Statistical Inference of Covariate-Adjusted Randomized Experiments

Feifang Hu

Department of Statistics

George Washington University

Joint research with Wei Ma, Yichen Qin and Yang Li

Email: feifang@gwu.edu

Nov 8, 2018 at IMA,

Outline

- Introduction
- Framework
- General Properties
- Implementation and Correction
- Numerical Studies
- Conclusion

1 Introduction

Covariate-adjusted randomization is frequently used because it utilizes the covariate information to form more balanced treatment groups.

- Balance categorical covariates: Pocock and Simon's minimization method and its extensions (Taves,1974; Pocock and Simon 1975; Hu and Hu 2012)
- Balance continuous covariates based on distribution characteristics, e.g., mean and variance (Frane 1998), quartiles (Su 2011), density function (Ma and Hu 2013).
- Balance continuous covariates based on models (Atkinson 1982, Smith 1984ab)
- Balance covariates available prior to the experiment onset (Morgan and Rubin, 2012, 2015, Qin et al. 2017)

Since covariate-adjusted randomizations inevitably use the covariate information in forming more balanced treatment groups, the subsequent statistical inference is usually affected and demonstrates undesirable properties, such as reduced type I errors and powers.

This phenomenon of conservativeness is particularly common for a working model including only a subset of covariates used in randomization, such as two sample t test.

It is ideal that the covariates used in randomization should be included in the subsequent analysis to achieve valid test.

However, unadjusted tests still dominate in practice (Sverdlov, 2015).

- Investigation sites
- Simplicity of the test procedure
- Robustness to model misspecification

As covariates are commonly used in comparative studies (biomarker analysis, precision medicine and crowdsourced-internet experimentation), understanding the impact of covariate-adjusted randomization on statistical inference is an increasingly pressing problem.

Existing work

- Birkett (1985), Forsythe (1987), etc.. mainly based on simulations.
- Shao et al. (2010) shows t-test is conservative for stratified biased coin design.
- Ma et al. (2015) studied tests under a linear model for discrete covariate-adjusted randomization by assuming that overall and marginal imbalances are bounded in probability.

Limitations

- Not applicable to randomizations directly balancing continuous covariates, e.g., Atkinson's D_A-Biased Coin Design.
- The assumed balancing properties are too strong, i.e., $O_p(1)$ marginal imbalances.
- Do not consider the scenario when covariate information are avialable before the experiment starts, e.g., Rerandomization, Pairwise Sequential Randomization.

Motivations

- Derive the statistical properties of inference under general covariate-adjusted randomization methods.
- Explicitly display the relationship between covariate balance and inference, and explain why inference behaves differently for various randomization methods.
- Obtain the results that have broad applications, including RR, PSR, and D_A-BCD, and compare these methods analytically.
- Propose a method to attain valid and powerful tests.

2 Framework

Suppose that n units are to be assigned to two treatment groups. T_i denotes the assignment of the i-th unit, i.e., $T_i = 1$ for treatment 1 and $T_i = 0$ for treatment 2. Let $\boldsymbol{x}_i = (x_{i,1}, ..., x_{i,p+q})^t$ represent $p+q$ covariates observed for the i-th unit, where $x_{i,j} \overset{iid}{\sim} X_j$ for $i = 1, ..., n$.

The underlying model:

$$Y_i = \mu_1 T_i + \mu_2(1 - T_i) + \sum_{j=1}^{p+q} \beta_j x_{i,j} + \epsilon_i,$$

where $\mu_1 - \mu_2$ is the treatment effect, $\boldsymbol{\beta} = (\beta_1, ..., \beta_{p+q})^t$ is the covariate effects, and ϵ_i is iid random error with mean zero and variance σ_ϵ^2, and is independent of covariates. Covariates are assumed independent of each other with $\mathbb{E}X_j = 0$ for $j = 1, ..., p+q$.

After allocating the units to treatment groups via covariate-adjusted randomization, a working model is used to estimate and test the treatment effect.

In such a working model, it is common in practice to include a subset of covariates used in randomization, or sometimes even no covariates at all (Shao et al. 2010, Ma et al. 2015, Sverdlov 2015).

The working model:

$$\mathbb{E}[Y_i] = \mu_1 T_i + \mu_2(1 - T_i) + \sum_{j=1}^{p} \beta_j x_{i,j}.$$

.

Let $\boldsymbol{Y} = (Y_1, ..., Y_n)^t$, $\boldsymbol{T} = (T_1, ..., T_n)^t$, $\mathbf{X} = [\mathbf{X}_{\text{in}}; \mathbf{X}_{\text{ex}}]$, where

$$\mathbf{X}_{\text{in}} = \begin{bmatrix} x_{1,1} & \cdots & x_{1,p} \\ \vdots & \ddots & \vdots \\ x_{n,1} & \cdots & x_{n,p} \end{bmatrix}, \ \mathbf{X}_{\text{ex}} = \begin{bmatrix} x_{1,p+1} & \cdots & x_{1,p+q} \\ \vdots & \ddots & \vdots \\ x_{n,p+1} & \cdots & x_{n,p+q} \end{bmatrix}.$$

Further let $\boldsymbol{\beta}_{\text{in}} = (\beta_1, ..., \beta_p)^t$, $\boldsymbol{\beta}_{\text{ex}} = (\beta_{p+1}, ..., \beta_{p+q})^t$, so that $\boldsymbol{\beta} = (\boldsymbol{\beta}_{\text{in}}^t, \boldsymbol{\beta}_{\text{ex}}^t)^t$. Then the working model can also be written as,

$$\mathbb{E}[\boldsymbol{Y}] = \mathbf{G}\boldsymbol{\theta},$$

where $\mathbf{G} = [\boldsymbol{T}; \mathbf{1}_n - \boldsymbol{T}; \mathbf{X}_{\text{in}}]$ is the design matrix, $\boldsymbol{\theta} = (\mu_1, \mu_2, \boldsymbol{\beta}_{\text{in}}^t)^t$ is the vector of parameters of interest, and $\mathbf{1}_n$ is the n-dimensional vector of ones.

The ordinary least squares (OLS) estimate of $\boldsymbol{\theta}$, $\hat{\boldsymbol{\theta}} = (\hat{\mu}_1, \hat{\mu}_2, \hat{\boldsymbol{\beta}}_{\text{in}}^t)^t$ is,

$$\hat{\boldsymbol{\theta}} = (\mathbf{G}^t\mathbf{G})^{-1}\mathbf{G}^t\boldsymbol{Y}.$$

Testing the treatment effect:

$$H_0 : \mu_1 - \mu_2 = 0 \text{ versus } H_1 : \mu_1 - \mu_2 \neq 0,$$

and the test statistic is

$$S = \frac{\boldsymbol{L}^t \hat{\boldsymbol{\theta}}}{\sqrt{\hat{\sigma}_w^2 \boldsymbol{L}^t (\mathbf{G}^t \mathbf{G})^{-1} \boldsymbol{L}}},$$

where $\boldsymbol{L} = (1, -1, 0, ..., 0)^t$ is a vector of length $p + 2$, and $\hat{\sigma}_w^2 = \|\boldsymbol{Y} - \mathbf{G}\hat{\boldsymbol{\theta}}\|^2/(n - p - 2)$ is the model-based estimate of the error variance $\sigma_w^2 = \sigma_\epsilon^2 + \sum_{j=1}^q \beta_{p+j}^2 \text{Var}(X_{p+j})$.

The traditional testing procedure is to reject the null hypothesis at the significance level α if $|S| > z_{1-\alpha/2}$, where $z_{1-\alpha/2}$ is $(1 - \alpha/2)$-th quantile of a standard normal distribution.

Testing the covariate effects:

Let $\mathbf{C}$ be an $m \times (p+2)$ matrix of rank m $(m \leq p)$ with entries in the first two columns all equal to zero (no treatment effect to test).

$$H_0 : \mathbf{C}\boldsymbol{\theta} = \boldsymbol{c}_0 \text{ versus } H_1 : \mathbf{C}\boldsymbol{\theta} = \boldsymbol{c}_1, \tag{1}$$

and the test statistic is,

$$S^* = \frac{(\mathbf{C}\hat{\boldsymbol{\theta}} - \boldsymbol{c}_0)^t [\mathbf{C}(\mathbf{G}^t\mathbf{G})^{-1}\mathbf{C}^t]^{-1}(\mathbf{C}\hat{\boldsymbol{\theta}} - \boldsymbol{c}_0)}{m\hat{\sigma}_w^2}$$

The traditional testing procedure is to reject the null hypothesis at the significance level α if $|S| > z_{1-\alpha/2}$, where $z_{1-\alpha/2}$ is $(1-\alpha/2)$-th quantile of a standard normal distribution.

3 General Properties

Assumption 1 *Global balance:* $n^{-1}\sum_{i=1}^{n}(2T_i - 1) \xrightarrow{p} 0$.

Assumption 2 *Covariate balance:* $n^{-1/2}\sum_{i=1}^{n}(2T_i - 1)\,\boldsymbol{x}_i \xrightarrow{d} \boldsymbol{\xi}$, *where* $\boldsymbol{\xi}$ *is a (p+q)-dimensional random vector with* $\mathbb{E}[\boldsymbol{\xi}] = \mathbf{0}$.

Consistency:

Theorem 3.1 *Given Assumptions 1 and 2, we have* $\hat{\boldsymbol{\theta}} \xrightarrow{p} \boldsymbol{\theta}$.

Testing the treatment effect:

We partition $\boldsymbol{\xi} = (\boldsymbol{\xi}_{\text{in}}^t, \boldsymbol{\xi}_{\text{ex}}^t)^t$ so that $\boldsymbol{\xi}_{\text{in}}$ represents the first p dimensions of $\boldsymbol{\xi}$, and $\boldsymbol{\xi}_{\text{ex}}$ the last q dimensions. Further let $\lambda_1 = \sigma_\epsilon/\sigma_w$, $\lambda_2 = 1/\sigma_w$, and Z be a standard normal random variable that is independent of $\boldsymbol{\xi}_{\text{ex}}$.

Theorem 3.2 *Given Assumptions 1 and 2, we have*

1. *Under* $H_0 : \mu_1 - \mu_2 = 0$, *then*

$$S \xrightarrow{d} \lambda_1 Z + \lambda_2 \boldsymbol{\beta}_{\text{ex}}^t \boldsymbol{\xi}_{\text{ex}}.$$

2. *Under* $H_1 : \mu_1 - \mu_2 \neq 0$, *consider a sequence of local alternatives with* $\mu_1 - \mu_2 = \delta/\sqrt{n}$ *for a fixed* $\delta \neq 0$, *then*

$$S \xrightarrow{d} \lambda_1 Z + \lambda_2 \boldsymbol{\beta}_{\text{ex}}^t \boldsymbol{\xi}_{\text{ex}} + \frac{1}{2}\lambda_2 \delta.$$

The asymptotic distribution of test statistic S under H_0 consists of two independent components, $\lambda_1 Z$ and $\lambda_2 \boldsymbol{\beta}_{\text{ex}}^t \boldsymbol{\xi}_{\text{ex}}$.

The first component is due to the random error ϵ_i in the underlying model, and remains invariant under different covariate-adjusted randomization.

The second component of S represents the impact of a covariate-adjusted randomization on the test statistic through the level of covariate balance.

Under covariate-adjusted randomization, $\boldsymbol{\xi}$ is more concentrated around 0 as opposed to complete randomization, leading to conservative tests.

Testing the covariate effects:

Theorem 3.3 *Given Assumptions 1 and 2, we have*

1. *Under* $H_0 : \mathbf{C}\boldsymbol{\theta} = \boldsymbol{c}_0$, *then*

$$S^* \xrightarrow{d} \chi^2_m / m.$$

2. *Under* $H_1 : \mathbf{C}\boldsymbol{\theta} = \boldsymbol{c}_1$, *consider a sequence of local alternatives with* $\boldsymbol{c}_1 - \boldsymbol{c}_0 = \boldsymbol{\Delta}/\sqrt{n}$ *for a fixed* $\boldsymbol{\Delta} \neq \mathbf{0}$, *then*

$$S^* \xrightarrow{d} \chi^2_m(\phi)/m, \ \phi = \boldsymbol{\Delta}^t[\mathbf{C}\mathbf{V}^{-1}\mathbf{C}^t]^{-1}\boldsymbol{\Delta}/\sigma^2_w.$$

where ϕ *is the non-central parameter, and* $\mathbf{V} = \text{diag}\,(1/2, 1/2, \text{Var}(X_1), ..., \text{Var}(X_p))$.

The type I error is maintained when testing the covariate effects under covariate-adjusted randomization.

The power, however, is reduced if not all covariate information is incorporated in the working model.

4 Implementation and Correction

4.1 Examples

Complete Randomization

Rerandomization (Morgan and Rubin, 2012, 2015)

- Repeat the traditional randomization process until a satisfactory configuration is achieved.

Pairwise Sequential Randomization (Qin et al, 2017)

- An alternative that achieves the optimal covariate balance and is computationally more efficient.

Atkinson's D_A-Biased Coin Design (Atkinson 1982, Smith 1984ab)

- Represent a large class of methods that take covariates into account in allocation rules based on certain optimality criteria.

Rerandomization

(1) Collect covariate data.

(2) Specify a balance criterion to determine when a randomization is acceptable. For example, the criterion could be defined as a threshold of $a > 0$ on some user-defined imbalance measure, denoted as M.

(3) Randomize the units into treatment groups using traditional randomization methods, such as CR.

(4) Check the balance criterion $M < a$. If the criterion is satisfied, go to Step (5); otherwise, return to Step (3).

(5) Perform the experiment using the final randomization obtained in Step (4).

Pairwise Sequential Randomization

(1) Collect covariate data.

(2) Choose the covariate imbalance measure for n units, denoted as $M(n)$.

(3) Randomly arrange all n units in a sequence $\boldsymbol{x}_1$, ... , $\boldsymbol{x}_n$.

(4) Separately assign the first two units to treatment 1 and treatment 2.

(5) Suppose that $2i$ units have been assigned to treatment groups ($i \geq 1$), for the $(2i+1)$-th and $(2i+2)$-th units:

(5a) If the $(2i+1)$-th unit is assigned to treatment 1 and the $(2i+2)$-th unit is assigned to treatment 2 (i.e., $T_{2i+1} = 1$ and $T_{2i+2} = 0$), then we can calculate the "potential" imbalance measure, $M_i^{(1)}$, between the updated treatment groups with $2i+2$ units.

(5b) Similarly, if the $(2i+1)$-th unit is assigned to treatment 2 and the $(2i+2)$-th unit is assigned to treatment 1 (i.e., $T_{2i+1} = 0$ and $T_{2i+2} = 1$), then we can calculate the "potential" imbalance measure, $M_i^{(2)}$, between the updated treatment groups with $2i+2$ units.

(6) Assign the $(2i+1)$-th and $(2i+2)$-th units to treatment groups according to the following probabilities:

$$\mathbb{P}(T_{2i+1} = 1 | \boldsymbol{x}_{2i}, ..., \boldsymbol{x}_1, T_{2i}, ..., T_1) = \begin{cases} \rho & \text{if } M_i^{(1)} < M_i^{(2)} \\ 1-\rho & \text{if } M_i^{(1)} > M_i^{(2)} \\ 0.5 & \text{if } M_i^{(1)} = M_i^{(2)} \end{cases},$$

where $0.5 < \rho < 1$, and assign $T_{2i+2} = 1 - T_{2i+1}$ to maintain the equal proportions.

(7) Repeat Steps (5) through (7) until all units are assigned.

Atkinson's D_A-Biased Coin Design

Suppose n units have been assigned to treatment groups, D_A-BCD assigns the $(n+1)$-th unit to treatment 1 with probability

$$\mathbb{P}(T_{n+1} = 1 | \boldsymbol{x}_{n+1}, ..., \boldsymbol{x}_1, T_n, ..., T_1)$$
$$= \frac{[1 - (1; \boldsymbol{x}_{n+1}^t)(\mathbf{F}_n^t \mathbf{F}_n)^{-1} \boldsymbol{b}_n]^2}{[1 - (1; \boldsymbol{x}_{n+1}^t)(\mathbf{F}_n^t \mathbf{F}_n)^{-1} \boldsymbol{b}_n]^2 + [1 + (1; \boldsymbol{x}_{n+1}^t)(\mathbf{F}_n^t \mathbf{F}_n)^{-1} \boldsymbol{b}_n]^2}.$$

where $\mathbf{F}_n = [\mathbf{1}_n; \mathbf{X}]$ and $\boldsymbol{b}_n^t = (2\boldsymbol{T} - \mathbf{1}_n)^t \mathbf{F}_n$.

Complete Randomization

$$\boldsymbol{\xi}^{\text{CR}} \sim N(\mathbf{0}, \boldsymbol{\Sigma})$$

Rerandomization

$$\boldsymbol{\xi}^{\text{RR}} \sim \boldsymbol{\Sigma}^{1/2}\boldsymbol{D} \;\Big|\; \boldsymbol{D}^t\boldsymbol{D} < a$$

Pairwise Sequential Randomization

$$\boldsymbol{\xi}^{\text{PSR}} = \boldsymbol{O}_p\left(\frac{1}{\sqrt{n}}\right)$$

Atkinson's D_A-Biased Coin Design

$$\boldsymbol{\xi}^{\text{D-BCD}} \sim N(\mathbf{0}, \frac{1}{5}\boldsymbol{\Sigma})$$

where $\boldsymbol{\Sigma} = \text{diag}(\text{Var}(X_1), ..., \text{Var}(X_{p+q}))$, $\boldsymbol{D} \sim N(\mathbf{0}, \mathbf{I}_{p+q})$ *and* $\mathbf{I}_{p+q}$ *is the* $(p+q)$*-dim identity matrix.*

Testing the Treatment Effect under Atkinson's D_A-Biased Coin Design

Theorem 4.1 *Under D_A-BCD, we have*

1. *Under $H_0 : \mu_1 - \mu_2 = 0$, then*

$$S \xrightarrow{d} N\left(0, \frac{\sigma_\epsilon^2 + \frac{1}{5}\sum_{j=1}^{q}\beta_{p+j}^2 \mathsf{Var}(X_{p+j})}{\sigma_\epsilon^2 + \sum_{j=1}^{q}\beta_{p+j}^2 \mathsf{Var}(X_{p+j})}\right).$$

2. *Under $H_1 : \mu_1 - \mu_2 \neq 0$, where $\mu_1 - \mu_2 = \delta/\sqrt{n}$ for a fixed $\delta \neq 0$,*

$$S \xrightarrow{d} N\left(\frac{1}{2}\lambda_2\delta, \frac{\sigma_\epsilon^2 + \frac{1}{5}\sum_{j=1}^{q}\beta_{p+j}^2 \mathsf{Var}(X_{p+j})}{\sigma_\epsilon^2 + \sum_{j=1}^{q}\beta_{p+j}^2 \mathsf{Var}(X_{p+j})}\right).$$

Testing the Treatment Effect under Pairwise Sequential Randomization

Theorem 4.2 *Under PSR, we have*

1. *Under* $H_0 : \mu_1 - \mu_2 = 0$*, then*

$$S \xrightarrow{d} N\left(0, \frac{\sigma_\epsilon^2}{\sigma_\epsilon^2 + \sum_{j=1}^{q} \beta_{p+j}^2 \mathsf{Var}(X_{p+j})}\right).$$

2. *Under* $H_1 : \mu_1 - \mu_2 \neq 0$*, where* $\mu_1 - \mu_2 = \delta/\sqrt{n}$ *for a fixed* $\delta \neq 0$*,*

$$S \xrightarrow{d} N\left(\frac{1}{2}\lambda_2\delta, \frac{\sigma_\epsilon^2}{\sigma_\epsilon^2 + \sum_{j=1}^{q} \beta_{p+j}^2 \mathsf{Var}(X_{p+j})}\right).$$

The variance from the covariates is completely eliminated out in the numerator of the asymptotic distribution of S, resulting in a distribution more concentrated around 0 than the standard normal distribution.

This can be considered as an extension of the results in Ma et al. (2015) that studied conservative tests for covariate-adaptive designs balancing discrete covariates.

4.2 Correction for Conservativeness

To correct conservativeness, we need to obtain the correct asymptotic critical values for valid tests.

- Based on the asymptotic distribution of S in Theorem 3.2. Need to estimate the unknown parameters.
- Or use Bootstrap method to do the correction. Computationally intensive.

Randomization	Covariate balance	Type I error of traditional test	Power of corrected test
CR	least balanced	valid	least powerful
RR	moderately balanced	moderately conservative	moderately powerful
D_A-BCD	moderately balanced	moderately conservative	moderately powerful
PSR	most balanced	most conservative	most powerful

Table 1: Comparison of different covariate-adjusted randomization procedures in terms of covariate balance, traditional tests' conservativeness, and corrected tests' powers.

5 Numerical Studies

Verification of Theoretical Results

Underlying model:

$$Y_i = \mu_1 T_i + \mu_2(1 - T_i) + \sum_{j=1}^{4} \beta_j x_{i,j} + \epsilon_i,$$

where $\mu_1 = \mu_2 = 0$, $\beta_j = 1$ for $j = 1, ..., 4$. $x_{i,j} \sim N(0, 1)$ for $j = 1, ..., 4$ and is independent of each other. The random error $\epsilon_i \sim N(0, 2^2)$ is independent of all $x_{i,j}$.

Working model:

$$\mathbb{E}[Y_i] = \mu_1 T_i + \mu_2(1 - T_i) + \beta_1 x_{i,1} + \beta_2 x_{i,2}$$

.

Verification of Theoretical Results

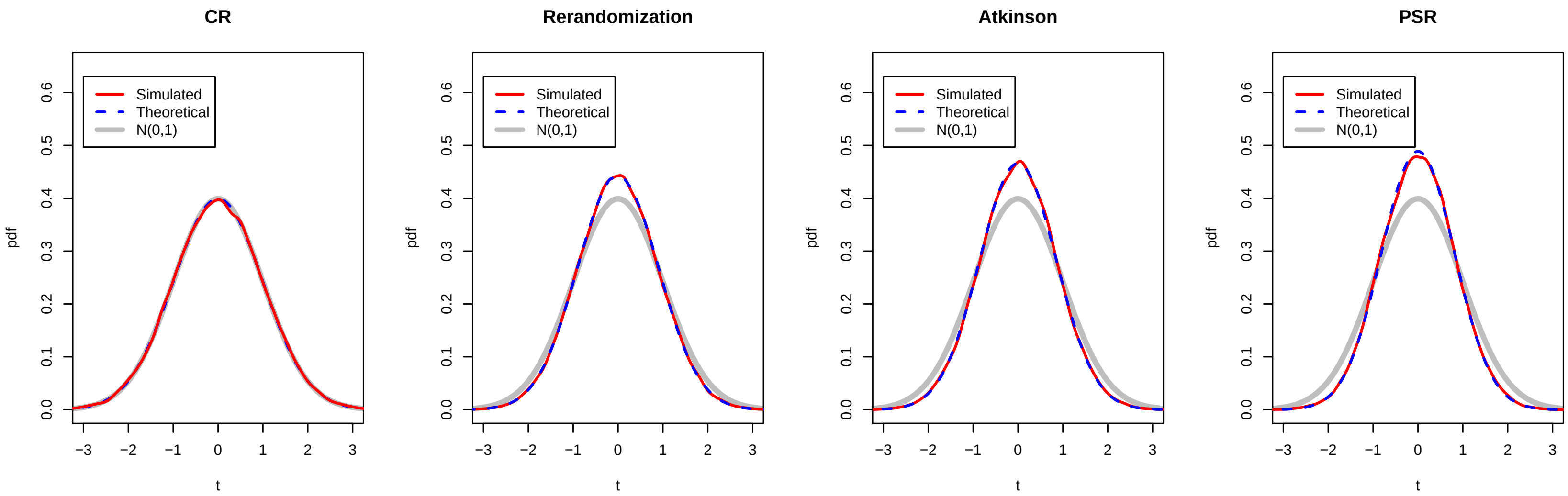

Figure 1: Comparison of theoretical distributions and simulated distributions of S. In each panel, red solid curve represents the simulated distribution, blue dash curve represents the theoretical distribution, and the gray bold curve is the standard normal density.

Conservative Hypothesis Testing for Treatment Effect

Underlying model:

$$Y_i = \mu_1 T_i + \mu_2(1 - T_i) + \sum_{j=1}^{6} \beta_j x_{i,j} + \epsilon_i, \qquad (2)$$

where $\beta_j = 1$ for $j = 1, ...6$. $x_{i,j} \sim N(0, 1)$ and is independent of each other. The random error $\epsilon_i \sim N(0, 2^2)$ is independent of all $x_{i,j}$.

Working model:

W1: $\mathbb{E}[Y_i] = \mu_1 T_i + \mu_2(1 - T_i)$.

W2: $\mathbb{E}[Y_i] = \mu_1 T_i + \mu_2(1 - T_i) + \sum_{j=1}^{2} \beta_j x_{i,j}$.

W3: $\mathbb{E}[Y_i] = \mu_1 T_i + \mu_2(1 - T_i) + \sum_{j=3}^{6} \beta_j x_{i,j}$.

W4: $\mathbb{E}[Y_i] = \mu_1 T_i + \mu_2(1 - T_i) + \sum_{j=1}^{6} \beta_j x_{i,j}$.

Conservative Hypothesis Testing for Treatment Effect: Type I error

Randomization	W1	W2	W3	W4
CR	0.0529	0.0512	0.0538	0.0513
RR	0.0114	0.0166	0.0259	0.0502
D_A-BCD	0.0071	0.0118	0.0249	0.0532
PSR	0.0018	0.0058	0.0178	0.0519

Table 2: Type I error of traditional tests for treatment effect using different working models and different randomization procedures.

Corrected Hypothesis Testing for Treatment Effect: Type I error

Randomization	W1	W2	W3	W4
CR	0.0477	0.0495	0.0459	0.0451
RR	0.0514	0.0498	0.0515	0.0510
D_A-BCD	0.0508	0.0518	0.0525	0.0511
PSR	0.0597	0.0584	0.0504	0.0477

Table 3: Type I error of hypothesis testing for treatment effect using estimated asymptotic distribution's critical values under different working models and different randomization procedures.

Corrected Hypothesis Testing for Treatment Effect: Power

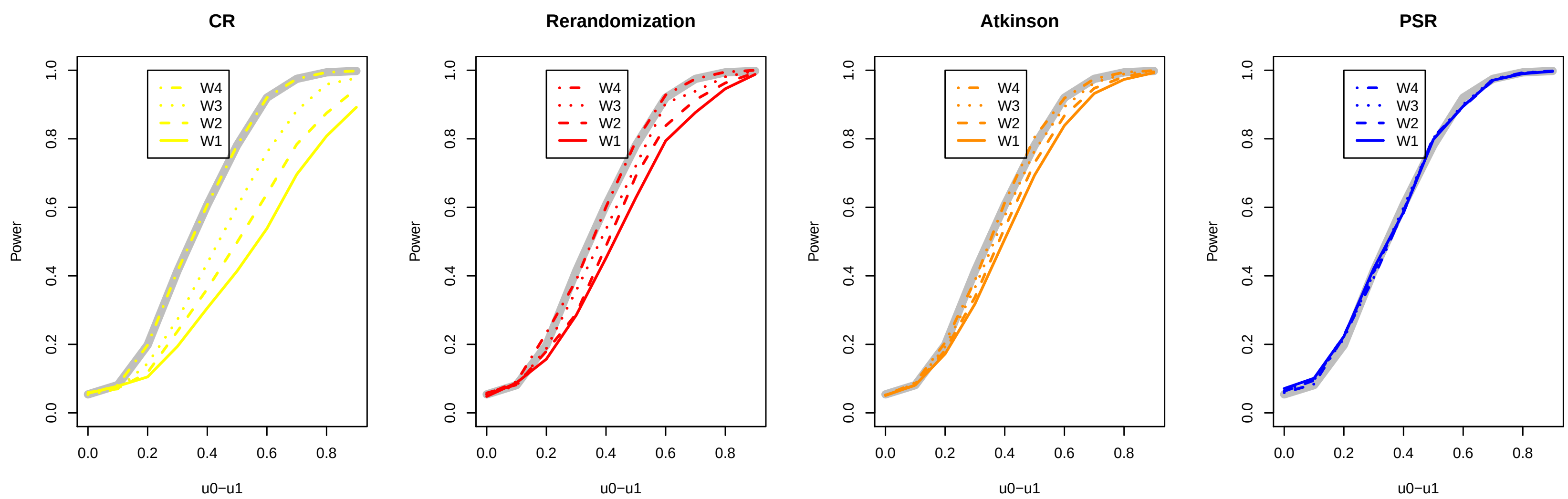

Figure 2: Power against $\mu_1 - \mu_2$ using estimated asymptotic distribution's critical values and p-values. Sample size $n = 500$. Note that we plot the power of W4 under CR in bold gray curves in all the panels for a better comparison among different randomizations.

6 Conclusion

- Derive inference properties under general covariate-adjusted randomization.
- Explicitly unveil the relationship between covariate-adjusted and inference properties.
- Apply the general theory to several important randomization methods.
- A correction approach is proposed to attain valid and powerful test.

Thank you!