Johns Hopkins University, Dept. of Biostatistics Working Papers

Fall 10-28-2019

Model-Robust Inference for Clinical Trials that Improve Precision by Stratified Randomization and Adjustment for Additional Baseline Variables

Bingkai Wang
Department of Biostatistics, Johns Hopkins Bloomberg chool of Public Health, bingkai.w@gmail.com

Ryoko Susukida
Department of Mental Health, Johns Hopkins Bloomberg School of Public Health, rsusukl1@jhu.edu

Ramin Mojtabai
Department of Mental Health, Johns Hopkins Bloomberg School of Public Health, rmojtab1@jhu.edu

Masoumeh Amin-Esmaeili
Department of Mental Health, Johns Hopkins Bloomberg School of Public Health and Iranian National Center for Addiction Studies (INCAS), Tehran University of Medical Sciences, mamines1@jhmi.edu

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

Follow this and additional works at: https://biostats.bepress.com/jhubiostat

Suggested Citation

Wang, Bingkai; Susukida, Ryoko; Mojtabai, Ramin; Amin-Esmaeili, Masoumeh; and Rosenblum, Michael, "Model-Robust Inference for Clinical Trials that Improve Precision by Stratified Randomization and Adjustment for Additional Baseline Variables" (October 2019). *Johns Hopkins University, Dept. of Biostatistics Working Papers.* Working Paper 293.
https://biostats.bepress.com/jhubiostat/paper293

This working paper is hosted by The Berkeley Electronic Press (bepress) and may not be commercially reproduced without the permission of the copyright holder.
Copyright © 2011 by the authors

Model-Robust Inference for Clinical Trials that Improve Precision by Stratified Randomization and Adjustment for Additional Baseline Variables

Bingkai Wang

Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health

Ryoko Susukida

Department of Mental Health, Johns Hopkins Bloomberg School of Public Health

Ramin Mojtabai

Department of Mental Health, Johns Hopkins Bloomberg School of Public Health

Masoumeh Amin-Esmaeili

Department of Mental Health, Johns Hopkins Bloomberg School of Public Health and

Iranian National Center for Addiction Studies (INCAS), Tehran University of Medical Sciences

and

Michael Rosenblum

Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health

Summary

We focus on estimating the average treatment effect in clinical trials that involve stratified randomization, which is commonly used. It is important to understand the large sample properties of estimators that adjust for stratum variables (those used in the randomization procedure) and additional baseline variables, since this can lead to substantial gains in precision and power.

COBRA
Collection of Biostatistics
Research Archive

Hosted by The Berkeley Electronic Press

Surprisingly, to the best of our knowledge, this is an open problem. It was only recently that a simpler problem was solved by Bugni et al. (2018) for the case with no additional baseline variables, continuous outcomes, the analysis of covariance (ANCOVA) estimator, and no missing data. We generalize their results in three directions. First, in addition to continuous outcomes, we handle binary and time-to-event outcomes; this broadens the applicability of the results. Second, we allow adjustment for an additional, preplanned set of baseline variables, which can improve precision. Third, we handle missing outcomes under the missing at random assumption. We prove that a wide class of estimators is asymptotically normally distributed under stratified randomization and has equal or smaller asymptotic variance than under simple randomization. For each estimator in this class, we give a consistent variance estimator. This is important in order to fully capitalize on the combined precision gains from stratified randomization and adjustment for additional baseline variables. The above results also hold for the biased-coin covariate-adaptive design. We demonstrate our results using completed trial data sets of treatments for substance use disorder, where adjustment for additional baseline variables brings substantial variance reduction.

Key words: Covariate-adaptive randomization, generalized linear model, robustness.

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

1. Introduction

Covariate-adaptive designs refer to randomization schemes that assign participants to study arms in a way that improves balance across arms in preselected strata of the baseline variables. E.g., balance on disease severity, a genetic marker, or another variable thought to be correlated with the primary outcome could be sought. The simplest form of covariate-adaptive randomization is stratified permuted block randomization (referred to as "stratified randomization" throughout, for conciseness).

Compared with simple randomization, covariate-adaptive randomization can be advantageous in minimizing imbalance and improving efficiency (Efron, 1971; Pocock and Simon, 1975; Wei, 1978). Due to these benefits, covariate-adaptive randomization has become a popular approach in clinical trials. According to a survey by Lin et al. (2015), 183 out of their sample of 224 randomized clinical trials published in 2014 in leading medical journals used some form of covariate-adaptive randomization. Stratified randomization (Zelen, 1974) was implemented by 70% of trials in this survey. Another method for covariate-adaptive randomization is the biased-coin design by Efron (1971). Other examples include Wei's urn design (Wei, 1978) and rerandomization (Morgan and Rubin, 2012). Our results only apply to stratified randomization and the biased-coin design.

Concerns have been raised regarding how to perform valid statistical analyses at the end of trials that use covariate-adaptive randomization. Adjusting for stratification variables is recommended (Lachin et al., 1988; Kahan and Morris, 2012; EMA, 2015), but, according to a survey by Kahan and Morris (2012) that sampled 65 published trials from major medical journals from March to May 2010, 41 implemented covariate-adaptive randomization (among which 29 used stratified randomization), but only 14 adjusted in the primary analysis for the variables used in the randomization procedure. Furthermore, there are few results on how to conduct the primary efficacy analysis in trials that use stratified randomization without making parametric model assumptions (which, if incorrect, can lead to bias).

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

Hosted by The Berkeley Electronic Press

Shao et al. (2010) proved the validity of the two-sample t-test under the biased-coin design assuming the outcome follows a linear model. Shao and Yu (2013) extended this result to a case where the outcome follows a generalized linear model. Ma et al. (2015, 2018) also assumed a linear model and derived the asymptotic distribution of the test statistic of the average treatment effect for the ANCOVA estimator and a class of covariate-adaptive designs. Bugni et al. (2018), who use the superpopulation inference framework as done here, established the asymptotic theory of the unadjusted estimator and the ANCOVA estimator (with adjustment for strata only) of the average treatment effect for a wide range of covariate-adaptive designs; their results, like ours in Sections 5.1-5.2, are robust to arbitrary misspecification of the regression model used in the estimator. Ye and Shao (2019) derived asymptotics for log-rank and score tests in survival analysis under covariate-adaptive randomization, and their methods can handle adjustment for additional baseline variables; however, estimation was not addressed. Li and Ding (2019) established the asymptotic theory for the ANCOVA estimator under covariate-adaptive randomization in the randomization inference framework.

For trials using stratified randomization or biased-coin covariate-adaptive randomization, to the best of our knowledge, it was an open problem to determine (without making parametric regression model assumptions) the large sample properties of covariate-adjusted estimators that involve any of the following: binary outcomes or time-to-event outcomes, adjustment for baseline variables other than those in the randomization procedure, or missing data. This is the problem that we address. The main challenge is that treatment assignment is not independent across participants.

Under regularity conditions, we prove that a large class of estimators is asymptotically normally distributed in randomized trials that use stratified randomization or the biased-coin design, and we give a formula for computing their asymptotic variance. This class of estimators includes ANCOVA (referred to as 'ANCOVA I' by Yang and Tsiatis, 2001) for continuous outcomes, the

COBRA
Collection of Biostatistics
Research Archive

standardized logistic regression estimator (Scharfstein et al., 1999; Moore and van der Laan, 2009) for binary outcomes, the doubly-robust weighted least squares (DR-WLS) estimator from Marshall Joffe as described by (Robins et al., 2007) for continuous and binary outcomes, and estimators of the restricted mean survival time for time-to-event outcomes (Van der Laan et al., 2003; Díaz et al., 2019). The key theoretical underpinning of our results is the empirical process result from Shorack and Wellner (2009), which was insightfully used by Bugni et al. (2018) to prove their fundamental results for the special case described above.

As in the above referenced work, we assume that the randomization scheme and analysis method have been completely specified before the trial starts, as is typically required by regulators such as the European Medicines Agency and the U.S. Food and Drug Administration (FDA and EMA, 1998; EMA, 2015; FDA, 2019).

In the next section, we describe three motivating trial examples to which we apply our methods. In Section 3, we describe our setup, notation and assumptions. We present our main results in Section 4. In Section 5, we give example estimators to which our general results apply. Trial applications are provided in Section 6. Practical recommendations and future directions are discussed in Section 7.

2. Three completed randomized trials that used stratified randomization

2.1 *Study of buprenorphine tapering schedule and illicit opioid use (NIDA-CTN-0003)*

The trial of "Buprenorphine tapering schedule and illicit opioid use", referred to as "NIDA-CTN-0003" in the National Drug Abuse Treatment Clinical Trials Network, is a phase-3 randomized trial completed in 2005 (Ling et al., 2009). The goal was to compare the effects of a short or long taper schedule after buprenorphine stabilization of patients with opioid use disorder. Patients were randomized into two arms: 28-day taper (control, 259 patients, 36% missing outcomes) and 7-day taper (treatment, 252 patients, 21% missing outcomes), stratified by maintenance dose (3

Hosted by The Berkeley Electronic Press

levels) measured at randomization. The outcome of interest is a binary indicator of whether a participant's urine tested at the end of study is opioid-free. Missing outcomes are encoded as "NA" and not imputed. In addition to the stratification variable, we adjust for the following baseline variables: sex, opioid urine toxicology results, the Adjective Rating Scale for Withdrawal (ARSW), the Clinical Opiate Withdrawal Scale (COWS) and the Visual Analog Scale (VAS).

2.2 *Prescription Opioid Addiction Treatment Study (NIDA-CTN-0030)*

The "Prescription Opioid Addiction Treatment Study", referred to as "NIDA-CTN-0030" is a phase-3 randomized trial completed in 2013 (Weiss et al., 2011). The goal is to determine whether adding individual drug counseling to the prescription of buprenorphine/naloxone will improve treatment outcome for patients with prescription opioid use disorder. Though this study adopted a 2-phase adaptive design, we focus on phase I, where patients were randomized into standard medical management (control, 330 patients, 10% missing outcomes) or standard medical management plus drug counseling (treatment, 335 patients, 13% missing outcomes). Randomization is stratified by the presence or absence of a history of heroin use and current chronic pain, resulting in 4 strata. The outcome of interest is the proportion of negative urine laboratory results among all tests. Among all 5 urine laboratory tests during the first 4 weeks of phase I, if a patient missed visits of more than 2 weeks, the outcome is regarded as missing. Baseline variables that are included in the analysis are strata, age, sex and urine laboratory result.

2.3 *Study of internet-delivered treatment for substance abuse (NIDA-CTN-0044)*

The phase-3 randomized trial "Internet-delivered treatment for substance abuse", referred to as "NIDA-CTN-0044", was completed in 2012 (Campbell et al., 2014). The goal was to evaluate the effectiveness of a web-delivered behavioral intervention, Therapeutic Education System (TES), in the treatment of substance abuse. Participants were randomly assigned to two arms: treatment

https://biostats.bepress.com/jhubiostat/paper293

as usual (control, 252 participants, 19% missing outcomes) and treatment as usual plus TES (treatment, 255 participants, 18% missing outcomes). Randomization was stratified by treatment site, patient's primary substance of abuse (sitmulant or nonstimulant) and abstinence status at baseline. Since we do not have access to the treatment site variable, only patient's primary substance of abuse and abstinence status at baseline were used to construct strata (4 levels) and treatment site was omitted from our analysis. After randomization, each participant was followed by 12 weeks with 2 urine laboratory tests per week. The outcome of interest is the proportion of negative urine lab results among all tests. If a participant missed visits of more than 6 weeks, the outcome is regarded as missing. We adjust for strata and additional baseline variables: age, sex and urine laboratory result.

In some cases, the outcomes in our analyses differ from the primary outcomes in the corresponding trials. The reason is that we wanted to use the same outcomes across trials for illustration. Our outcomes are all considered clinically meaningful in the field of substance use disorder treatments. All trial data were drawn from the National Institute on Drug Abuse Clinical Trials Network Data Share Website (CITE).

3. Definitions and assumptions

We focus on randomized trials that use stratified randomization (or the biased-coin design) with n participants. For each participant $i = 1, \ldots, n$, let Y_i denote the observed outcome, M_i denote whether Y_i is observed ($M_i = 1$) or not ($M_i = 0$), A_i denote the treatment allocation ($A_i = 1$ if assigned to treatment and $A_i = 0$ if assigned to control), and $\boldsymbol{X}_i$ denote a vector of baseline covariates.

We adopt the Neyman-Rubin causal model, which assumes $Y_i = Y_i(1)A_i + Y_i(0)(1 - A_i)$, where $Y_i(a)$ is the potential outcome for participant i were they in treatment group $a, a = 0, 1$. Analogous to the definition of potential outcomes, we define "potential missingness" $M_i(a)$, which

Hosted by The Berkeley Electronic Press

is the indicator of whether $Y_i(a)$ would be observed for participant i were they assigned to study arm $a \in \{0, 1\}$; we assume that the observed (non)-missingness variable M_i is connected to the potential missingness variables as follows: $M_i = M_i(1)A_i + M_i(0)(1 - A_i)$. Though potential missingness is not commonly used in the literature, it fits into the Neyman-Rubin causal model if missingness (e.g., whether a participant completes follow-up) is also an outcome of interest, and the outcome vector is defined as $(M_i, Y_i M_i)$. We emphasize that our asymptotic results involve neither the joint distribution of $(M_i(1), M_i(0))$ nor $(Y_i(1), Y_i(0))$, and that the targets of estimation are identifiable using only the observed data distribution. Denote participant i's vector of potential outcomes as $\boldsymbol{W}_i = (Y_i(1), Y_i(0), M_i(1), M_i(0), \boldsymbol{X}_i)$ (some of which are not observed) and their observed vector as $(A_i, \boldsymbol{X}_i, Y_i M_i, M_i)$. We make the following assumptions on the distribution of potential outcome vectors $\boldsymbol{W}_i$:

Assumption 1.

(1) $\boldsymbol{W}_i, i = 1, \ldots, n$ are independent samples from an unknown joint distribution P on $\boldsymbol{W} = (Y(1), Y(0), M(1), M(0), \boldsymbol{X})$.

(2) Missing (censoring) at random (MAR): $M(a) \perp\!\!\!\perp Y(a) | \boldsymbol{X}$ for each arm $a \in \{0, 1\}$, where $\perp\!\!\!\perp$ denotes independence.

For stratified randomization or the biased-coin design, treatment allocation depends on strata, such as gender, location or race. Strata are represented by a single, categorical, stratification variable S taking K possible values, which is encoded using $K - 1$ dummy variables in the covariate vector $\boldsymbol{X}$. For example, if strata are determined by 4 sites and a binary indicator of high disease severity, then S has $K = 8$ possible values. We denote S_i as the stratification variable for participant i and $\mathcal{S}$ the set of all K strata. The goal of stratified randomization or the biased-coin design is to achieve "balance" in each stratum; that is, the proportion of participants assigned to the treatment arm is targeted to a prespecified proportion $\pi \in (0, 1)$, e.g. $\pi = 0.5$.

Stratified randomization uses permuted blocks to assign treatment. In each stratum, a ran-

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

domly permuted block with fraction π 1's (representing treatment) and $(1-\pi)$ 0's (representing control) is used for sequential allocation. When a block is exhausted, a new block is used.

The biased-coin design allocates participants sequentially by the following rule for $k = 1, \ldots, n$:

$$P(A_k = 1 | S_1, \ldots, S_k, A_1, \ldots, A_{k-1}) = \begin{cases} \pi, & \text{if } \sum_{i=1}^{k-1}(A_i - \pi) I\{S_i = S_k\} = 0 \\ \lambda, & \text{if } \sum_{i=1}^{k-1}(A_i - \pi) I\{S_i = S_k\} < 0 \\ 1-\lambda, & \text{if } \sum_{i=1}^{k-1}(A_i - \pi) I\{S_i = S_k\} > 0 \end{cases}$$

with $\lambda \in (\pi, 1]$, e.g., $\lambda = 0.8$.

For these two designs, it follows by construction (as proved by Examples 3.2 and 3.4 of Bugni et al., 2018) that $(\boldsymbol{W}_1, \ldots, \boldsymbol{W}_n) \perp\!\!\!\perp (A_1, \ldots, A_n) | (S_1, \ldots, S_n)$, since treatment allocation only depends on strata and exogenous randomness. This and the assumptions above imply for each $s \in \mathcal{S}$ that

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} (A_i - \pi) I\{S_i = s\} | (S_1, \ldots, S_n) \xrightarrow{d} 0 \text{ a.s.}$$

Different from these two designs, simple randomization assigns treatments by independently flipping a coin for each participant with $P(A_i = 1) = \pi$. This randomization scheme implies that $A_i \perp\!\!\!\perp A_j$ for $i \neq j$ and $A_i \perp\!\!\!\perp \boldsymbol{W}_i$ for all $i \leqslant n$.

Given the observed data $(A_i, \boldsymbol{X}_i, Y_i M_i, M_i)$ for each participant i, our goal is to estimate a population parameter Δ^*, which is a contrast between the marginal distributions of $Y(1)$ and $Y(0)$. For example, the parameter of interest can be the population average treatment effect, defined as $\Delta^* = E[Y(1)] - E[Y(0)]$, or it can be the restricted mean survival time when the outcome is a time-to-event.

We assume that the estimator of Δ^* can be expressed as a solution to estimating equations of the form:

$$\sum_{i=1}^{n} \boldsymbol{\psi}(A_i, \boldsymbol{X}_i, Y_i M_i, M_i; \boldsymbol{\theta}) = 0, \tag{1}$$

where $\boldsymbol{\psi}$ is a column vector of known functions with dimension $p+1$, $\boldsymbol{\theta} = (\Delta, \boldsymbol{\beta}^t)^t$ is a column vector of parameters with dimension $p+1$ and $\boldsymbol{\beta}$ is a column vector of nuisance parameters with

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

dimension p. We denote the solution to equation (1) as $\hat{\boldsymbol{\theta}} = (\widehat{\Delta}, \hat{\boldsymbol{\beta}}^t)^t$. Then $\widehat{\Delta}$ is regarded as the estimator of Δ^*. Many estimators used in clinical trials, including all estimators defined in Section 5, can be expressed as solutions to estimating equations (1) for an appropriately chosen estimating function $\boldsymbol{\psi}$.

We assume regularity conditions similar to the classical conditions that are used for proving consistency and asymptotic linearity of M-estimators for independent, identically distributed (i.i.d.) data, as given in Section 5.3 of van der Vaart (1998). One of the conditions is that the expectation of the estimating equations $E^*[\boldsymbol{\psi}(A, \boldsymbol{X}, YM, M; \boldsymbol{\theta})] = 0$ has a unique solution in $\boldsymbol{\theta}$; in our framework, the expectation E^* on the observed data vector $\boldsymbol{\psi}(A, \boldsymbol{X}, YM, M; \boldsymbol{\theta})$ is defined as being with respect to the distribution induced by first drawing $\boldsymbol{W}$ from the joint distribution P on potential outcomes (see Assumption 1), then drawing A as an independent Bernoulli draw with probability π of being 1, and lastly applying the consistency assumptions $Y = Y(1)A + Y(0)(1-A)$ and $M = M(1)A + M(0)(1-A)$. Equivalently, this condition is that

$$\pi E[\boldsymbol{\psi}(1, \boldsymbol{X}, Y(1)M(1), M(1); \boldsymbol{\theta})] + (1-\pi)E[\boldsymbol{\psi}(0, \boldsymbol{X}, Y(0)M(0), M(0); \boldsymbol{\theta})] = \mathbf{0} \tag{2}$$

has a unique solution in $\boldsymbol{\theta}$, which is denoted as $\underline{\boldsymbol{\theta}} = (\underline{\Delta}, \underline{\boldsymbol{\beta}}^t)^t$, where E denotes expectation with respect to the joint distribution P on the potential outcomes. The full set of regularity conditions is given in Appendix A.

Results in Section 5.3 of van der Vaart (1998) imply that under simple randomization, given Assumption 1 and the regularity conditions in Appendix A, $\widehat{\Delta}$ converges in probability to $\underline{\Delta}$ and is asymptotically normally distributed with asymptotic variance that we denote by $\tilde{V}$. We focus on determining what happens under stratified randomization or the biased-coin design, where our main result (Section 4) is that consistency and asymptotic normality still hold but the asymptotic variance may be smaller (and a consistent variance estimator is given).

https://biostats.bepress.com/jhubiostat/paper293

4. Main result

Theorem 4.1 Let $\widehat{\Delta}$ be the estimator of Δ^* corresponding to solving (1) for a given set of estimating functions $\boldsymbol{\psi}$ with parameter $\boldsymbol{\theta} = (\Delta, \boldsymbol{\beta}^t)^t$, and $\underline{\boldsymbol{\theta}} = (\underline{\Delta}, \underline{\boldsymbol{\beta}}^t)^t$ be the solution to the corresponding expectation of estimating equations (2). Then given Assumption 1 and regularity conditions in Appendix A, under stratified randomization or the biased-coin design,

$$\sqrt{n}(\widehat{\Delta} - \underline{\Delta}) \xrightarrow{d} N(0, V),$$

where V is the first-row, first-column entry of the matrix $\boldsymbol{B}^{-1}(\boldsymbol{C} + \boldsymbol{D})(\boldsymbol{B}^{-1})^t$, where

$$\boldsymbol{B} = \pi E\left[\frac{\partial}{\partial \boldsymbol{\theta}} \boldsymbol{\psi}(1, \boldsymbol{X}, Y(1)M(1), M(1); \boldsymbol{\theta})\Big|_{\boldsymbol{\theta}=\underline{\boldsymbol{\theta}}}\right] + (1-\pi)E\left[\frac{\partial}{\partial \boldsymbol{\theta}} \boldsymbol{\psi}(0, \boldsymbol{X}, Y(0)M(0), M(0); \boldsymbol{\theta})\Big|_{\boldsymbol{\theta}=\underline{\boldsymbol{\theta}}}\right],$$

$$\boldsymbol{C} = \pi E[Var\{\boldsymbol{\psi}(1, \boldsymbol{X}, Y(1)M(1), M(1); \underline{\boldsymbol{\theta}})|S\}] + (1-\pi)E[Var\{\boldsymbol{\psi}(0, \boldsymbol{X}, Y(0)M(0), M(0); \underline{\boldsymbol{\theta}})|S\}],$$

$$\boldsymbol{D} = E[E[\pi \boldsymbol{\psi}(1, \boldsymbol{X}, Y(1)M(1), M(1); \underline{\boldsymbol{\theta}}) - (1-\pi)\boldsymbol{\psi}(0, \boldsymbol{X}, Y(0)M(0), M(0); \underline{\boldsymbol{\theta}})|S]^2].$$

Furthermore, $V \leqslant \tilde{V}$, where $\tilde{V}$ is the asymptotic variance under simple randomization. V can be consistently estimated using the observed data distribution as described in Appendix B.

Theorem 4.1 implies that, given our setup and assumptions, whenever an estimator $\widehat{\Delta}$ is consistent and asymptotically normally distributed under simple randomization, then it is consistent and asymptotically normally distributed under stratified randomization or the biased-coin design with equal or smaller asymptotic variance.

For the unadjusted estimator, our Theorem 4.1 is equivalent to Theorem 4.1 of Bugni et al. (2018) under stratified randomization or the biased-coin design. In the special case of continuous outcomes, if the ANCOVA estimator is used with $X = S$, then Theorem 4.1 is equivalent to the result of Bugni et al. (2018) in section 4.2 under stratified randomization or the biased-coin design, though their results also handle other types of covariate-adaptive randomization. Our proof relies on a generalization of a key Lemma (Lemma C.2 in Appendix C) from Bugni et al. (2018) that is based on the empirical process result of Shorack and Wellner (2009).

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

Hosted by The Berkeley Electronic Press

5. Examples of estimators

We give several examples of estimators that our theorem above applies to. For estimators defined in Sections 5.1-5.3, the parameter of interest, i.e. Δ^*, is the average treatment effect defined as $E[Y(1)] - E[Y(0)]$. In the first two subsections, we assume no missing data.

5.1 *The ANCOVA estimator for continuous outcomes; no missing data*

The ANCOVA estimator $\widehat{\Delta}_{ancova}$ for Δ^* involves first fitting a linear regression working model

$$E[Y|A, \boldsymbol{X}] = \beta_0 + \beta_A A + \boldsymbol{\beta}_{\boldsymbol{X}}^t \boldsymbol{X} \tag{3}$$

and then letting $\widehat{\Delta}_{ancova}$ be the ordinary least square estimate of β_A in model (3). The working model (3) can be arbitrarily misspecified. The ANCOVA estimator can be equivalently calculated by solving estimating equations (1) letting

$$\boldsymbol{\psi}(A_i, \boldsymbol{X}_i, Y_i, M_i; \boldsymbol{\theta}) = \begin{pmatrix} \beta_A - \Delta \\ \{Y_i - (\beta_0 + \beta_A A_i + \boldsymbol{\beta}_{\boldsymbol{X}}^t \boldsymbol{X}_i)\}\boldsymbol{Z}_i \end{pmatrix},$$

where $\boldsymbol{Z}_i = (1, A_i, \boldsymbol{X}_i^t)^t$.

5.2 *The standardized logistic regression estimator for binary outcomes; no missing data*

The standardized logistic regression estimator $\widehat{\Delta}_{logistic}$ is calculated by first fitting a working model:

$$P(Y = 1|A, \boldsymbol{X}) = expit(\beta_0 + \beta_A A + \boldsymbol{\beta}_{\boldsymbol{X}}^t \boldsymbol{X}), \tag{4}$$

where $expit(x) = 1/(1+e^{-x})$, and getting the maximum likelihood estimates (MLE) $(\hat{\beta}_0, \hat{\beta}_A, \hat{\boldsymbol{\beta}}_{\boldsymbol{X}}^t)^t$. Then we have

$$\widehat{\Delta}_{logistic} = \frac{1}{n}\sum_{i=1}^{n}\{expit(\hat{\beta}_0 + \hat{\beta}_A + \hat{\boldsymbol{\beta}}_{\boldsymbol{X}}^t \boldsymbol{X}_i) - expit(\hat{\beta}_0 + \hat{\boldsymbol{\beta}}_{\boldsymbol{X}}^t \boldsymbol{X}_i)\}.$$

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

We do not assume the logistic regression model (4) to be correctly specified. The estimator $\widehat{\Delta}_{logistic}$ can be equivalently calculated by solving estimating equations (1) letting

$$\boldsymbol{\psi}(A_i, \boldsymbol{X}_i, Y_i, M_i; \boldsymbol{\theta}) = \begin{pmatrix} expit(\beta_0 + \beta_A + \boldsymbol{\beta}_{\boldsymbol{X}}^t \boldsymbol{X}_i) - expit(\beta_0 + \boldsymbol{\beta}_{\boldsymbol{X}}^t \boldsymbol{X}_i) - \Delta \\ \{Y_i - expit(\beta_0 + \beta_A A_i + \boldsymbol{\beta}_{\boldsymbol{X}}^t \boldsymbol{X}_i)\} \boldsymbol{Z}_i \end{pmatrix}.$$

Another estimator is the logistic coefficient estimator, defined as $\hat{\beta}_A$. Unlike the standardized logistic regression estimator, the logistic coefficient estimator estimates a conditional effect and can lead to invalid inference if there is treatment effect heterogeneity. A detailed comparison of the two estimators can be found in Steingrimsson et al. (2016). We hence do not consider the logistic coefficient estimator in this paper.

5.3 *The DR-WLS estimator for continuous and binary outcomes; outcomes missing at random*

When missing outcome data present, we estimate Δ^* by the DR-WLS estimator. The estimator is calculated by the following steps. (1) Fit a logistic regression model:

$$P(M = 1|A, \boldsymbol{X}) = expit(\alpha_0 + \alpha_A A + \boldsymbol{\alpha}_{\boldsymbol{X}}^t \boldsymbol{X}) \tag{5}$$

and get MLE $(\hat{\alpha}_0, \hat{\alpha}_A, \hat{\boldsymbol{\alpha}}_{\boldsymbol{X}}^t)^t$ of parameters $(\alpha_0, \alpha_A, \boldsymbol{\alpha}_{\boldsymbol{X}}^t)^t$. (2) Fit a generalized linear model

$$E[Y|A, \boldsymbol{X}] = g(\beta_0 + \beta_A A + \boldsymbol{\beta}_{\boldsymbol{X}}^t \boldsymbol{X}) \tag{6}$$

with weights $1/expit(\hat{\alpha}_0 + \hat{\alpha}_A A_i + \hat{\boldsymbol{\alpha}}_{\boldsymbol{X}}^t \boldsymbol{X}_i)$ using data with $M_i = 1$. Here the inverse link function is $g(x) = x$ for continuous outcomes and $g(x) = expit(x)$ for binary outcomes. Denote the model prediction for $E[Y|A, \boldsymbol{X}]$ as $\hat{E}[Y|A, \boldsymbol{X}]$. (3) The DR-WLS estimator is

$$\widehat{\Delta}_{DR-WLS} = \frac{1}{n} \sum_{i=1}^{n} \{\hat{E}[Y|A = 1, \boldsymbol{X}_i] - \hat{E}[Y|A = 0, \boldsymbol{X}_i]\}.$$

The DR-WLS estimator can be equivalently calculated by solving estimating equations (1) with

$$\boldsymbol{\psi}(A_i, \boldsymbol{X}_i, Y_i, M_i; \boldsymbol{\theta}) = \begin{pmatrix} \{g(\beta_0 + \beta_A + \boldsymbol{\beta}_{\boldsymbol{X}}^t \boldsymbol{X}_i) - g(\beta_0 + \boldsymbol{\beta}_{\boldsymbol{X}}^t \boldsymbol{X}_i)\} - \Delta \\ \frac{M_i}{expit(\alpha_0 + \alpha_A A_i + \boldsymbol{\alpha}_{\boldsymbol{X}}^t \boldsymbol{X}_i)} \{Y_i - g(\beta_0 + \beta_A A_i + \boldsymbol{\beta}_{\boldsymbol{X}}^t \boldsymbol{X}_i)\} \boldsymbol{Z}_i \\ \{M_i - expit(\alpha_0 + \alpha_A A_i + \boldsymbol{\alpha}_{\boldsymbol{X}}^t \boldsymbol{X}_i)\} \boldsymbol{Z}_i \end{pmatrix}.$$

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

The ANCOVA estimator and the standardized logistic regression estimator are special cases of the DR-WLS estimator. If there are no missing data, which means $M_i \equiv 1$ for $a = 0, 1$ and $i = 1, \ldots, n$, then $\widehat{\Delta}_{DR-WLS}$ reduces to $\widehat{\Delta}_{ancova}$ for continuous outcomes and to $\widehat{\Delta}_{logistic}$ for binary outcomes.

5.4 *Estimators for time-to-event outcomes*

We give several examples of estimators for time-to-event outcomes to which our Theorem 4.1 applies, under our Assumption 1 and regularity conditions in Appendix A. All of these estimators can be represented as M-estimators, which is why our general approach applies to them.

In survival analysis, Y_i is the failure time and M_i is the censoring time. One parameter of interest is the restricted mean survival time, defined as $\Delta^* = E[\min\{Y(1), \tau\} - \min\{Y(0), \tau\}]$, where τ is a restriction time. Estimators of Δ^* can be found in Van der Laan et al. (2003); Díaz et al. (2019).

5.5 *Results for the ANCOVA estimator, standardized logistic regression estimator and DR-WLS estimator*

In Appendix C, we prove several results that apply to the estimators Sections 5.1-5.3. In Corollary 1, we show that the ANCOVA estimator and standardized logistic regression estimator remain model-robust and the DR-WLS estimator is doubly-robust under stratified randomization or the biased-coin design.

The sandwich variance estimator of $\hat{\boldsymbol{\theta}}$ (typically used when data are assumed i.i.d.) is defined as

$$\widehat{Var}(\hat{\boldsymbol{\theta}}) = \left\{ \frac{1}{n} \sum_{i=1}^{n} \frac{\partial}{\partial \boldsymbol{\theta}} \boldsymbol{\psi}(A_i, \boldsymbol{X}_i, Y_i M_i, M_i; \boldsymbol{\theta}) \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} \right\}^{-1} \left\{ \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{\psi}(A_i, \boldsymbol{X}_i, Y_i M_i, M_i; \hat{\boldsymbol{\theta}}) \boldsymbol{\psi}(A_i, \boldsymbol{X}_i, Y_i M_i, M_i; \hat{\boldsymbol{\theta}})^t \right\}$$
$$\left\{ \frac{1}{n} \sum_{i=1}^{n} \frac{\partial}{\partial \boldsymbol{\theta}} \boldsymbol{\psi}(A_i, \boldsymbol{X}_i, Y_i M_i, M_i; \boldsymbol{\theta}) \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} \right\}^{-1\,t}$$

https://biostats.bepress.com/jhubiostat/paper293

from Section 3.2 of Tsiatis (2007) or Theorem 5.21 of van der Vaart (1998); the sandwich variance estimator for Δ^* is then the first-row first-column entry of $\widehat{Var}(\hat{\boldsymbol{\theta}})$.

The following corollary gives conditions for when the asymptotic distribution of the estimators in the subsections above are the same regardless of whether simple randomization, stratified randomization, or the biased-coin design is used. Under such conditions, the estimators and their corresponding sandwich variance estimators can be used to perform valid hypothesis tests and construct confidence intervals (without being conservative).

Corollary 1 For the ANCOVA estimator, standardized logistic regression estimator and DR-WLS estimator, we assume their estimating equations satisfy Assumption 1 and regularity conditions in Appendix A. For the DR-WLS estimator, we further assume at least one of the two working models (5) and (6) is correctly specified. If (1) $\pi = 0.5$, or (2) the outcome regression model includes a treatment-by-strata interaction term, or (3) the outcome regression model is correctly specified, then under stratified randomization or the biased-coin design, then these estimators are consistent and asymptotically normally distributed with asymptotic variance $V = \tilde{V}$. Furthermore, the sandwich variance estimator is consistent (for the true asymptotic variance).

6. Clinical trial applications

Table 1 summaries our data analyses for each application (NIDA-CTN-0003, NIDA-CTN-0030, NIDA-CTN-0044). All missing baseline values were imputed by the median for continuous variables and mode for binary or categorical variables. When implementing the ANCOVA estimator or standardized logistic regression estimator, all participants with missing outcomes were removed from the analysis. Estimates and standard errors are rounded to the nearest 0.01. "Confidence Interval" is abbreviated as "CI". For all of the three trials, negative (positive) estimates are in the direction of clinical benefit (harm).

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

Table 1. Summary of clinical trial data analyses. The first column is the study ID. The second column gives the estimator with 95% confidence intervals (CI) adjusting for strata only. The third column gives the estimator with 95% confidence intervals (CI) adjusting for strata and additional baseline variables. The fourth column gives the DR-WLS estimator with 95% confidence intervals (CI) adjusting for strata and additional baseline variables. The fifth column shows the variance reduction due to adjustment for additional baseline variables comparing the second and third column.

Study	Estimator adjusting for strata only (95% CI)	Estimator adjusting for all baseline variables (95% CI)	DR-WLS estimator (95% CI)	Proportional variance reduction
NIDA-CTN-0003	-0.11(-0.21, -0.01)	-0.10(-0.19, -0.02)	-0.10(-0.18, -0.02)	35%
NIDA-CTN-0030	0.02(-0.02, 0.05)	0.01(-0.02, 0.05)	0.01(-0.02, 0.05)	17%
NIDA-CTN-0044	-0.09(-0.14, -0.03)	-0.09(-0.14, -0.03)	-0.09(-0.15, -0.03)	2%

For NIDA-CTN-0003, the standardized logistic regression estimator was used since the outcome was binary. If strata were adjusted only, the estimated absolute risk difference of getting negative urine lab result was -0.11 with 95% CI $(-0.21, -0.01)$; if strata as well as additional baseline variables were adjusted, the point estimate became -0.10 with 95% CI $(-0.19, -0.02)$. Though the point estimates are similar, the variance reduction due to adjusting for additional baseline variables is 35%, indicating that researchers planning to perform adjustment for strata and additional baseline variables could achieve the same precision as adjusting for strata only with approximately 35% fewer participants. The DR-WLS estimator was -0.10 with 95% CI $(-0.18, -0.02)$, which had 1% wider 95% CI than the standardized logistic regression estimator with adjustment for strata and additional baseline variables. This precision loss is typical; that is the price paid for added robustness to model misspecification, since the DR-WLS estimator is consistent under missing at random when at least one of its working models is correct, while the other estimators are generally only consistent when outcomes are missing completely at random or when the outcome regression working model is correct.

NIDA-CTN-0030 and NIDA-CTN-0044 had continuous outcomes. Adjustment for additional baseline variables brings 17% and 2% variance reduction for NIDA-CTN-0030 and NIDA-CTN-0044, respectively, compared to adjusting only for the strata. When additional baseline variables are not strongly prognostic, such as in NIDA-CTN-0044, the variance reduction from additional

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

baseline variables can be small.

7. Discussion

There is potential to substantially improve precision by adjusting for additional baseline variables than the ones used in the randomization procedure. For example, for NIDA-CTN-0003 and NIDA-CTN-0030, adjustment for additional baseline variables brings 35% and 17% variance reduction respectively. Our results show that there is no problem to implement this, and that many estimators used in randomized trials are consistent, asymptotically normal with variance that can be consistently estimated using our formula in Theorem 4.1. This asymptotic variance may be less than in the i.i.d. case, and our variance formula captures any added precision gain from stratified randomization (asymptotically).

The key to improving precision is adjusting for strongly prognostic baseline variables, if they exist. At the outset, one could use previous trials or observational data from a similar population to measure the prognostic value added by a set of baseline variables. This can be done by fitting two models, with one adjusting for the set of baseline variables and the other one not, and comparing their sandwich variance estimates.

Our asymptotics, as essentially all asymptotic results under the commonly used superpopulation inference framework, assume that the number of strata is fixed and the number of participants in each stratum goes to infinity. This may be a reasonable approximation when no stratum has a small number of participants. In our data examples, the smallest stratum has 49 participants. An area of future research is to consider cases where some strata have few participants.

Acknowledgement

This project was supported by a research award from Arnold Ventures. The content is solely the responsibility of the authors and does not necessarily represent the official views of Arnold

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

Hosted by The Berkeley Electronic Press

Ventures. The information reported here results from secondary analyses of data from clinical trials conducted by the National Institute on Drug Abuse (NIDA). Specifically, data from NIDA–CTN-0003 (Suboxone (Buprenorphine/Naloxone) Taper: A Comparison of Two Schedules), NIDA-CTN-0030 (Prescription Opioid Addiction Treatment Study) and NIDA-CTN-0044 (Web-delivery of Evidence-Based, Psychosocial Treatment for Substance Use Disorders) were included. NIDA databases and information are available at (https://datashare.nida.nih.gov).

References

Bugni, F. A., Canay, I. A., and Shaikh, A. M. (2018). Inference under covariate-adaptive randomization. *Journal of the American Statistical Association* **113,** 1784–1796.

Campbell, A. N., Nunes, E. V., Matthews, A. G., Stitzer, M., Miele, G. M., Polsky, D., Turrigiano, E., Walters, S., McClure, E. A., Kyle, T. L., Wahle, A., Van Veldhuisen, P., Goldman, B., Babcock, D., Stabile, P. Q., Winhusen, T., and Ghitza, U. E. (2014). Internet-delivered treatment for substance abuse: A multisite randomized controlled trial. *American Journal of Psychiatry* **171,** 683–690.

Díaz, I., Colantuoni, E., Hanley, D. F., and Rosenblum, M. (2019). Improved precision in the analysis of randomized trials with survival outcomes, without assuming proportional hazards. *Lifetime Data Analysis* **25,** 439–468.

Efron, B. (1971). Forcing a sequential experiment to be balanced. *Biometrika* **58,** 403–417.

EMA (2015). European Medicines Agency Guideline on Adjustment for Baseline Covariates in Clinical Trials. Reference number EMA/CHMP/295050/2013. Committee for Medicinal Products for Human Use (CHMP).

FDA (2019). Adjusting for Covariates in Randomized Clinical Trials for Drugs and Biologics

COBRA
Collection of Biostatistics
Research Archive

with Continuous Outcomes Guidance for Industry. Docket Number:2019-08470. Center for Drug Evaluation and Research.

FDA and EMA (1998). E9 statistical principles for clinical trials. *U.S. Food and Drug Administration: CDER/CBER. European Medicines Agency: CPMP/ICH/363/96* .

Kahan, B. C. and Morris, T. P. (2012). Improper analysis of trials randomised using stratified blocks or minimisation. *Statistics in Medicine* **31,** 328–340.

Lachin, J., Matts, J., and Wei, L. (1988). Randomization in clinical trials: Conclusions and recommendations. *Controlled Clinical Trials* **9,** 365 – 374.

Li, X. and Ding, P. (2019). Rerandomization and regression adjustment. *arXiv* page https://arxiv.org/abs/1906.11291.

Lin, Y., Zhu, M., and Su, Z. (2015). The pursuit of balance: An overview of covariate-adaptive randomization techniques in clinical trials. *Contemporary Clinical Trials* **45,** 21 – 25. 10th Anniversary Special Issue.

Ling, W., Hillhouse, M., Domier, C., Doraimani, G., Hunter, J., Thomas, C., Jenkins, J., Hasson, A., Annon, J., Saxon, A., Selzer, J., Boverman, J., and Bilangi, R. (2009). Buprenorphine tapering schedule and illicit opioid use. *Addiction* **104,** 256–265.

Ma, W., Hu, F., and Zhang, L. (2015). Testing hypotheses of covariate-adaptive randomized clinical trials. *Journal of the American Statistical Association* **110,** 669–680.

Ma, W., Qin, Y., Li, Y., and Hu, F. (2018). Statistical inference of covariate-adjusted randomized experiments. *arXiv* page https://arxiv.org/abs/1807.09678.

Moore, K. and van der Laan, M. (2009). Covariate adjustment in randomized trials with binary outcomes: Targeted maximum likelihood estimation. *Statistics in Medicine* **28,** 39–64.

Hosted by The Berkeley Electronic Press

Morgan, K. L. and Rubin, D. B. (2012). Rerandomization to improve covariate balance in experiments. *The Annals of Statistics* **40,** 1263–1282.

Pocock, S. J. and Simon, R. (1975). Sequential treatment assignment with balancing for prognostic factors in the controlled clinical trial. *Biometrics* **31,** 103–115.

Robins, J., Sued, M., Lei-Gomez, Q., and Rotnitzky, A. (2007). Comment: Performance of double-robust estimators when "inverse probability" weights are highly variable. *Statist. Sci.* **22,** 544–559.

Scharfstein, D. O., Rotnitzky, A., and Robins, J. M. (1999). Adjusting for nonignorable drop-out using semiparametric nonresponse models. *Journal of the American Statistical Association* **94,** 1096–1120.

Shao, J. and Yu, X. (2013). Validity of tests under covariate-adaptive biased coin randomization and generalized linear models. *Biometrics* **69,** 960–969.

Shao, J., YU, X., and ZHONG, B. (2010). A theory for testing hypotheses under covariate-adaptive randomization. *Biometrika* **97,** 347–360.

Shorack, G. R. and Wellner, J. A. (2009). *Empirical processes with applications to statistics.* Society for Industrial and Applied Mathematics.

Steingrimsson, J., Hanley, D., and Rosenblum, M. (2016). Improving precision by adjusting for baseline variables in randomized trials with binary outcomes, without regression model assumptions. *Contemporary Clinical Trials* **54,** 18–24.

Tsiatis, A. (2007). *Semiparametric theory and missing data.* Springer Science & Business Media.

Van der Laan, M. J., Laan, M., and Robins, J. M. (2003). *Unified methods for censored longitudinal data and causality.* Springer Science & Business Media.

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

van der Vaart, A. (1998). *Asymptotic Statistics.* Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press.

Wei, L. J. (1978). The adaptive biased coin design for sequential experiments. *The Annals of Statistics* **6,** 92–100.

Weiss, R. D., Potter, J. S., Fiellin, D. A., Byrne, M., Connery, H. S., Dickinson, W., Gardin, J., Griffin, M. L., Gourevitch, M. N., Haller, D. L., Hasson, A. L., Huang, Z., Jacobs, P., Kosinski, A. S., Lindblad, R., McCance-Katz, E. F., Provost, S. E., Selzer, J., Somoza, E. C., Sonne, S. C., and Ling, W. (2011). Adjunctive Counseling During Brief and Extended Buprenorphine-Naloxone Treatment for Prescription Opioid Dependence: A 2-Phase Randomized Controlled Trial. *JAMA Psychiatry* **68,** 1238–1246.

Yang, L. and Tsiatis, A. (2001). Efficiency study of estimators for a treatment effect in a pretest-posttest trial. *The American Statistician* **55,** 314–321.

Ye, T. and Shao, J. (2019). Validity and robustness of tests in survival analysis under covariate-adaptive randomization. *arXiv* page https://arxiv.org/abs/1811.07232.

Zelen, M. (1974). The randomization and stratification of patients to clinical trials. *Journal of Chronic Diseases* **27,** 365 – 375.

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

APPENDIX

A. Regularity conditions

The regularity conditions similar to Section 5.3 of van der Vaart (1998) are specifed below:

(1) $\boldsymbol{\theta} \in \boldsymbol{\theta}$, a compact set in $\mathbb{R}^{p+1}$.

(2) $E[\boldsymbol{\psi}(a, \boldsymbol{X}, Y(a)M(a), M(a); \boldsymbol{\theta})^2] < \infty$ for any $\boldsymbol{\theta} \in \boldsymbol{\theta}$ and $a \in \{0, 1\}$.

(3) There exists a zero, denoted as $\underline{\boldsymbol{\theta}} = (\underline{\Delta}, \underline{\beta}^t)^t$, of the expectation of estimating equations

$$\pi E[\boldsymbol{\psi}(1, \boldsymbol{X}, Y(1)M(1), M(1); \boldsymbol{\theta})] + (1-\pi)E[\boldsymbol{\psi}(0, \boldsymbol{X}, Y(0)M(0), M(0); \boldsymbol{\theta})] = 0.$$

Furthermore, there is a $\eta > 0$ such that $\underline{\boldsymbol{\theta}}$ is the only zero in its neighborhood $\{\boldsymbol{\theta} : ||\boldsymbol{\theta} - \underline{\boldsymbol{\theta}}|| < \eta\}$.

(4) functions $\boldsymbol{\theta} \mapsto \boldsymbol{\psi}(a, x, y, m; \boldsymbol{\theta})$ are twice continuously differentiable for every (a, x, y, m) in the support of $(A, \boldsymbol{X}, Y(a)M(a), M(a))$ and dominated by an integrable function $u(A, \boldsymbol{X}, Y(a)M(a), M(a))$.

(5) There exist a $C > 0$ and and integrable function $v(\boldsymbol{X}, Y(a)M(a), M(a))$, such that $|\frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^t}\boldsymbol{\psi}(a, x, y, m; \boldsymbol{\theta})| < v(x, y, m)$ element-wise for every (a, x, y, m) in the support of $(A, \boldsymbol{X}, Y(a)M(a), M(a))$ and $||\boldsymbol{\theta} - \underline{\boldsymbol{\theta}}|| < C$.

(6) $\frac{\partial}{\partial\boldsymbol{\theta}}\boldsymbol{\psi}(a, \boldsymbol{X}, Y(a)M(a), M(a); \boldsymbol{\theta})\Big|_{\boldsymbol{\theta}=\underline{\boldsymbol{\theta}}}$ has finite second moment for $a = 0, 1$ and

$$\pi E\left[\frac{\partial}{\partial\boldsymbol{\theta}}\boldsymbol{\psi}(1, \boldsymbol{X}, Y(1)M(1), M(1); \boldsymbol{\theta})\Big|_{\boldsymbol{\theta}=\underline{\boldsymbol{\theta}}}\right] + (1-\pi)E\left[\frac{\partial}{\partial\boldsymbol{\theta}}\boldsymbol{\psi}(0, \boldsymbol{X}, Y(0)M(1), M(0); \boldsymbol{\theta})\Big|_{\boldsymbol{\theta}=\underline{\boldsymbol{\theta}}}\right]$$

is invertible.

B. Consistent estimator of $\boldsymbol{B}, \boldsymbol{C}$ and $\boldsymbol{D}$ in Theorem 4.1

$$\hat{\boldsymbol{B}} = \frac{1}{n}\sum_{i=1}^{n}\frac{\partial}{\partial\boldsymbol{\theta}}\boldsymbol{\psi}(A_i, \boldsymbol{X}_i, Y_iM_i, M_i; \boldsymbol{\theta})\Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}$$

$$\hat{\boldsymbol{C}} = \frac{1}{|\mathcal{S}|}\sum_{s\in\mathcal{S}}\left[\sum_{i=1}^{n}\frac{I\{S_i = s\}}{\sum_{i=1}^{n} I\{S_i = s\}}\boldsymbol{\psi}(A_i, \boldsymbol{X}_i, Y_iM_i, M_i; \boldsymbol{\theta})^2 - \left\{\sum_{i=1}^{n}\frac{I\{S_i = s\}}{\sum_{i=1}^{n} I\{S_i = s\}}\boldsymbol{\psi}(A_i, \boldsymbol{X}_i, Y_iM_i, M_i; \boldsymbol{\theta})\right\}^2\right]$$

$$\hat{\boldsymbol{D}} = \frac{1}{|\mathcal{S}|}\sum_{s\in\mathcal{S}}\left[\sum_{i=1}^{n}\frac{I\{S_i = s\}}{\sum_{i=1}^{n} I\{S_i = s\}}\{A_i\boldsymbol{\psi}(1, \boldsymbol{X}_i, Y_iM_i, M_i; \boldsymbol{\theta}) - (1-A_i)\boldsymbol{\psi}(0, \boldsymbol{X}_i, Y_iM_i, M_i; \boldsymbol{\theta})\}\right]^2$$

Consistency of $\hat{\boldsymbol{B}}, \hat{\boldsymbol{C}}$ and $\hat{\boldsymbol{D}}$ to $\boldsymbol{B}, \boldsymbol{C}$ and $\boldsymbol{D}$ are proved in Appendix C.

https://biostats.bepress.com/jhubiostat/paper293

C. Proof of main results

In this section, we first present two lemmas that are critical for proving our main results. These two lemmas are adapted from results of Bugni et al. (2018) and the proofs of them are very similar to Bugni et al. (2018). Then we prove Theorem 4.1 based on Lemmas C.1 and C.2. To prove Corollary 1, we first introduce and prove Corollary 2, which gives the asymptotic distribution of the DR-WLS estimator and partially indicates Corollary 1.

Lemma C.1 Given Assumption 1, let $Z_i = h(Y_i(1), Y_i(0), M_i(1), M_i(0), \boldsymbol{X}_i)$ such that $E[|Z_i|] < \infty$. Then under stratified randomization or the biased-coin design, $\frac{1}{n}\sum_{i=1}^{n} Z_i A_i \xrightarrow{P} \pi E[Z_i]$.

Proof. See Lemma B.3 in the supplementary material of Bugni et al. (2018). The only difference is that we replace $(Y_i(1), Y_i(0), S_i)$ by $(Y_i(1), Y_i(0), M_i(1), M_i(0), \boldsymbol{X}_i)$ and all deduction still holds. □

Lemma C.2 Given Assumption 1, let $Z_i(1) = h_1(Y_i(1), M_i(1), \boldsymbol{X}_i)$ and $Z_i(0) = h_2(Y_i(0), M_i(0), \boldsymbol{X}_i)$ such that $Var(Z_i(a)) < \infty$ for $a = 0, 1$. Then under stratified randomization or the biased-coin design,

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\{(Z_i(1) - E[Z_i(1)])(1-\pi)A_i - (Z_i(0) - E[Z_i(0)])\pi(1-A_i)\} \xrightarrow{d} \pi(1-\pi)N(0, \sigma_1^2 + \sigma_2^2),$$

where

$$\sigma_1^2 = \frac{1}{\pi}Var(Z(1) - E[Z(1)|S]) + \frac{1}{1-\pi}Var(Z(0) - E[Z(0)|S]),$$
$$\sigma_2^2 = Var(E[Z(1) - Z(0)|S]).$$

Proof. See Lemma B.1 and Lemma B.2 in the supplementary material of Bugni et al. (2018). The only difference is that we replace $Y_i(a)$ by $Z_i(a)$ and all deduction still holds. □

Proof of Theorem 4.1. Using the fact that $Y_i = Y_i(1)A_i + Y_i(0)(1 - A_i)$ and $M_i = M_i(1)A_i +$

Hosted by The Berkeley Electronic Press

$M_i(0)(1-A_i)$, the estimating equations (1) can be re-written as

$$\frac{1}{n}\sum_{i=1}^{n}\{A_i\boldsymbol{\psi}(1,\boldsymbol{X}_i,Y_i(1)M_i(1),M_i(1);\boldsymbol{\theta})+(1-A_i)\boldsymbol{\psi}(0,\boldsymbol{X}_i,Y_i(0)M_i(0),M_i(0);\boldsymbol{\theta})\}=0.$$

We first show that $\hat{\boldsymbol{\theta}}\xrightarrow{P}\underline{\boldsymbol{\theta}}$, where $\underline{\boldsymbol{\theta}}$ is a vector that solves

$$\pi E[\boldsymbol{\psi}(1,\boldsymbol{X},Y(1)M(1),M(1);\boldsymbol{\theta})]+(1-\pi)E[\boldsymbol{\psi}(0,\boldsymbol{X},Y(0)M(0),M(0);\boldsymbol{\theta})]=0.$$

Regularity condition (3) in Appendix A implies that $\underline{\boldsymbol{\theta}}$ exists and is the only zero in its neighborhood. By Lemma C.1 and regularity condition (2), we have

$$\begin{aligned}&\frac{1}{n}\sum_{i=1}^{n}\{A_i\boldsymbol{\psi}(1,\boldsymbol{X}_i,Y_i(1)M_i(1),M_i(1);\boldsymbol{\theta})+(1-A_i)\boldsymbol{\psi}(0,\boldsymbol{X}_i,Y_i(0)M_i(0),M_i(0);\boldsymbol{\theta})\}\\&\quad\xrightarrow{P}\pi E[\boldsymbol{\psi}(1,\boldsymbol{X},Y(1)M(1),M(1);\boldsymbol{\theta})]+(1-\pi)E[\boldsymbol{\psi}(0,\boldsymbol{X},Y(0)M(0),M(0);\boldsymbol{\theta})].\end{aligned}$$

Combined with regularity conditions 2 (1) and (4), by Theorem 5.9 of Van der Vaart (1999), the above results imply $\hat{\boldsymbol{\theta}}$ converge in probability to $\underline{\boldsymbol{\theta}}$.

We then show $\hat{\boldsymbol{\theta}}$ is asymptotically linear. By multivariate Taylor expasion,

$$\begin{aligned}0&=\frac{1}{n}\sum_{i=1}^{n}\Big\{A_i\boldsymbol{\psi}(1,\boldsymbol{X}_i,Y_i(1)M_i(1),M_i(1);\hat{\boldsymbol{\theta}})+(1-A_i)\boldsymbol{\psi}(0,\boldsymbol{X}_i,Y_i(0)M_i(0),M_i(0);\hat{\boldsymbol{\theta}})\Big\}\\&=\frac{1}{n}\sum_{i=1}^{n}\{A_i\boldsymbol{\psi}(1,\boldsymbol{X}_i,Y_i(1)M_i(1),M_i(1);\underline{\boldsymbol{\theta}})+(1-A_i)\boldsymbol{\psi}(0,\boldsymbol{X}_i,Y_i(0)M_i(0),M_i(0);\underline{\boldsymbol{\theta}})\}\\&\quad+\frac{1}{n}\sum_{i=1}^{n}\Big\{A_i\dot{\boldsymbol{\psi}}(1,\boldsymbol{X}_i,Y_i(1)M_i(1),M_i(1);\underline{\boldsymbol{\theta}})+(1-A_i)\dot{\boldsymbol{\psi}}(0,\boldsymbol{X}_i,Y_i(0)M_i(0),M_i(0);\underline{\boldsymbol{\theta}})\Big\}(\hat{\boldsymbol{\theta}}-\underline{\boldsymbol{\theta}})\\&\quad+\frac{1}{2}(\hat{\boldsymbol{\theta}}-\underline{\boldsymbol{\theta}})^t\frac{1}{n}\sum_{i=1}^{n}\Big\{A_i\ddot{\boldsymbol{\psi}}(1,\boldsymbol{X}_i,Y_i(1)M_i(1),M_i(1);\tilde{\boldsymbol{\theta}})+(1-A_i)\ddot{\boldsymbol{\psi}}(0,\boldsymbol{X}_i,Y_i(0)M_i(0),M_i(0);\tilde{\boldsymbol{\theta}})\Big\}(\hat{\boldsymbol{\theta}}-\underline{\boldsymbol{\theta}}),\end{aligned}$$

where

$$\begin{aligned}\dot{\boldsymbol{\psi}}(a,\boldsymbol{X}_i,Y_i(a)M_i(a),M_i(a);\underline{\boldsymbol{\theta}})&=\frac{\partial}{\partial\boldsymbol{\theta}}\boldsymbol{\psi}(a,\boldsymbol{X}_i,Y_i(a)M_i(a),M_i(a);\boldsymbol{\theta})\Big|_{\boldsymbol{\theta}=\underline{\boldsymbol{\theta}}}\\\ddot{\boldsymbol{\psi}}(a,\boldsymbol{X}_i,Y_i(a)M_i(a),M_i(a);\tilde{\boldsymbol{\theta}})&=\frac{\partial}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^t}\boldsymbol{\psi}(a,\boldsymbol{X}_i,Y_i(a)M_i(a),M_i(a);\boldsymbol{\theta})\Big|_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}},\end{aligned}$$

for $a=0,1$ and $\tilde{\boldsymbol{\theta}}$ is a random point on the line segment between $\hat{\boldsymbol{\theta}}$ and $\underline{\boldsymbol{\theta}}$.

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

According to regularity condition (5), there exists a ball B around $\underline{\boldsymbol{\theta}}$ such that $\ddot{\boldsymbol{\psi}}(a, \boldsymbol{X}_i, Y_i(a)M_i(a), M_i(a); \boldsymbol{\theta})$ is dominated by a function $v(\boldsymbol{X}_i, Y_i(a)M_i(a), M_i(a))$. Hence, if $\tilde{\boldsymbol{\theta}} \in B$, then

$$\begin{aligned}&\Big|\Big|\frac{1}{n}\sum_{i=1}^{n}\Big\{A_i\ddot{\boldsymbol{\psi}}(1, \boldsymbol{X}_i, Y_i(1)M_i(1), M_i(1); \tilde{\boldsymbol{\theta}}) + (1-A_i)\ddot{\boldsymbol{\psi}}(0, \boldsymbol{X}_i, Y_i(0)M_i(0), M_i(0); \tilde{\boldsymbol{\theta}})\Big\}\Big|\Big| \\ &\quad \leqslant \frac{1}{n}\sum_{i=1}^{n}\{||v(\boldsymbol{X}_i, Y_i(1)M_i(1), M_i(1))|| + ||v(\boldsymbol{X}_i, Y_i(0)M_i(0), M_i(0))||\},\end{aligned}$$

which is bounded in probability by the law of large numbers. Furthermore, since $P(\tilde{\boldsymbol{\theta}} \in B) \to 1$, we have

$$\frac{1}{n}\sum_{i=1}^{n}\Big\{A_i\ddot{\boldsymbol{\psi}}(1, \boldsymbol{X}_i, Y_i(1)M_i(1), M_i(1); \tilde{\boldsymbol{\theta}}) + (1-A_i)\ddot{\boldsymbol{\psi}}(0, \boldsymbol{X}_i, Y_i(0)M_i(0), M_i(0); \tilde{\boldsymbol{\theta}})\Big\} = O_p(1).$$

Denoting $B = \pi E[\dot{\boldsymbol{\psi}}(1, \boldsymbol{X}, Y(1)M(1), M(1); \underline{\boldsymbol{\theta}})] + (1-\pi)E[\dot{\boldsymbol{\psi}}(0, \boldsymbol{X}, Y(0)M(0), M(0); \underline{\boldsymbol{\theta}})]$, by Lemma C.1 and regularity condition (6), we have

$$\frac{1}{n}\sum_{i=1}^{n}\Big\{A_i\dot{\boldsymbol{\psi}}(1, \boldsymbol{X}_i, Y_i(1)M_i(1), M_i(1); \underline{\boldsymbol{\theta}}) + (1-A_i)\dot{\boldsymbol{\psi}}(0, \boldsymbol{X}_i, Y_i(0)M_i(0), M_i(0); \underline{\boldsymbol{\theta}})\Big\} \xrightarrow{P} \boldsymbol{B}.$$

Combination of above facts implies that

$$\begin{aligned}&\frac{1}{n}\sum_{i=1}^{n}\{A_i\boldsymbol{\psi}(1, \boldsymbol{X}_i, Y_i(1)M_i(1), M_i(1); \underline{\boldsymbol{\theta}}) + (1-A_i)\boldsymbol{\psi}(0, \boldsymbol{X}_i, Y_i(0)M_i(0), M_i(0); \underline{\boldsymbol{\theta}})\} \\ &\quad = (\hat{\boldsymbol{\theta}} - \underline{\boldsymbol{\theta}})(\boldsymbol{B} + o_p(1)) + (\hat{\boldsymbol{\theta}} - \underline{\boldsymbol{\theta}})^t O_p(1)(\hat{\boldsymbol{\theta}} - \underline{\boldsymbol{\theta}}).\end{aligned}$$

Since $\hat{\boldsymbol{\theta}} \xrightarrow{P} \underline{\boldsymbol{\theta}}$, we get

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \underline{\boldsymbol{\theta}}) = -\boldsymbol{B}^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\{A_i\boldsymbol{\psi}(1, \boldsymbol{X}_i, Y_i(1)M_i(1), M_i(1); \underline{\boldsymbol{\theta}}) + (1-A_i)\boldsymbol{\psi}(0, \boldsymbol{X}_i, Y_i(0)M_i(0), M_i(0); \underline{\boldsymbol{\theta}})\} + o_p(1).$$

Letting $\boldsymbol{Z}_i(1) = \pi\boldsymbol{\psi}(1, \boldsymbol{X}_i, Y_i(1)M_i(1), M_i(1); \underline{\boldsymbol{\theta}})$ and $\boldsymbol{Z}_i(0) = (1-\pi)\boldsymbol{\psi}(0, \boldsymbol{X}_i, Y_i(0)M_i(0), M_i(0); \underline{\boldsymbol{\theta}})$, by applying Lemma C.2 to the first entry of $\hat{\boldsymbol{\theta}}$, we get the desired asymptotic normality.

To show that $V \leqslant \tilde{V}$, we use the result by Theorem 5.21 of van der Vaart (1998) that $\tilde{V}$ is the first-row first-column entry of $\boldsymbol{B}^{-1}\boldsymbol{E}\boldsymbol{B}^{-1t}$, where

$$\boldsymbol{E} = \pi Var\{\boldsymbol{\psi}(1, \boldsymbol{X}, Y(1)M(1), M(1); \underline{\boldsymbol{\theta}})\} + (1-\pi)Var\{\boldsymbol{\psi}(0, \boldsymbol{X}, Y(0)M(0), M(0); \underline{\boldsymbol{\theta}})\}.$$

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

Since $\boldsymbol{C}, \boldsymbol{D}, \boldsymbol{E}$ are all semi-positive definite, it suffices to show that $\boldsymbol{E} - \boldsymbol{C} - \boldsymbol{D}$ is semi-positive definite. This is implied by the following derivation:

$$\begin{aligned}
\boldsymbol{E} - \boldsymbol{C} - \boldsymbol{D} &= \pi Var\{\boldsymbol{\psi}(1, \boldsymbol{X}, Y(1)M(1), M(1); \underline{\boldsymbol{\theta}})\} - \pi E[Var\{\boldsymbol{\psi}(1, \boldsymbol{X}, Y(1)M(1), M(1); \underline{\boldsymbol{\theta}})|S\}] \\
&\quad + (1-\pi)Var\{\boldsymbol{\psi}(0, \boldsymbol{X}, Y(0)M(0), M(0); \underline{\boldsymbol{\theta}})\} - (1-\pi)E[Var\{\boldsymbol{\psi}(0, \boldsymbol{X}, Y(0)M(0), M(0); \underline{\boldsymbol{\theta}})|S\}] - \boldsymbol{D} \\
&= \pi Var\{E[\boldsymbol{\psi}(1, \boldsymbol{X}, Y(1)M(1), M(1); \underline{\boldsymbol{\theta}})|S]\} + (1-\pi)Var\{E[\boldsymbol{\psi}(0, \boldsymbol{X}, Y(0)M(0), M(0); \underline{\boldsymbol{\theta}})|S]\} \\
&\quad - E[E[\pi\boldsymbol{\psi}(1, \boldsymbol{X}, Y(1)M(1), M(1); \underline{\boldsymbol{\theta}}) - (1-\pi)\boldsymbol{\psi}(0, \boldsymbol{X}, Y(0)M(0), M(0); \underline{\boldsymbol{\theta}})|S]^2] \\
&= \pi(1-\pi)Var\left(\frac{1}{\pi}E[\boldsymbol{\psi}(1, \boldsymbol{X}_i, Y(1)M(1), M(1); \underline{\boldsymbol{\theta}})|S] + \frac{1}{1-\pi}E[\boldsymbol{\psi}(0, \boldsymbol{X}, Y(0)M(0), M(0); \underline{\boldsymbol{\theta}})|S]\right).
\end{aligned}$$

To prove that $\boldsymbol{B}, \boldsymbol{C}, \boldsymbol{D}$ can be consistently estimated by $\hat{\boldsymbol{B}}, \hat{\boldsymbol{C}}, \hat{\boldsymbol{D}}$ defined in Section B respectively, we show that $\hat{\boldsymbol{B}} \xrightarrow{P} \boldsymbol{B}$ for demonstration and other quantities converge with a similar proof. By multivariate Taylor's expansion,

$$\begin{aligned}
&\frac{1}{n\pi}\sum_{i=1}^{n} A_i \frac{\partial}{\partial \boldsymbol{\theta}}\boldsymbol{\psi}(1, \boldsymbol{X}_i, Y_iM_i, M_i; \boldsymbol{\theta})\Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} - \frac{1}{n\pi}\sum_{i=1}^{n} A_i \frac{\partial}{\partial \boldsymbol{\theta}}\boldsymbol{\psi}(1, \boldsymbol{X}_i, Y_iM_i, M_i; \boldsymbol{\theta})\Big|_{\boldsymbol{\theta}=\underline{\boldsymbol{\theta}}} \\
&\quad = \frac{1}{n\pi}\sum_{i=1}^{n} A_i \ddot{\boldsymbol{\psi}}(1, \boldsymbol{X}_i, Y_iM_i, M_i; \tilde{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}} - \underline{\boldsymbol{\theta}}) \\
&\quad = \frac{1}{n\pi}\sum_{i=1}^{n} A_i \ddot{\boldsymbol{\psi}}(1, \boldsymbol{X}_i, Y_i(1)M_i(1), M_i(1); \tilde{\boldsymbol{\theta}})(\hat{\boldsymbol{\theta}} - \underline{\boldsymbol{\theta}})
\end{aligned}$$

Regularity condition (5) indicates that $\frac{1}{n}\sum_{i=1}^{n} A_i \ddot{\boldsymbol{\psi}}(1, \boldsymbol{X}_i, Y_i(1)M_i(1), M_i(1); \tilde{\boldsymbol{\theta}}) = O_p(1)$ since $\hat{\boldsymbol{\theta}} \xrightarrow{P} \underline{\boldsymbol{\theta}}$. As a result, by Lemma C.1,

$$\begin{aligned}
\frac{1}{n\pi}\sum_{i=1}^{n} A_i \frac{\partial}{\partial \boldsymbol{\theta}}\boldsymbol{\psi}(1, \boldsymbol{X}_i, Y_iM_i, M_i; \boldsymbol{\theta})\Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} &= \frac{1}{n\pi}\sum_{i=1}^{n} A_i \frac{\partial}{\partial \boldsymbol{\theta}}\boldsymbol{\psi}(1, \boldsymbol{X}_i, Y_i(1)M_i(1), M_i(1); \boldsymbol{\theta})\Big|_{\boldsymbol{\theta}=\underline{\boldsymbol{\theta}}} + o_p(1) \\
&= E\left[\frac{\partial}{\partial \boldsymbol{\theta}}\boldsymbol{\psi}(1, \boldsymbol{X}, Y(1)M(1), M(1); \boldsymbol{\theta})\Big|_{\boldsymbol{\theta}=\underline{\boldsymbol{\theta}}}\right] + o_p(1).
\end{aligned}$$

Similarly, we have

$$\frac{1}{n(1-\pi)}\sum_{i=1}^{n}(1-A_i)\frac{\partial}{\partial \boldsymbol{\theta}}\boldsymbol{\psi}(0, \boldsymbol{X}_i, Y_iM_i, M_i; \boldsymbol{\theta})\Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} = E\left[\frac{\partial}{\partial \boldsymbol{\theta}}\boldsymbol{\psi}(0, \boldsymbol{X}, Y(0)M(0), M(0); \boldsymbol{\theta})\Big|_{\boldsymbol{\theta}=\underline{\boldsymbol{\theta}}}\right] + o_p(1).$$

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

Then we get

$$\begin{aligned}
\hat{\boldsymbol{B}} &= \frac{1}{n}\sum_{i=1}^{n}\frac{\partial}{\partial\boldsymbol{\theta}}\boldsymbol{\psi}(A_i,\boldsymbol{X}_i,Y_iM_i,M_i;\boldsymbol{\theta})\Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} \\
&= \frac{1}{n}\sum_{i=1}^{n}A_i\frac{\partial}{\partial\boldsymbol{\theta}}\boldsymbol{\psi}(1,\boldsymbol{X}_i,Y_iM_i,M_i;\boldsymbol{\theta})\Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} + \frac{1}{n}\sum_{i=1}^{n}(1-A_i)\frac{\partial}{\partial\boldsymbol{\theta}}\boldsymbol{\psi}(0,\boldsymbol{X}_i,Y_iM_i,M_i;\boldsymbol{\theta})\Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} \\
&= \pi E\left[\frac{\partial}{\partial\boldsymbol{\theta}}\boldsymbol{\psi}(1,\boldsymbol{X},Y(1)M(1),M(1);\boldsymbol{\theta})\Big|_{\boldsymbol{\theta}=\underline{\boldsymbol{\theta}}}\right] + (1-\pi)E\left[\frac{\partial}{\partial\boldsymbol{\theta}}\boldsymbol{\psi}(0,\boldsymbol{X},Y(0)M(0),M(0);\boldsymbol{\theta})\Big|_{\boldsymbol{\theta}=\underline{\boldsymbol{\theta}}}\right] + o_p(1) \\
&= \boldsymbol{B} + o_p(1).
\end{aligned}$$

For $\hat{\boldsymbol{C}}$ and $\hat{\boldsymbol{D}}$, regularity conditions (2), (4), (5) and (6) bounded the derivatives in probability and Lemma C.1 implies their consistency. □

Corollary 2 Given Assumption 1 and regularity conditions in Appendix A, if at least one of the two working models (5) and (6) is correctly specified, then under stratified randomization or the biased-coin design,

$$\sqrt{n}(\widehat{\Delta}_{DR-WLS} - \Delta^*) \xrightarrow{d} N(0, V_{DR-WLS}),$$

where

$$V_{DR-WLS} = \tilde{V}_{DR-WLS} - \frac{(1-2\pi)^2}{\pi(1-\pi)}Var\left\{E\left[M(1)\frac{Y(1)-h(1,\boldsymbol{X})}{e(1,\boldsymbol{X})} - M(0)\frac{Y(0)-h(0,\boldsymbol{X})}{e(0,\boldsymbol{X})}\Big|S\right]\right\},$$

where $\tilde{V}_{DR-WLS}$ is defined in Equation (C.1) of Appendix C, $h(a,\boldsymbol{X}) = g(\underline{\beta}_0 + \underline{\beta}_A a + \underline{\boldsymbol{\beta}}_{\boldsymbol{X}}^t\boldsymbol{X})$, $e(a,\boldsymbol{X}) = expit(\underline{\alpha}_0 + \underline{\alpha}_A a + \underline{\boldsymbol{\alpha}}_{\boldsymbol{X}}^t\boldsymbol{X})$ for $a = 0,1$ and $\underline{\beta}_0, \underline{\beta}_A, \underline{\boldsymbol{\beta}}_{\boldsymbol{X}}, \underline{\alpha}_0, \underline{\alpha}_A, \underline{\boldsymbol{\alpha}}_{\boldsymbol{X}}$ are probability limits of $\hat{\beta}_0, \hat{\beta}_A, \hat{\boldsymbol{\beta}}_{\boldsymbol{X}}, \hat{\alpha}_0, \hat{\alpha}_A, \hat{\boldsymbol{\alpha}}_{\boldsymbol{X}}$ respectively.

Furthermore, the sandwich variance estimator of $\widehat{\Delta}_{DR-WLS}$ converges in probability to $\tilde{V}_{DR-WLS}$.

Proof of Corollary 2. For the DR-WLS estimator, we have

$$\boldsymbol{\psi}(a,\boldsymbol{X},Y(a)M(a),M(a);\boldsymbol{\theta}) = \begin{pmatrix} \{g(\beta_0+\beta_A+\boldsymbol{\beta}_{\boldsymbol{X}}^t\boldsymbol{X}) - g(\beta_0+\boldsymbol{\beta}_{\boldsymbol{X}}^t\boldsymbol{X})\} - \Delta \\ \frac{M(a)}{expit(\alpha_0+\alpha_A a+\boldsymbol{\alpha}_{\boldsymbol{X}}^t\boldsymbol{X})}\{Y(a) - g(\beta_0+\beta_A a+\boldsymbol{\beta}_{\boldsymbol{X}}^t\boldsymbol{X})\}Z \\ \{M(a) - expit(\alpha_0+\alpha_A a+\boldsymbol{\alpha}_{\boldsymbol{X}}^t\boldsymbol{X})\}Z \end{pmatrix},$$

where $Z = (1, A, \boldsymbol{X}^t)^t$ is a p-dimensional column vector.

COBRA
Collection of Biostatistics
Research Archive

Given Assumption 1 and regularity conditions in Appendix A, Theorem 4.1 shows that

$$\sqrt{n}(\widehat{\Delta}_{DR-WLS} - \underline{\Delta}) \xrightarrow{d} N(0, V_{DR-WLS}),$$

where V_{DR-WLS} is the first-row first-column entry of matrix $\boldsymbol{B}^{-1}(\boldsymbol{C}+\boldsymbol{D})(\boldsymbol{B}^{-1})^t$, where $\boldsymbol{B}, \boldsymbol{C}, \boldsymbol{D}$ are defined in Theorem 4.1.

We first show that $\widehat{\Delta}_{DR-WLS}$ is doubly robust. If $E[M(a)|\boldsymbol{X}] = e(a, \boldsymbol{X})$ for $a = 0, 1$ then

$$0 = E[M(a)\frac{Y(a) - h(a, \boldsymbol{X})}{e(a, \boldsymbol{X})}] = E\left[E[M(a)|\boldsymbol{X}]\frac{E[Y(a)|\boldsymbol{X}] - h(a, \boldsymbol{X})}{e(a, \boldsymbol{X})}\right] = E[Y(a)] - E[h(a, \boldsymbol{X})],$$

where the missing at random assumption is used. If $E[Y(a)|\boldsymbol{X}] = h(a, \boldsymbol{X})$ for $a = 0, 1$, then $E[Y(a)] = E[E[Y(a)|\boldsymbol{X}]] = E[h(a, \boldsymbol{X})]$. As a result, if at least one of the two working models is correctly specified, then $\underline{\Delta} = E[h(1, \boldsymbol{X})] - E[h(0, \boldsymbol{X})] = \Delta^*$.

We then write out the expression of V_{DR-WLS}. Theorem 4.1 and tedious algebra show that

$$\sqrt{n}(\widehat{\Delta}_{DR-WLS} - \underline{\Delta}) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{\frac{A_i}{\pi}f_1(\boldsymbol{X}_i, Y_i(1)M_i(1), M_i(1)) - \frac{1-A_i}{1-\pi}f_0(\boldsymbol{X}_i, Y_i(0)M_i(0), M_i(0)) - \Delta^*\right\} + o_p(1),$$

where

$$\begin{aligned} f_1(\boldsymbol{X}, Y(1)M(1), M(1)) &= \{1 - \pi \boldsymbol{c}_1 \boldsymbol{Z}(1)\}M(1)\frac{Y(1) - h(1, \boldsymbol{X})}{e(1, \boldsymbol{X})} + \pi h(1, \boldsymbol{X}) - \pi h(0, \boldsymbol{X}) \\ &\quad - \pi \boldsymbol{c}_2 \boldsymbol{Z}(1)(M(1) - e(1, \boldsymbol{X})), \\ f_0(\boldsymbol{X}, Y(0)M(0), M(0)) &= \{1 + (1-\pi)\boldsymbol{c}_1 \boldsymbol{Z}(0)\}M(0)\frac{Y(0) - h(0, \boldsymbol{X})}{e(0, \boldsymbol{X})} - (1-\pi)h(1, \boldsymbol{X}) + (1-\pi)h(0, \boldsymbol{X}) \\ &\quad + (1-\pi)\boldsymbol{c}_2 \boldsymbol{Z}(0)(M(0) - e(0, \boldsymbol{X})). \end{aligned}$$

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

with

$$
\begin{aligned}
\boldsymbol{Z}(a) &= (1, a, \boldsymbol{X}^t)^t \\
\boldsymbol{c}_1 &= E\left[\left(\frac{M(1)}{e(1,\boldsymbol{X})} - 1\right) h_{\underline{\beta}}(1,\boldsymbol{X}) - \left(\frac{M(0)}{e(0,\boldsymbol{X})} - 1\right) h_{\underline{\beta}}(0,\boldsymbol{X})\right] \\
&\quad \times \left\{E\left[\pi \frac{M(1)}{e(1,\boldsymbol{X})} h_{\underline{\beta}}(1,\boldsymbol{X})\boldsymbol{Z}(1)^t + (1-\pi)\frac{M(0)}{e(0,\boldsymbol{X})} h_{\underline{\beta}}(0,\boldsymbol{X})\boldsymbol{Z}(0)^t\right]\right\}^{-1} \\
\boldsymbol{c}_2 &= E[h_{\underline{\beta}}(1,\boldsymbol{X}) - h_{\underline{\beta}}(0,\boldsymbol{X})]\left\{E\left[\pi \frac{M(1)}{e(1,\boldsymbol{X})} h_{\underline{\beta}}(1,\boldsymbol{X})\boldsymbol{Z}(1)^t + (1-\pi)\frac{M(0)}{e(0,\boldsymbol{X})} h_{\underline{\beta}}(0,\boldsymbol{X})\boldsymbol{Z}(0)^t\right]\right\}^{-1} \\
&\quad \times E\left[\pi \frac{M(1)e_{\underline{\alpha}}(1,\boldsymbol{X})}{e(1,\boldsymbol{X})^2}(Y(1) - h(1,\boldsymbol{X}))\boldsymbol{Z}(1)^t + (1-\pi)\frac{M(0)e_{\underline{\alpha}}(0,\boldsymbol{X})}{e(0,\boldsymbol{X})^2}(Y(0) - h(0,\boldsymbol{X}))\boldsymbol{Z}(0)^t\right] \\
&\quad \times \{E[\pi e_{\underline{\alpha}}(1,\boldsymbol{X})\boldsymbol{Z}(1)^t + (1-\pi)e_{\underline{\alpha}}(0,\boldsymbol{X})\boldsymbol{Z}(0)^t]\}^{-1} \\
h_{\underline{\beta}}(a,\boldsymbol{X}) &= \frac{\partial}{\partial\beta} g(\beta_0 + \beta_A a + \boldsymbol{\beta}_{\boldsymbol{X}}^t \boldsymbol{X})\Big|_{\beta=\underline{\beta}} \\
e_{\underline{\alpha}}(a,\boldsymbol{X}) &= \frac{\partial}{\partial\alpha} expit(\alpha_0 + \alpha_A a + \boldsymbol{\alpha}_{\boldsymbol{X}}^t \boldsymbol{X})\Big|_{\alpha=\underline{\alpha}}.
\end{aligned}
$$

We define

$$
\tilde{V}_{DR-WLS} = \frac{1}{\pi} Var(f_1(\boldsymbol{X}, Y(1)M(1), M(1)) + \frac{1}{1-\pi} Var(f_0(\boldsymbol{X}, Y(0)M(0), M(0)). \qquad \text{(C.1)}
$$

Lemma C.2 implies $V_{DR-WLS} = \sigma_1^2 + \sigma_2^2$ with

$$
\begin{aligned}
\sigma_1^2 &= \frac{1}{\pi} E[Var(f_1(\boldsymbol{X}, Y(1)M(1), M(1))|S)] + \frac{1}{1-\pi} E[Var(f_0(\boldsymbol{X}, Y(0)M(0), M(0))|S)] \\
\sigma_2^2 &= Var\{E[f_1(\boldsymbol{X}, Y(1)M(1), M(1)) - f_0(\boldsymbol{X}, Y(0)M(0), M(0))|S]\}.
\end{aligned}
$$

The desired variance formula then comes from the following derivation.

$$
\begin{aligned}
V_{DR-WLS} &= \sigma_1^2 + \sigma_2^2 \\
&= \frac{1}{\pi} Var(f_1(\boldsymbol{X}, Y(1)M(1), M(1))) - \frac{1}{\pi} Var(E[f_1(\boldsymbol{X}, Y(1)M(1), M(1))|S]) \\
&\quad + \frac{1}{1-\pi} Var(f_0(\boldsymbol{X}, Y(0)M(0), M(0))) - \frac{1}{1-\pi} Var(E[f_0(\boldsymbol{X}, Y(0)M(0), M(0))|S]) + \sigma_2^2 \\
&= \tilde{V}_{DR-WLS} - \pi(1-\pi) Var\left(\frac{1}{\pi} E[f_1(\boldsymbol{X}, Y(1)M(1), M(1))|S] + \frac{1}{1-\pi} E[f_0(\boldsymbol{X}, Y(0)M(0), M(0))|S]\right) \\
&= \tilde{V}_{DR-WLS} - \pi(1-\pi) Var\left(\frac{1}{\pi} E[M(1)\frac{Y(1) - h(1,\boldsymbol{X})}{e(1,\boldsymbol{X})}|S] + \frac{1}{1-\pi} E[M(0)\frac{Y(0) - h(0,\boldsymbol{X})}{e(0,\boldsymbol{X})}|S]\right) \\
&= \tilde{V}_{DR-WLS} - \frac{(1-2\pi)^2}{\pi(1-\pi)} Var\left(E[M(1)\frac{Y(1) - h(1,\boldsymbol{X})}{e(1,\boldsymbol{X})}|S] - E[M(0)\frac{Y(0) - h(0,\boldsymbol{X})}{e(0,\boldsymbol{X})}|S]\right).
\end{aligned}
$$

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

The third equation comes from the fact that, if the outcome regression model is correctly specified, then $c_2 = 0$ and $E[Z(a)M(a)\frac{Y(a)-h(a,\boldsymbol{X})}{e(a,\boldsymbol{X})}|S] = 0$; if the missing model is correctly specified, then $c_1 = 0$ and $E[Z(a)(M(a) - e(a,\boldsymbol{X}))|S] = 0$. Hence, if at least one of the two models is correct, then

$$E[f_1(\boldsymbol{X}, Y(1)M(1), M(1))|S] = E\left[M(1)\frac{Y(1)-h(1,\boldsymbol{X})}{e(1,\boldsymbol{X})} + \pi h(1,\boldsymbol{X}) - \pi h(0,\boldsymbol{X})\middle|S\right]$$
$$E[f_0(\boldsymbol{X}, Y(0)M(0), M(0))|S] = E\left[M(0)\frac{Y(0)-h(0,\boldsymbol{X})}{e(0,\boldsymbol{X})} - (1-\pi)h(1,\boldsymbol{X}) + (1-\pi)h(0,\boldsymbol{X})\middle|S\right].$$

The last equation above comes from the fact that

$$\pi E[M(1)\frac{Y(1)-h(1,\boldsymbol{X})}{e(1,\boldsymbol{X})}|S] + (1-\pi)E[M(0)\frac{Y(0)-h(0,\boldsymbol{X})}{e(0,\boldsymbol{X})}|S] = 0,$$

which is implied by the definition of $\underline{\boldsymbol{\theta}}$.

At last, we prove $\widehat{Var}(\widehat{\Delta}_{DR-WLS}) \xrightarrow{P} \tilde{V}_{DR-WLS}$. The sandwich variance estimator defined in Section 3 can be rewritten into

$$\widehat{Var}(\widehat{\Delta}_{DR-WLS}) = \frac{1}{n}\sum_{i=1}^{n}\left\{\frac{A_i}{\pi}\hat{f}_1(\boldsymbol{X}_i, Y_iM_i, M_i) - \frac{(1-A_i)}{(1-\pi)}\hat{f}_0(\boldsymbol{X}_i, Y_iM_i, M_i) - \widehat{\Delta}_{DR-WLS}\right\}^2,$$

where $\hat{f}_1, \hat{f}_0$ are finite sample estimates of f_1, f_0 respectively, by substituting $\hat{\beta}, \hat{\alpha}$ for $\underline{\beta}, \underline{\alpha}$. Using the same technique as in proof of Theorem 4.1 for consistent estimation of asymptotic variance, we have

$$\widehat{Var}(\widehat{\Delta}_{DR-WLS}) = \frac{1}{n}\sum_{i=1}^{n}\left\{\frac{A_i}{\pi}f_1(\boldsymbol{X}_i, Y_i(1)M_i(1), M_i(1)) - \frac{(1-A_i)}{(1-\pi)}f_0(\boldsymbol{X}_i, Y_i(0)M_i(0), M_i(0)) - \Delta^*\right\}^2 + o_p(1).$$

We apply Lemma C.1 and get

$$\begin{aligned}\widehat{Var}(\widehat{\Delta}_{DR-WLS}) &= \frac{1}{n}\sum_{i=1}^{n}\frac{A_i}{\pi^2}\left\{f_1(\boldsymbol{X}_i, Y_i(1)M_i(1), M_i(1)) - \pi\Delta^*\right\}^2 \\ &\quad + \frac{1}{n}\sum_{i=1}^{n}\frac{1-A_i}{(1-\pi)^2}\left\{f_0(\boldsymbol{X}_i, Y_i(0)M_i(0), M_i(0)) - (1-\pi)\Delta^*\right\}^2 + o_p(1) \\ &= \frac{1}{\pi}Var(f_1(\boldsymbol{X}, Y(1)M(1), M(1)) + \frac{1}{1-\pi}Var(f_0(\boldsymbol{X}, Y(0)M(0), M(0)) + o_p(1) \\ &= \tilde{V}_{DR-WLS} + o_p(1).\end{aligned}$$

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive

□

Proof of Corollary 1. Given the asymptotic variance formula of Corollary 2, it is easy to verify that when $\pi = 0.5$ or $E[Y(a)|\boldsymbol{X}] = g(\beta_0 + \beta_A a + \boldsymbol{\beta}_{\boldsymbol{X}}^t \boldsymbol{X})$, we have

$$\frac{(1-2\pi)^2}{\pi(1-\pi)} Var\left\{E\left[M(1)\frac{Y(1)-h(1,\boldsymbol{X})}{e(1,\boldsymbol{X})} - M(0)\frac{Y(0)-h(0,\boldsymbol{X})}{e(0,\boldsymbol{X})}\middle| S\right]\right\} = 0$$

and hence $V_{DR-WLS} = \tilde{V}_{DR-WLS}$. Since the sandwich variance estimator is consistent to $\tilde{V}_{DR-WLS}$ we get the desired result for these two cases.

For the last case where the model includes a treatment-by-strata term, we define $\tilde{h}(a, \boldsymbol{X}) = g(\beta_0 + \beta_A a + \boldsymbol{\beta}_{\boldsymbol{X}}^t X + \beta_{AS}^t a \times S)$ and $\tilde{f}_a(\boldsymbol{X}, Y(a)M(a), M(a))$ by substituting $\tilde{h}(a, \boldsymbol{X})$ for $h(a, \boldsymbol{X})$ in $f_a(\boldsymbol{X}, Y(a)M(a), M(a))$. Following the proof of Corollary 2, if the model includes a treatment-by-strata interaction term, under stratified randomization or the biased-coin design, we have

$$\sqrt{n}(\widehat{\Delta}_{DR-WLS} - \Delta^*) \xrightarrow{d} N(0, \overline{V}_{DR-WLS}),$$

where

$$\begin{aligned}\overline{V}_{DR-WLS} = &\frac{1}{\pi} Var(\tilde{f}_1(\boldsymbol{X}, Y(1)M(1), M(1))) + \frac{1}{1-\pi} Var(\tilde{f}_0(\boldsymbol{X}, Y(0)M(0), M(0))) \\ &- \frac{(1-2\pi)^2}{\pi(1-\pi)} Var\{E[M(1)\frac{Y(1)-\tilde{h}(1,\boldsymbol{X})}{e(1,\boldsymbol{X})} - M(0)\frac{Y(0)-\tilde{h}(0,\boldsymbol{X})}{e(0,\boldsymbol{X})}|S]\}.\end{aligned}$$

According to the expected estimating equation

$$\begin{aligned}\pi E[M(1)\frac{Y(1)-\tilde{h}(1,\boldsymbol{X})}{e(1,\boldsymbol{X})}|S] + (1-\pi)E[M(0)\frac{Y(0)-\tilde{h}(0,\boldsymbol{X})}{e(0,\boldsymbol{X})}|S] &= 0, \\ E[M(1)\frac{Y(1)-\tilde{h}(1,\boldsymbol{X})}{e(1,\boldsymbol{X})}|S] &= 0,\end{aligned}$$

we have $E[M(1)\frac{Y(1)-\tilde{h}(1,\boldsymbol{X})}{e(1,\boldsymbol{X})} - M(0)\frac{Y(0)-\tilde{h}(0,\boldsymbol{X})}{e(0,\boldsymbol{X})}|S] = 0$ and hence

$$\overline{V}_{DR-WLS} = \frac{1}{\pi} Var(\tilde{f}_1(\boldsymbol{X}, Y(1)M(1), M(1))) + \frac{1}{1-\pi} Var(\tilde{f}_0(\boldsymbol{X}, Y(0)M(0), M(0))).$$

Then following the proof of Corollary 2, we have the sandwich variance estimator is consistent to $\overline{V}_{DR-WLS}$. □

COBRA
A BEPRESS REPOSITORY
Collection of Biostatistics
Research Archive