3-26-2018

ROBUST ESTIMATION OF THE AVERAGE TREATMENT EFFECT IN ALZHEIMER'S DISEASE CLINICAL TRIALS

Michael Rosenblum
Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, mrosenbl@jhsph.edu

Aidan McDermont
Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health

Elizabeth Colantuoni
Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health

Robust Estimation of the Average Treatment Effect in Alzheimer's Disease Clinical Trials

Michael Rosenblum[1], Aidan McDermott[1], and Elizabeth Colantuoni[1]

[1]Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA.

March 26, 2018

Abstract

The primary analysis of Alzheimer's disease clinical trials often involves a mixed-model repeated measure (MMRM) approach. We consider another estimator of the average treatment effect, called targeted minimum loss based estimation (TMLE). This estimator is more robust to violations of assumptions about missing data than MMRM. We compare TMLE versus MMRM by analyzing data from a completed Alzheimer's disease trial data set and by simulation studies. The simulations involved different missing data distributions, where loss to followup at a given visit could depend on baseline variables, treatment assignment, and the outcome measured at previous visits. The TMLE generally has improved robustness in our simulated settings, i.e., less bias and mean squared error, and better confidence interval coverage probability. The robustness is due to the TMLE correctly modeling the dropout distribution. We illustrate the tradeoffs between these estimators and give recommendations for how to use these estimators in practice.

1 Introduction

In Alzheimer's disease clinical trials, the mixed-model repeated measure (MMRM) approach is often used in the primary analysis (Chen et al., 2018). This method for estimating the average treatment effect adjusts for potential bias due to patient dropout by leveraging information in baseline variables and longitudinal measurements of the primary outcome. (van der Laan and Gruber, 2012) developed another statistical method for achieving this goal, called targeted minimum loss based estimation (TMLE).

We assess the performance of TMLE in estimating the average treatment effect in the setting of randomized clinical trials of individuals with mild cognitive impairment. TMLE has theoretical advantages over more commonly used estimation methods such as MMRM, including greater robustness to model misspecification. This estimator has not been compared head-to-head versus commonly used estimation methods in clinical trials of treatments for mild cognitive impairment.

In Section 2, we compare TMLE versus MMRM in analyzing data from a completed phase 3 trial of vitamin E and donepezil (Petersen et al., 2005). Implementation of the TMLE estimator is presented in Section 3. We conduct simulation studies using distributions that mimic key features from the completed trial in Section 4.

2 Data Analysis Using Completed Phase 3 Trial

We reanalyzed data from the completed phase 3 randomized trial of vitamin E and donepezil for treating mild cognitive impairment (Petersen et al., 2005). We refer to this trial as the donepezil MCI trial and we focus on the following two arms of the trial: 10mg of donepezil daily (treatment) versus placebo (control).

We focus on the Clinical Dementia Rating scale sum of boxes (CDR_{SB}), which was measured at baseline and then every 6 months up to 18 months. We define our primary outcome to be the difference between CDR_{SB} measured at 18 months and baseline. (This was a secondary outcome in the trial, but is important since it is the primary outcome for other, ongoing trials.) The average treatment effect of interest is the difference in the primary outcome comparing assignment to the donepezil arm versus placebo. The estimators described below are defined in later sections, but we wanted to first present results to show how these methods compare in the data analysis.

The unadjusted estimate (difference between sample means comparing treatment versus control for those who completed the 18 months of follow-up) of the treatment effect is -0.19 (standard error (se): 0.16, 95% Confidence Interval (CI): -0.51 to 0.11). When applying MMRM that adjusts for baseline CDR_{SB} and CDR_{SB} over time using longitudinal data on all participants, the estimated treatment effect is -0.19 (se: 0.15, 95% CI: -0.49 to 0.10), providing a 5% reduction in the width of the 95% confidence interval compared to the unadjusted estimator.

We next consider estimators that adjust for more baseline variables than just CDR_{SB}. The study measured a set of potentially prognostic baseline (i.e., measured before randomization) variables including: age, gender, APOE4 carrier status, Alzheimer's Disease Assessment Scale (ADAS) cognitive score, Mini Mental State Exam (MMSE) score, Activities of Daily Living Scale (ADLS) score, and the Global Deterioration Scale (GDS) score. Including this additional set of baseline variables in the MMRM model, the estimated treatment effect was -0.20 (se: 0.14, 95% CI: -0.48, 0.07). The targeted minimum loss based estimator (TMLE) using the same set of baseline variables produced an estimate of -0.20 (se: 0.14, 95% CI: -0.50 to 0.06). The MMRM and TMLE results are very similar to each other when applied to the donepezil MCI trial. The added value of the TMLE estimator, i.e., more robustness to model assumption violations, will be demonstrated in simulation studies in Section 4.

We next compare the impact of adjusting for the larger set of baseline variables (as in the previous paragraph) compared to adjusting only for baseline CDR_{SB} (as done two paragraphs above). Adjusting for the larger set of prognostic variables using either MMRM or TMLE led to an approximately 23% smaller variance compared to the unadjusted estimator. When adjusting only for baseline CDR_{SB}, the variance reduction was approximately 9%. These variance reductions are equivalent to the reductions in the required sample size to achieve a desired power, when using the adjusted versus unadjusted estimators. (Adjusted estimators, such as MMRM and TMLE, are those that adjust for baseline and/or post-randomization variables.)

Table 1: Results of applying 4 estimators of the average treatment effect to the data from the donepezil MCI trial of Petersen et al. (2005). The average treatment effect is defined as the difference in the population mean change in CDR_{SB} between baseline and 18-months comparing assignment to treatment versus control. The results include the estimate, 95% BCa bootstrap confidence interval (CI), estimator variance, and variance reduction (i.e., 1 minus the variance ratio comparing each estimator (numerator) to the unadjusted estimator (denominator)). "Baseline only" indicates that estimator adjusts for only the baseline CDR_{SB} whereas "Baseline +" includes adjustment for the additional prognostic baseline variables described in Section 2.

Estimator	Estimate	95% BCa CI	Variance	Variance Reduction
Unadjusted	-0.19	-0.51 to 0.11	0.024	0
MMRM, baseline only	-0.19	-0.49 to 0.10	0.022	9%
MMRM, baseline +	-0.20	-0.48 to 0.07	0.019	23%
TMLE, baseline +	-0.20	-0.50 to 0.06	0.019	23%

3 Estimator Definitions

3.1 Overview

Consider a randomized trial where participants have measurements recorded at baseline, K interim visits, and then a final visit. The following variables are measured on each participant (if he/she completes the trial): $L_0, A, L_1, L_2, \ldots, L_K, L_{K+1}$, where L_0 is a vector of baseline variables, A is the study arm ($A = 0$ for control and $A = 1$ for treatment), $L_1, \ldots, L_K$ are variables based on measurements at visits $1, \ldots, K$ after enrollment, respectively, and L_{K+1} is the primary outcome measured at the final visit $K + 1$. In the donepezil MCI trial, $K = 2$ and L_1, L_2, and L_3 represent the change in CDR_{SB} from baseline to 6-, 12- and 18-months of follow-up, respectively.

Each participant may be censored due to loss to follow-up (dropout). Let R_t be the indicator variable that the participant attends the next visit, i.e., the visit at time $t+1$. That is, $R_t = 1$ if L_{t+1} is observed, and $R_t = 0$ otherwise. We assume monotone censoring (i.e., right censoring), that is, if $R_t = 0$ then $R_{t+1} = 0$. We make the missing at random assumption, that is, participant dropout is independent of the primary outcome conditioned on the observed data at all visits before dropout. We assume that the baseline variable vector L_0 is independent of the study arm assignment A, which holds by design since study arm assignment is randomized.

We code each participant's data vector as

$$V = (L_0, A, R_0, L_1, R_1, L_2, \ldots, R_{K-1}, L_K, R_K, L_{K+1}). \tag{1}$$

Each variable L_t has value NA if the corresponding $R_{t-1} = 0$ (i.e., if L_t is censored). By convention, define $R_{-1} = 1$; this is important below, in iteration $t = 0$ of step 2 of the TMLE. We use the notation $\bar{L}_k = (L_0, L_1, \ldots, L_k)$ to denote the history up to time k. Each participant's data vector is assumed to be an independent, identically distributed draw from an unknown joint distribution on V.

The goal is to estimate the average treatment effect $\theta = E(L_{K+1}|A = 1) - E(L_{K+1}|A = 0)$.

3.2 Mixed Model for Repeated Measures Estimator

A commonly used estimator of the average treatment effect is based on the mixed model for repeated measures (MMRM) approach. The MMRM approach specifies a longitudinal linear model for the vector of outcomes $L_{1,i}$, ..., $L_{K+1,i}$ for participant $i = 1, ..., N$ as $L_{t,i} = \mu_{t,i} + \epsilon_{t,i}$ for $t = 1, ..., K+1$ where

$$\mu_{t,i} = E(L_{t,i}|A_i, L_{0,i}) = \theta_1 + \sum_{j=2}^{K+1} \theta_j I(t=j) + \alpha_1 \times I(A_i = 1) + \sum_{j=2}^{K+1} \alpha_j I(t=j) \times I(A_i = 1) + \gamma_{\mathbf{B}} L_{0,i}$$

and $\epsilon_{t,i} \sim N(0, \sigma_t^2)$ with $Corr(\epsilon_{t,i}, \epsilon_{t',i}) = \rho_{t,t'}$ for $t \neq t'$. The indicator function, $I(x)$, is defined as 1 if x is true and 0 otherwise. When L_0 is a vector of baseline variables, then γ_B is the corresponding vector of regression coefficients.

The MMRM estimator of the average treatment effect is the maximum likelihood estimator of $\alpha_1 + \alpha_{K+1}$ using the participants data vector (1).

3.3 Targeted Minimum Loss Based Estimator

We define an estimator $\hat{E}(L_{K+1}|A = a)$ of $E(L_{K+1}|A = a)$ for each arm $a \in \{0, 1\}$. The final estimator of the average treatment effect is $\hat{E}(L_{K+1}|A = 1) - \hat{E}(L_{K+1}|A = 0)$. The estimator is a special case of the sequential regression targeted minimum loss based estimator from (van der Laan and Gruber, 2012, Section 3.6, last paragraph).

Step 1 Fit logistic regression models for the probability of attending a given visit conditioned on the observed history prior to that visit (restricting to those who attended the previous visit). For each $t = 0, \ldots, K$, fit a logistic regression model

$$P(R_t = 1|\bar{L}_t, A, R_{t-1} = 1) = \pi_t(\bar{L}_t, A, \boldsymbol{\gamma}^{(t)}),$$

using all participants with $R_{t-1} = 1$, where $\boldsymbol{\gamma}^{(t)}$ is a vector of coefficients (which can differ arbitrarily for each $t = 0, \ldots, K$). This can be done in R using **glm** with the argument **subset** set to only use those with $R_{t-1} = 1$. For example, one could use the logistic regression model $\pi_t(\bar{L}_t, A, \boldsymbol{\gamma}^{(t)}) = \text{logit}^{-1}(\gamma_0^{(t)} + \gamma_1^{(t)} A + \gamma_2^{(t)} L_0 + \cdots + \gamma_{t+2}^{(t)} L_t)$. Another option is to include interaction terms with A as well. Any terms can be included (or not included) that are functions of the vector $\bar{L}_t, A$. Denote the fit coefficients by $\hat{\boldsymbol{\gamma}}^{(t)}$. Also, fit a logistic regression model $P(A = 1|L_0) = q(L_0, \eta)$; for each $a \in \{0, 1\}$, define $q_a(L_0) = q(L_0, \hat{\eta})^a \{1 - q(L_0, \hat{\eta})\}^{1-a}$.

Step 2 involves a sequence of regression model fits, whose purpose at iteration t is to construct a new variable Y_t. We iterate backwards from $t = K+1, K, \ldots, 1, 0$.

Iteration $K+1$ consists of defining $Y_{K+1} = L_{K+1}$ for all participants with $R_K = 1$ (i.e., for all participants who are uncensored at the final visit). That is, add a new column Y_{K+1} to the data set, and set $Y_{K+1} \leftarrow L_{K+1}$ for each participant with $R_K = 1$; set $Y_{K+1} \leftarrow$ NA otherwise.

Loop over $t = K, K-1, \ldots, 1, 0$, where at each iteration t we construct new variables W_t, Y_t by doing the following steps:

1. For each participant with $R_t = 1$, construct the weight:

$$W_t = \left\{ \prod_{t'=0}^{t} \pi_{t'}(\bar{L}_{t'}, A, \hat{\boldsymbol{\gamma}}^{(t')}) \right\}^{-1} \{q_A(L_0)\}^{-1}.$$

This is based on the model fits from step 1. The value $\pi_{t'}(\bar{L}_{t'}, A, \hat{\boldsymbol{\gamma}}^{(t')})$ for a participant with $R_t = 1$ is computed in R by using **glm.predict (type="response")** with a **newdata** row consisting of that participant's values of $\bar{L}_{t'}$. If the aforementioned example of a logistic regression model from step 1 is used, then for a participant with $R_t = 1$, her/his value of $\pi_{t'}(\bar{L}_{t'}, A, \hat{\boldsymbol{\gamma}}^{(t')})$ equals $\text{logit}^{-1}(\hat{\gamma}_0^{(t')} + \hat{\gamma}_1^{(t')} A + \hat{\gamma}_2^{(t')} L_0 + \cdots + \hat{\gamma}_{t'+2}^{(t')} L_{t'})$ substituting the participant's values for each component of $\bar{L}_{t'}, A$.

2. Fit a weighted regression model $E(Y_{t+1}|\bar{L}_t, R_t = 1, A) = m_t(\bar{L}_t, A, \boldsymbol{\beta}^{(t)})$ using weights W_t, among participants with $R_t = 1$; denote the fit coefficients by $\hat{\boldsymbol{\beta}}^{(t)}$. This can be implemented in R using **glm** with the arguments **weight** and **subset** set appropriately. [If the primary outcome L_{K+1} is continuous, linear regression models are used for each m_t; if the primary outcome L_{K+1} is binary, then logistic regression models are used for each m_t.]

3. If $t > 0$, construct the new covariate $Y_t = m_t(\bar{L}_t, A, \hat{\boldsymbol{\beta}}^{(t)})$ for all participants with $R_{t-1} = 1$. In R you generate this by using **glm.predict (type="response")** applied to the model fit, with a **newdata** row consisting of that participant's values of $\bar{L}_t, A$. **This should be done for all participants with $R_{t-1} = 1$ (not just those with $R_t = 1$).** Intuitively, Y_t represents a prediction of the final outcome using only the measurements $\bar{L}_t$ up through time t, for every participant with $R_{t-1} = 1$.

   If $t = 0$, for each $a \in \{0, 1\}$, set $Y_{0,a}$ to be the predicted value $m_0(L_0, a, \hat{\boldsymbol{\beta}}^{(0)})$ **for every participant (not just those with $A = a$).**

4. Decrement t by 1 and iterate the above procedure starting at (1).

When the above loop is completed, each participant has a new variable Y_0 defined.

Step 3 For each $a \in \{0, 1\}$, set $\hat{E}(Y|A = a)$ to be the sample mean of $Y_{a,0}$ across all participants (including those with $A \neq a$).

The TMLE estimator of the average treatment effect is $\hat{\theta} = \hat{E}(Y|A = 1) - \hat{E}(Y|A = 0)$.

4 Simulation Studies Mimicking Features of Donepezil MCI Trial

4.1 Overview

We compared the performance of MMRM, TMLE, and several other adjusted estimators by simulating randomized trials. The simulation distributions were chosen to mimic the joint distribution of the baseline variables and longitudinal outcomes from the donepezil MCI trial. Different dropout distributions were considered in order to explore their impact on estimator bias, variance, mean squared error, and coverage probability of the 95% BCa confidence interval based on 1000 bootstrap samples.

We generated simulated trials with zero average treatment effect. We also generated simulated trials with a beneficial treatment effect $\theta = -0.19$. This value was chosen since it is the unadjusted average treatment effect estimate from the donepezil MCI trial; specifically, at the 18-month follow-up, the mean change in CDR_{SB} was 0.19 units less in the donepezil arm compared to placebo, which is in the direction of a slower decline in cognitive function among patients receiving donepezil.

4.2 Simulation Distributions Part I: Full Data Vectors Without Censoring

Each simulated trial is generated first by resampling 500 participant data vectors (with replacement) from the donepezil MCI trial from those in both arms who completed the trial, i.e. not having any missing CDR_{SB} at the 6-, 12- and 18-month follow-up visits. By resampling patients from the trial, the relationship between the baseline variables and outcomes observed within the donepezil MCI trial is retained. The functional form of this relationship does not follow any simple statistical model (such as the outcome regression models used by the MMRM and TMLE estimators). We consider this an advantage of our simulation study distributions, since in practice these models will be at least somewhat misspecified.

Next, study arm assignment A was overwritten (replaced) for each simulated participant by an independent Bernoulli draw with probability 0.5 to receive donepezil (A = 1) vs. placebo (A = 0). This was to ensure study arm assignment A is drawn independent of the baseline variables L_0. If we had used the original values of A from the data set, this would have induced small correlations with L_0, which violates the randomization assumption and could have distorted our findings.

The third step, described next, only applies when generating simulated trials corresponding to the beneficial average treatment effect. For each simulated participant in the treatment arm $A = 1$, her/his change in CDR_{SB} scores at 6-, 12- and 18-months (L_1, L_2, L_3) were decreased by 0.063, 0.127 and 0.19, respectively.

The fourth step in generating simulated trial data sets was to set patient dropout (the variables R_k). Patient dropout was generated using one of four distributions, called scenarios. These are described below.

For each combination of the four dropout scenarios and two average treatment effects, 10,000 trials were simulated. Each estimator was applied to each simulated data set, and their performance was summarized across data sets under each scenario and treatment effect.

4.3 Simulation Distributions Part II: Data Vectors Incorporating Censoring

Each participant data vector in our simulated trials was first constructed as in Section 4.2, and then censoring (dropout) was added using distributions that are summarized next.

We constructed four dropout scenarios that, similar to the donepezil MCI trial, produce roughly 30% patient dropout over the course of the 18-month follow-up. Precise specifications of the corresponding dropout distributions are given in Appendix A. We summarize key features of these distributions below.

Dropout scenarios A through C induce the following dropout rates over time: 16%, 8% and 7% of patients dropout out just after baseline, 6- and 12-months, respectively, which is similar to what occurred in the donepezil MCI trial. Dropout scenario D, in contrast, has increasing dropout

rates over time; specifically, 7%, 10% and 13% of patients dropout just after baseline, 6-, and 12-months, respectively. These are not cumulative dropout rates; they are the probabilities of dropout among those who have not dropped out previously.

In each of scenarios A and B, dropout is independent of study arm assignment. This type of censoring generally leads to reduced estimator precision compared to no censoring, but does not cause bias in estimating the average treatment effect, since any bias in the estimated mean outcome in each arm cancels out when taking the difference between arms. In scenario A, dropout is completely at random, i.e., independent of all components of the participant data vector. In scenario B, dropout depends on CDR_{SB} and ADAS cognitive score, both measured at baseline, such that patients with higher cognitive impairment at baseline (i.e. higher scores on both scales) are more likely to dropout.

Scenarios C and D were constructed so that dropout is dependent on the treatment arm. The dropout rates are higher in the placebo arm. In scenarios C, dropout depends on some baseline variables and study arm, while in scenario D it also depends on longitudinal data (CDR_{SB} change over time). Specifically, in scenario C, patient dropout is a function of baseline CDR_{SB} and ADAS cognitive score and study arm assignment A. The dropout rates are 22%, 7% and 7% in the placebo arm and 9%, 5% and 6% in the donepezil arm just after baseline, 6-, and 12-months, respectively. In scenario D, dropout probability depends on longitudinal history of CDR_{SB} and study arm A. The resulting proportion of patients that dropout just after each follow-up visit are 6%, 15% and 21% in the placebo arm and 3%, 5% and 9% in the donepezil arm just after baseline, 6-, and 12-month visits, respectively.

4.4 Estimators

The unadjusted estimator of the average treatment effect is the difference in the sample mean change in CDR_{SB} from baseline to 18-months comparing donepezil to placebo, among patients who were followed to the 18-month follow-up. In addition, we considered 5 adjusted estimators for the average treatment effect. Unless otherwise noted, each adjusted estimator accounts for the baseline CDR_{SB} and the following additional baseline covariates described in Section 2: age, gender, APOE4 carrier status, ADAS cognitive score, MMSE score, ADLS score, and the GDS score.

The adjusted estimators include the following for the average treatment effect:

- ANCOVA: The analysis of covariance estimator is the coefficient for the main term of treatment from a linear regression model for the 18-month change in CDR_{SB} with main terms for treatment and for each of the baseline variables, applied only to patients who completed the 18-month follow-up. If patient dropout is generated completely at random, this estimator is consistent, regardless of whether or not the ANCOVA model is correct. This estimator is also consistent for the average treatment effect under patient dropout missing at random if the ANCOVA model is correctly specified.

- PLEASE: The precise locally-efficient augmented simple estimator (Colantuoni and Rosenblum, 2015), is similar to the ANCOVA estimator in that it considers only the 18-month change in CDR_{SB}. This estimator requires specification of three models: a propensity score model for the treatment assignment, a dropout model and an outcome regression model. The

outcome regression model, fit separately to each treatment arm, is a linear model for the 18-month change in CDR_{SB} as a function of main terms for the baseline variables. The propensity score and dropout models are logistic regression models for the probability of $A = 1$ and $R_2 = 1$, respectively, and include main terms for the baseline variables. The dropout model is fit separately for each treatment arm. Due to randomized assignment of treatment, the propensity score model is correct; however, the dropout model is only correct for scenarios A, B and C; the PLEASE estimator is consistent in these scenarios even under misspecification of the outcome regression model.

- MMRM: The MMRM estimator (Section 3.2) involves a longitudinal linear model that includes main terms for treatment and time (factor) plus the interaction for treatment and time and a main term for the baseline CDR_{SB} (and no other baseline variables). This estimator is consistent when dropout is completely at random or when the longitudinal linear model is correctly specified.

- MMRM+: The MMRM+ approach is the same as the MMRM approach but includes main terms for all the baseline variables described above.

- TMLE: The TMLE (Section 3.3) involves fitting three sets of models: a propensity score model for the treatment assignment, dropout models and outcome regression models. The propensity score model is a logistic regression model for $A = 1$ with main terms for the baseline variables. The dropout models are logistic models for $R_0 = 1$, $R_1 = 1$ and $R_2 = 1$ that include main terms for baseline CDR_{SB} and the baseline variables, as well as the prior follow-up CDR_{SB} in the models for R_1 and R_2; fit separately for each treatment arm. The outcome regression models are linear models that include all baseline variables as main terms, as well as the previously measured CDR_{SB} change scores in the models for L_1, L_2 and L_3 (as main terms). The propensity score model is correct due to randomization. The dropout models are correctly specified in all scenarios, which means this estimator is consistent.

4.5 Simulation Results

A comparison of the unadjusted and adjusted estimators of the average treatment effect are presented in Table 2. The performance of the estimators are similar regardless of whether the average treatment effect is zero or beneficial. Table 3 compares the unadjusted and adjusted estimators of the mean 18-month change in CDR_{SB}, separately by study arm, for the setting with treatment effect being beneficial.

Under missing completely at random (scenario A), all of the estimators are approximately unbiased for the average treatment effect and achieve the targeted 95% coverage probability. Among the adjusted estimators, the TMLE yields the greatest precision gain with roughly a 42 percent reduction in the variance of the estimated average treatment effect comparing the unadjusted estimator to the TMLE (Table 2). However, the other adjusted estimators that accounted for the full set of baseline variables (i.e. ANCOVA, PLEASE, and MMRM+) also yield substantial precision gains ranging from 30 to 34 percent when compared to the unadjusted estimator. The MMRM approach that accounts for only the baseline CDR_{SB} yielded the smallest precision gain (roughly 9 percent reduction in the variance compared to the unadjusted estimator).

Analogous patterns of bias, coverage probabilities and precision gains are observed when comparing the performance of estimators of the treatment arm specific mean 18-month change in CDR_{SB} (Table 3) in scenario A. The TMLE yields a 26 percent reduction in variance compared to reductions ranging from 16 to 22 percent for the ANCOVA, PLEASE and MMRM+ and 10 percent for the MMRM compared to the unadjusted estimator.

When patient dropout depends on baseline variables, but not differentially across treatment arms (scenario B), all of the estimators yield approximately unbiased estimates of the average treatment effect and roughly achieve the targeted 95% coverage probability. In this case, the PLEASE and TMLE exhibit a loss of precision relative to the ANCOVA and MMRM+ estimators. However, MMRM and MMRM+ yield biased estimates of the treatment arm specific mean 18-month change in CDR_{SB} (Table 3); whereas, the ANCOVA, PLEASE and TMLE provide approximately unbiased estimates. The MMRM and MMRM+ estimators have roughly the same bias for each treatment arm, and the bias cancels out when estimating the average treatment effect. The bias in estimates of the treatment arm specific mean 18-month change in CDR_{SB} translate to large reductions from the targeted 95% coverage probability; 95% coverage probabilities of roughly 70%, 75% and 87% for the unadjusted, MMRM and MMRM+, respectively. For estimating the treatment arm specific mean change, the ANCOVA estimator has highest relative MSE (roughly 3.07) followed by the TMLE (roughly 2.65) and PLEASE (roughly 2.35) estimators.

Under scenario C, dropout depends on baseline variables and study arm assignment, with risk of dropout greater in the placebo compared to the donepezil arm. There is substantial bias for the unadjusted (roughly 0.19), MMRM (roughly 0.18) and MMRM+ (roughly 0.10) estimators, with corresponding low coverage for the 95% confidence intervals. The ANCOVA estimator bias is considerably smaller; roughly 0.02. From Table 3, the bias in the average treatment effect is attributable to bias in estimating the mean 18-month change in CDR_{SB} for the placebo arm. The PLEASE and TMLE are approximately unbiased for estimating the average treatment effect and treatment arm specific mean changes. Under scenario C, the MSE for the average treatment effect is smallest for the ANCOVA estimator; the performance of the TMLE and MMRM+ estimators are similar and the TMLE estimator yields approximately unbiased estimates.

In scenario D, dropout depends on baseline variables, study arm, and CDR_{SB} measured over time. Under this dropout model, the unadjusted, ANCOVA and PLEASE are biased for estimating the average treatment effect and the treatment arm specific means. Given the bias in the unadjusted, ANCOVA, and PLEASE estimators, the 95% coverage probabilities are reduced and range from 71% to 82% for the average treatment effect and 81% to 89% for the treatment arm specific means. The MMRM and TMLE are approximately unbiased and the MMRM+ has small bias (roughly 0.02) for estimating the average treatment effect. The TMLE has smallest MSE compared to the MMRM estimators.

Table 2: Comparison of the bias, variance, mean squared error and 95% coverage probability for the average treatment effect based on 10,000 hypothetical trials for each treatment effect (zero vs. beneficial) and dropout scenarios A through D.

		Zero treatment effect ($\theta = 0$)					Beneficial treatment effect ($\theta = -0.19$)				
Dropout Scenario	Estimator	Bias	VAR	MSE	RMSE	CP	Bias	VAR	MSE	RMSE	CP
A	Unadj	0.003	0.025	0.025	1.00	0.94	-0.000	0.025	0.025	1.00	0.95
	ANCOVA	0.002	0.019	0.019	1.34	0.94	0.001	0.019	0.019	1.33	0.94
	PLEASE	0.002	0.019	0.019	1.31	0.94	0.001	0.019	0.019	1.30	0.94
	MMRM	0.003	0.023	0.023	1.09	0.94	-0.001	0.023	0.023	1.10	0.95
	MMRM+	0.003	0.019	0.019	1.32	0.94	-0.000	0.019	0.019	1.33	0.95
	TMLE	0.002	0.018	0.018	1.43	0.94	0.000	0.017	0.017	1.42	0.95
B	Unadj	0.002	0.021	0.021	1.00	0.94	0.001	0.020	0.020	1.00	0.95
	ANCOVA	0.002	0.015	0.015	1.35	0.94	0.001	0.015	0.015	1.33	0.95
	PLEASE	0.002	0.025	0.025	0.82	0.93	0.003	0.025	0.025	0.80	0.93
	MMRM	0.002	0.020	0.020	1.06	0.95	0.000	0.019	0.019	1.06	0.95
	MMRM+	0.001	0.016	0.016	1.32	0.94	0.001	0.016	0.016	1.31	0.94
	TMLE	0.001	0.022	0.022	0.93	0.94	0.002	0.022	0.022	0.91	0.93
C	Unadj	0.194	0.020	0.058	1.00	0.72	0.192	0.019	0.056	1.00	0.71
	ANCOVA	0.019	0.015	0.015	3.73	0.94	0.018	0.015	0.015	3.76	0.94
	PLEASE	-0.005	0.034	0.034	1.69	0.92	-0.006	0.033	0.033	1.68	0.92
	MMRM	0.180	0.019	0.051	1.12	0.74	0.178	0.018	0.050	1.12	0.73
	MMRM+	0.107	0.015	0.027	2.15	0.86	0.105	0.015	0.026	2.16	0.85
	TMLE	-0.003	0.029	0.029	1.99	0.92	-0.004	0.028	0.028	2.01	0.92
D	Unadj	0.208	0.024	0.067	1.00	0.72	0.215	0.024	0.070	1.00	0.71
	ANCOVA	0.137	0.018	0.037	1.82	0.81	0.142	0.018	0.038	1.84	0.81
	PLEASE	0.137	0.019	0.037	1.79	0.82	0.143	0.018	0.039	1.81	0.81
	MMRM	0.004	0.022	0.022	3.08	0.94	0.004	0.022	0.022	3.22	0.94
	MMRM+	0.023	0.018	0.018	3.63	0.94	0.024	0.018	0.018	3.85	0.94
	TMLE	-0.000	0.017	0.017	3.91	0.94	0.001	0.017	0.017	4.19	0.95

Table 3: Comparison of the bias, variance, mean squared error and 95% coverage probability for the mean 18-month change in CDR_{SB} for the placebo ($E(Y|A = 0) = 0.567$) and donepezil ($E(Y|A = 1) = 0.377$) arms under beneficial treatment effect ($\theta = -0.19$), based on 10,000 hypothetical trials for dropout scenarios A through D.

		$E(Y\|A=0)=0.567$					$E(Y\|A=1)=0.377$				
Dropout Scenario	Estimator	Bias	VAR	MSE	RMSE	CP	Bias	VAR	MSE	RMSE	CP
A	Unadj	-0.004	0.012	0.012	1.00	0.94	-0.005	0.013	0.013	1.00	0.95
	ANCOVA	-0.005	0.010	0.010	1.19	0.94	-0.004	0.010	0.010	1.21	0.95
	PLEASE	-0.005	0.011	0.011	1.16	0.94	-0.004	0.011	0.011	1.17	0.94
	MMRM	-0.004	0.011	0.011	1.10	0.95	-0.005	0.012	0.012	1.10	0.95
	MMRM+	-0.005	0.010	0.010	1.22	0.94	-0.005	0.010	0.010	1.22	0.94
	TMLE	-0.006	0.010	0.010	1.26	0.94	-0.005	0.010	0.010	1.26	0.94
B	Unadj	-0.151	0.010	0.033	1.00	0.70	-0.150	0.010	0.033	1.00	0.71
	ANCOVA	-0.007	0.011	0.011	3.12	0.94	-0.006	0.011	0.011	3.07	0.94
	PLEASE	0.002	0.014	0.014	2.41	0.93	0.005	0.014	0.014	2.35	0.93
	MMRM	-0.131	0.010	0.027	1.23	0.75	-0.130	0.010	0.027	1.23	0.77
	MMRM+	-0.078	0.009	0.015	2.19	0.87	-0.078	0.009	0.015	2.20	0.87
	TMLE	-0.003	0.012	0.012	2.69	0.93	-0.001	0.012	0.012	2.65	0.94
C	Unadj	-0.259	0.010	0.077	1.00	0.34	-0.067	0.010	0.015	1.00	0.90
	ANCOVA	-0.024	0.012	0.012	6.34	0.93	-0.006	0.010	0.010	1.53	0.94
	PLEASE	0.003	0.026	0.026	2.94	0.91	-0.003	0.010	0.010	1.43	0.94
	MMRM	-0.228	0.009	0.061	1.25	0.42	-0.050	0.010	0.012	1.20	0.92
	MMRM+	-0.140	0.009	0.028	2.71	0.71	-0.035	0.009	0.010	1.47	0.93
	TMLE	-0.002	0.021	0.021	3.62	0.91	-0.006	0.009	0.009	1.56	0.94
D	Unadj	-0.117	0.011	0.025	1.00	0.81	0.097	0.013	0.023	1.00	0.84
	ANCOVA	-0.078	0.010	0.016	1.57	0.87	0.064	0.011	0.015	1.52	0.89
	PLEASE	-0.080	0.010	0.017	1.48	0.87	0.062	0.011	0.015	1.54	0.89
	MMRM	-0.004	0.011	0.011	2.24	0.94	0.000	0.011	0.011	2.07	0.94
	MMRM+	-0.015	0.010	0.010	2.50	0.94	0.009	0.010	0.010	2.25	0.94
	TMLE	-0.001	0.010	0.010	2.44	0.94	-0.000	0.009	0.009	2.44	0.94

5 Discussion

We compared the commonly used MMRM approach and novel TMLE approach for estimating the average treatment effect within a longitudinal randomized controlled trial evaluating the effect of donepezil on cognitive function, as measured by the CDR_{SB} score, among MCI patients. When using the MMRM approach, precision in the estimated average treatment effect can be gained by adjusting for additional prognostic baseline variables beyond the baseline CDR_{SB} score. Researchers can utilize available observational studies or earlier phase clinical trial data and the methods described in (Colantuoni and Rosenblum, 2015) to evaluate the prognostic ability of available baseline variables.

The TMLE approach offers advantages over the MMRM. Specifically, estimates of the aver-

age treatment effect are unbiased, achieve specified coverage probabilities and have lower mean-squared error compared to the MMRM approach under a broad range of distributions generating patient dropout; whereas, the estimated average treatment effects based on the MMRM can suffer from large bias when patient dropout is generated under different distributions within each treatment arm. Similar advantages of the TMLE hold when estimating the treatment arm specific mean in the primary outcome.

6 Acknowledgments

Data collection and sharing for this project was funded by the Alzheimer's Disease Cooperative Study (ADCS) (National Institutes of Health Grant U19 5U19AG010483). This research was funded under NIH/NIA grant R01AG048349. This publication's contents are solely the responsibility of the authors and do not represent the views of the above organizations.

References

Chen, Y.-F., X. Ni, A. S. Fleisher, W. Zhou, P. Aisen, and R. Mohs (2018). A simulation study comparing slope model with mixed-model repeated measure to assess cognitive data in clinical trials of alzheimer's disease. *Alzheimer's Dementia: Translational Research Clinical Interventions 4*, 46 – 53.

Colantuoni, E. and M. Rosenblum (2015). Leveraging prognostic baseline variables to gain precision in randomized trials. *Statistics in Medicine 34*(18), 2602–2617. sim.6507.

Petersen, R. C., R. G. Thomas, M. Grundman, D. Bennett, R. Doody, S. Ferris, D. Galasko, S. Jin, J. Kaye, A. Levey, E. Pfeiffer, M. Sano, C. H. van Dyck, and L. J. Thal (2005). Vitamin e and donepezil for the treatment of mild cognitive impairment. *New England Journal of Medicine 352*(23), 2379–2388. PMID: 15829527.

van der Laan, M. J. and S. Gruber (2012). Targeted minimum loss based estimation of causal effects of multiple time point interventions. *The International Journal of Biostatistics 8*(1), Article 9.

A Simulation: Dropout models

In the donepezil MCI trial, roughly 30% of the patients dropped out of the study by 18-months (16%, 8% and 7% dropped out after baseline, 6- and 12-months, respectively). We considered three dropout scenarios that produced roughly the same marginal dropout rates in our hypothetical trials. In scenarios A and B, patient dropout was independent of treatment assignment but varied over follow-up. In scenario A, patient dropout was completely at random. In scenario B, patient dropout depended on both the baseline CDR_{SB} (denoted Y_0) and ADAS cognitive score (patients with higher scores or more cognitive impairment were more likely to dropout).

In scenarios C and D, the patterns of patient dropout varied by treatment arm with greater proportion of patients dropping out of the placebo arm. Patient dropout in scenario C depends on baseline CDR_{SB} and the ADAS cognitive score, separately for each treatment arm. In Scenario D, a total of roughly 30% of the patients dropout (similar to the donepizil trial); however the proportion of dropout after baseline, 6- and 12-months is 6%, 15% and 21% in the placebo arm and 3%, 5% and 9% in the donepezil arm.

The specific models for R_0, R_1 and R_2 for each scenario are provided below. Recall that Y_0 is CDR_{SB} measured at baseline and Y_k, for $k > 0$ is defined as the $CDR_{SB,k} - CDR_{SB,0}$ (i.e. change in CDR_{SB} comparing current follow-up to baseline).

1. Scenario A:

$$P(R_0 = 1) = \text{logit}^{-1}(\text{logit}(0.84))$$
$$P(R_1 = 1|R_0 = 1) = \text{logit}^{-1}(\text{logit}(0.90))$$
$$P(R_2 = 1|R_1 = 1) = \text{logit}^{-1}(\text{logit}(0.90))$$

2. Scenario B:

$$P(R_0 = 1|L_0) = \text{logit}^{-1}(\text{logit}(0.91) - 0.25 \times (ADAS - \bar{X}_{ADAS}) - 0.10 \times Y_0)$$
$$P(R_1 = 1|R_0 = 1, L_0) = \text{logit}^{-1}(\text{logit}(0.93) - 0.15 \times (ADAS - \bar{X}_{ADAS}) - 0.10 \times Y_0)$$
$$P(R_2 = 1|R_1 = 1, L_0) = \text{logit}^{-1}(\text{logit}(0.93) - 0.10 \times (ADAS - \bar{X}_{ADAS}) - 0.10 \times Y_0)$$

3. Scenario C:

$$P(R_0 = 1|L_0, A) = \text{logit}^{-1}(\text{logit}(0.93) + (-0.35(ADAS - \bar{X}_{ADAS}) - 0.40Y_0) \times (1 - A)$$
$$+(-0.10(ADAS - \bar{X}_{ADAS}) - 0.10Y_0) \times A)$$
$$P(R_1 = 1|R_0 = 1, L_0, A) = \text{logit}^{-1}(\text{logit}(0.95)+(-0.25(ADAS-\bar{ADAS})-0.35Y_0)\times(1-A)$$
$$+(-0.10(ADAS - \bar{ADAS}) - 0.10Y_0) \times A)$$
$$P(R_2 = 1|R_1 = 1, L_0, A) = \text{logit}^{-1}(\text{logit}(0.94)+(-0.20(ADAS-\bar{ADAS})-0.30Y_0)\times(1-A)$$
$$+(-0.10(ADAS - \bar{ADAS}) - 0.10Y_0) \times A)$$

4. Scenario D:

$$P(R_0 = 1|L_0, A) = \text{logit}^{-1}(\text{logit}(0.93) + 0.20Y_0 \times (2A - 1))$$
$$P(R_1 = 1|R_0 = 1, L_0 1A) = \text{logit}^{-1}((\text{logit}(0.89) + 0.30\epsilon_1 \times (2A - 1))$$
$$P(R_2 = 1|R_1 = 1, L_2, A) = \text{logit}^{-1}((\text{logit}(0.85) + 0.40\epsilon_2 \times (2A - 1))$$

where ϵ_1 are the residuals from the linear regression fit of Y_1 on Y_0, i.e. the information in the 6-month change in CDR_{SB} that is not explained by the baseline CDR_{SB}). Similarly, ϵ_2 are the residuals from the linear regression fit of Y_2 on Y_0, i.e. the information in the 12-month change in CDR_{SB} that is not explained by the baseline CDR_{SB}.