<i>n</i> = 50								
β_2	-0.01	0.89	0.94	0.93	-0.01	0.88	0.92	0.92
β_4	0.00	0.13	0.15	0.92	0.01	0.12	0.14	0.93
β_5	0.00	0.20	0.22	0.91	0.00	0.20	0.21	0.93
<i>n</i> = 100								
β_2	-0.03	0.45	0.45	0.95	-0.03	0.43	0.43	0.94
β_4	0.01	0.07	0.07	0.95	0.01	0.07	0.07	0.95
β_5	0.02	0.11	0.12	0.92	0.02	0.11	0.11	0.93
<i>n</i> = 300								
β_2	-0.02	0.15	0.15	0.95	-0.02	0.14	0.14	0.95
β_4	-0.00	0.022	0.023	0.95	-0.00	0.016	0.017	0.94
β_5	0.01	0.04	0.04	0.94	0.01	0.04	0.04	0.94

V1: mean estimated variance; V2: empirical variance; coverage: empirical coverage rate for the 95% confidence interval.

Table 2

Simulation results for selected parameters in the piece-wise linear trajectory model.

	<u>true weights</u>				<u>estimated weights</u>			
	bias	V1	V2	coverage	bias	V1	V2	coverage
<u>“independence” working correlation structure</u>								
<i>n</i> = 50								
β_2	0.01	0.89	0.90	0.93	-0.01	0.89	0.87	0.97
β_4	0.00	0.13	0.14	0.93	0.01	0.15	0.13	0.97
β_6	-0.01	0.15	0.16	0.94	-0.01	0.16	0.15	0.96
β_7	0.01	0.05	0.06	0.93	0.01	0.07	0.06	0.96
<i>n</i> = 100								
β_2	-0.03	0.47	0.46	0.95	-0.03	0.46	0.44	0.96
β_4	0.01	0.071	0.073	0.94	0.01	0.072	0.069	0.97
β_6	0.01	0.078	0.080	0.93	0.01	0.077	0.075	0.96
β_7	0.02	0.027	0.030	0.92	0.02	0.029	0.028	0.95
<i>n</i> = 300								
β_2	-0.02	0.16	0.16	0.95	-0.02	0.16	0.15	0.96
β_4	0.00	0.024	0.024	0.95	-0.00	0.023	0.022	0.96
β_6	0.01	0.026	0.025	0.95	0.01	0.024	0.023	0.96
β_7	0.01	0.01	0.01	0.94	0.01	0.01	0.01	0.94
<u>“true” working correlation structure</u>								
<i>n</i> = 50								
β_2	-0.01	0.86	0.88	0.93	-0.01	0.87	0.84	0.97
β_4	0.00	0.12	0.13	0.92	0.01	0.13	0.12	0.96
β_6	0.01	0.14	0.15	0.92	0.01	0.15	0.14	0.97
β_7	0.00	0.05	0.06	0.91	0.00	0.06	0.06	0.93
<i>n</i> = 100								
β_2	-0.03	0.45	0.45	0.95	-0.03	0.43	0.43	0.96
β_4	0.01	0.072	0.071	0.95	0.01	0.069	0.067	0.96
β_6	0.01	0.075	0.077	0.92	0.01	0.076	0.074	0.96
β_7	0.02	0.027	0.028	0.93	0.02	0.029	0.028	0.95
<i>n</i> = 300								
β_2	-0.02	0.15	0.15	0.95	-0.02	0.14	0.14	0.96
β_4	-0.00	0.023	0.024	0.94	-0.00	0.024	0.023	0.96
β_6	0.01	0.025	0.026	0.93	0.01	0.025	0.024	0.96
β_7	0.01	0.01	0.01	0.94	0.01	0.01	0.01	0.94

V1: mean estimated variance; V2: empirical variance; coverage: empirical coverage rate for the 95% confidence interval.