Covariate-adaptive Randomization: New Procedures and Inference

Feifang Hu

Department of Statistics

George Washington University

Email: feifang@gwu.edu

Partially based on joint projects with

Ping Li, Yang Li, Yang Liu, Wei Ma, Yichen Qin, Fan Wang, Lixin Zhang, Yifan Zhou and Zhixin Zhou

Sept 17, 2021, USA

Outline

- Introduction
- New Covariate-Adaptive Designs
- Statistical Inference
- 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; Taves 2010; 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. 2016)

1.1 Some motivated examples

Example 1: Remdesivir-COVID-19 trial (China). Remdesivir in adults with severe COVID-19 trial (Wang *et al.* 2020) is a randomized, double-blind, placebo-controlled, multicentre trial that aimed to compare Remvesivir with placebo. There were 236 patients in the trial. There are about 20 baseline covariates for each patient, including 10 continuous variables (e.g. age and White blood cell count) and 10 discrete variables (e.g. gender and Hypertension). The stratified (according to the level of respiratory support) permuted block (30 patients per block) randomization procedure were implemented. At the end of this trial, some important imbalances existed at enrollment between the groups, including more patients with hypertension, diabetes, or coronary artery disease in the Remdesivir group than the placebo group.

Example 2. The Project GATE (Growing America Through Entrepreneurship), sponsored by the U.S. Department of Labor, was designed to evaluate the impact of offering tuition-free entrepreneurship training services (GATE services) on helping clients create, sustain or expand their own business. (https://www.doleta.gov/reports/projectgate/)

The cornerstone is complete randomization. Members of the treatment group were offered GATE services; members of the control group were not.

- $n = 4,198$ participants
- $p = 105$ covariates

Example 3. Online A/B testing. (Kohavi and Thomke, 2017, Harvard Business Review) Microsoft, Amazon, Facebook and Google conduct more than 10,000 online controlled experiments annually, with many tests engaging millions of users.

Amazon's experiment.
Treatment A: Credit card offers on front page.
Treatment B: Credit card offers on the shopping cart page.

This (change from A to B) boosted profits by tens of millions of US Dollars annually.

Often Network (Dependent and Interference) Data, How to Design these studies?

2 New covariate-adaptive designs

2.1 Balancing many covariates

Advantages of covariate balance:

- Improve accuracy and efficiency of inference.
- Remove the bias and increase the power.
- Increases the interpretability of results by making the units more comparable, enhance the credibility.
- More robust against model misspecification.
- Rubin (2008): the greatest possible efforts should be made during the design phase rather than the analysis stage.

- Randomization: an essential tool for evaluating treatment effect.
- Traditional randomization methods (e.g., complete randomization (CR)): unsatisfactory, **unbalanced** prognostic or baseline covariates.

"Most of experimenters on carrying out a random assignment of plots will be shocked to find out how far from equally the plots distribute themselves." —Fisher (1926)

What if large p and large n?

- The phenomenon of covariate imbalance is exacerbated as p and n increase.
- Ubiquitous in the era of big data.
- Example: the probability of one particular covariate being unbalanced is $\alpha = 5\%$. For a study with 10 covariates, the chance of at least one covariate exhibiting imbalance is $1 - (1 - \alpha)^p = 40\%$. With 100 covariates, the chance is $1 - (1 - \alpha)^{100} = 1$.

Morgan and Rubin (2012) proposed rerandomization.

(1) Collect covariate data.

(2) Specify a balance criterion, $M < a$, i.e., threshold on the Mahalanobis distance,

$$M = (\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2)^T [\mathrm{cov}(\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2)]^{-1} (\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2),$$

where $\bar{\boldsymbol{x}}_1$ and $\bar{\boldsymbol{x}}_2$ are the sample means for treatment groups.

(3) Randomize the units using the complete randomization (CR).

(4) Check the balance criterion, $M < a$.

- If satisfied, go to Step (5); otherwise, return to Step (3).

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

Advantages:

- Desirable properties for causal inference:
  - Reduction in variance of estimated treatment effect.
- Work well with a few covariates.

Drawbacks:

- Not for sequential experiments
- Incapable to scale up for massive data.
- As p increases, the probability of acceptance $p_a = P(M < a)$ decreases, causing the RR to remain in the loop for a long time.

2.2 CAM and Its Properties

Covariate-Adaptive Randomization via Mahalanobis Distance (CAM)

$\boldsymbol{x}_i \in \mathbb{R}^p$: covariate of the i-th unit.

$T_i \in \{1, 0\}$: treatment assignment of the i-th unit.

- $T_i = 1$: treatment 1.
- $T_i = 0$: treatment 2.

$i = 1, ..., n$

(1) Use the new defined Mahalanobis distance

$$M(n) = 0.25(\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2)^T [\mathrm{cov}(\bar{\boldsymbol{x}})]^{-1} (\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2).$$

(2) Randomly arrange units in a sequence

$$\underbrace{\boldsymbol{x}_1, \boldsymbol{x}_2}_{\textit{1st pair}}, \underbrace{\boldsymbol{x}_3, \boldsymbol{x}_4}_{\textit{2nd pair}}, \underbrace{\boldsymbol{x}_5, \boldsymbol{x}_6}_{\textit{3rd pair}}, ..., \boldsymbol{x}_n.$$

(3) Assign the 1st pair, $T_1 = 1$, $T_2 = 0$.

(4) For the next pair, i.e., $2i+1$-th and $2i+2$-th units, $(i > 1)$

(4a) If $T_{2i+1} = 1$ and $T_{2i+2} = 0$, obtain the "potential" $M_i^{(1)}$.

(4b) If $T_{2i+1} = 0$ and $T_{2i+2} = 1$, obtain the "potential" $M_i^{(2)}$.

(5) Assign the $(2i+1)$-th and $(2i+2)$-th units by

$$P(T_{2i+1}=1, T_{2i+2}=0|\boldsymbol{x}_{2i}, T_{2i}...) = \begin{cases} q & \text{if } M_i^{(1)} < M_i^{(2)}, \\ 1-q & \text{if } M_i^{(1)} > M_i^{(2)}, \\ 0.5 & \text{if } M_i^{(1)} = M_i^{(2)}, \end{cases}$$

$$P(T_{2i+1}=0, T_{2i+2}=1|\boldsymbol{x}_{2i}, T_{2i}...) = 1 - P(T_{2i+1}=1, T_{2i+1}=0|\boldsymbol{x}_{2i}, T_{2i}...),$$

where

- $0.5 < q < 1$.
- Note: $T_{2i+1} = T_{2i+2} = 0, 1$ is not allowed.

(6) Repeat Steps (4) and (5) until finish.

- A **smaller** value of $M(n)$ indicates a **better** covariate balance.
- $q = 0.75$. More discussion in Hu and Hu (2012).
- Units are not observed sequentially; however, we allocate them sequentially (in pairs).
- Better covariate balance.
- $n!$ different possible sequences. Similar performance.

Properties of CAM

Under CAM, suppose $\boldsymbol{x}_i$ is i.i.d. multivariate normal; then

$$M(n) = O_p(n^{-1}).$$

Note:

- Under CR, $M_{\mathsf{CR}}(n) \sim \chi^2_{df=p}$, a stationary distribution of a Chi-square distribution with p degrees of freedom, regardless of n.

- Under RR, $M_{\mathsf{RR}}(n) \sim \chi^2_{df=p} | \chi^2_{df=p} < a$, a stationary distribution of a Chi-square distribution with p degrees of freedom conditional on $M_{\mathsf{RR}}(n) < a$, regardless of n.

- Under CAM, $M(n) \to 0$ at the rate of $1/n$.
  - More units, better balance.
  - Advantages of CAM in large n.

Properties of CAM

As p increases,

- Under CR, the stationary distribution becomes flatter, poorer covariate balance.
- Under RR, the stationary distribution becomes flatter, poorer covariate balance.
- Under CAM, $M(n) \to 0$ at the rate of $1/n$, regardless of p.
  - The effect of p on $M(n)$ is less severe than CR and RR.

Properties of CAM

- Adaptive based on covariates.
- Works for sequential experiments, just estimate the covariance matrix sequentially.
- Capable for large p and large n.
- Better covariate balance.
- Less computational time.

2.2 Estimating treatment effect

A natural setup of A/B testing:

- The observed outcome y_i, $i = 1, ..., n$, for each unit.
- Let $y_i(T_i)$ represents the potential outcome of the i-th unit under the treatment T_i.
- $y_i = y_i(1)T_i + y_i(0)(1 - T_i)$.
- The average treatment effect (ATE) is

$$\tau = \frac{\sum_{i=1}^{n} y_i(1)}{n} - \frac{\sum_{i=1}^{n} y_i(0)}{n}.$$

- The fundamental problem: only observe $y_i(T_i)$ for one particular T_i, therefore, τ cannot be calculated directly.

A natural estimate, $\hat{\tau}$:

$$\hat{\tau} = \frac{\sum_{i=1}^{n} T_i y_i}{\sum_{i=1}^{n} T_i} - \frac{\sum_{i=1}^{n} (1 - T_i) y_i}{\sum_{i=1}^{n} (1 - T_i)},$$

- $\hat{\tau}$ could be bad with imbalance in covariates.
- Example: estimate the drug effect using treatment groups with predominately male and female patients. Cannot remove the gender effect.

Theoretical properties:

(1) Unbiasedness: under CAM, $E(\hat{\tau}) = \tau$.

(2) Under CAM, $Var(\hat{\tau})$ attains the lower bound asymptotically.

(3) This implies that

$$Var_{CAM}(\hat{\tau}) < Var_{RR}(\hat{\tau}) < Var_{CR}(\hat{\tau}).$$

2.3 Examples

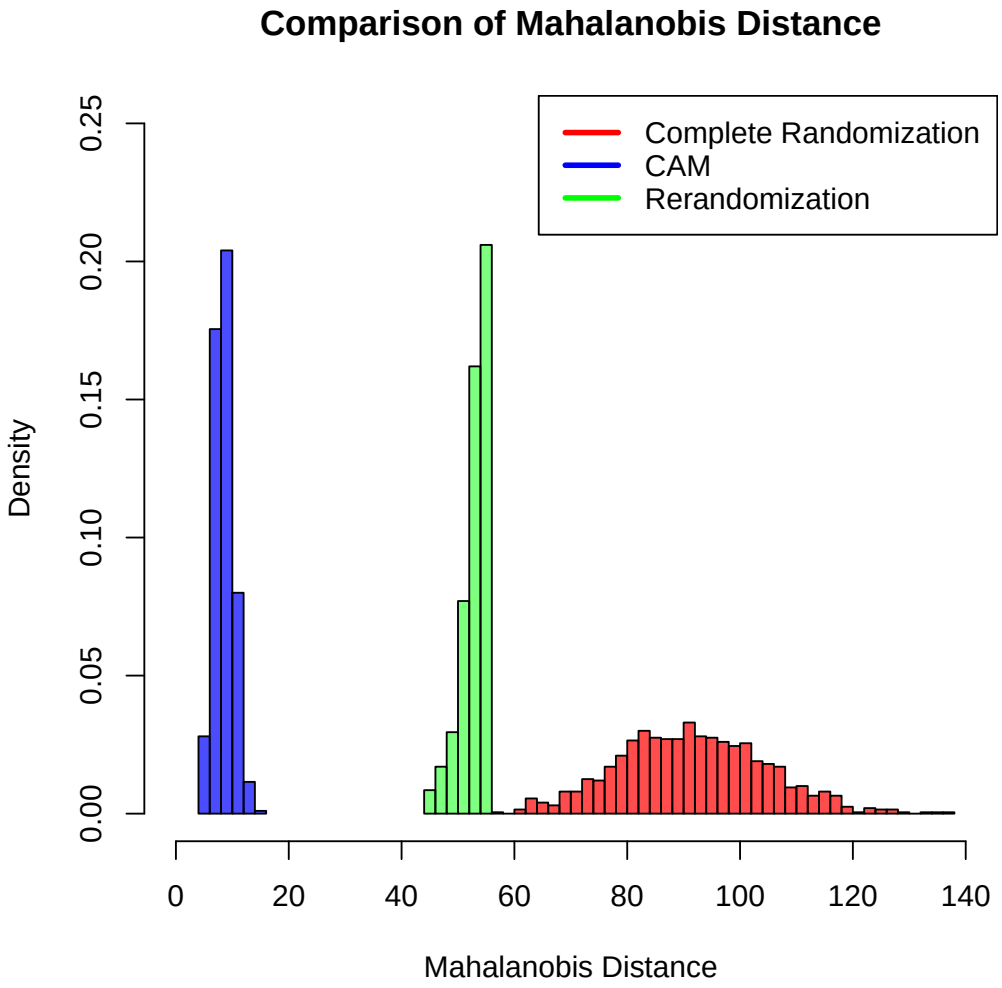
Example 2: Project GATE

- Two treatment groups:
  Treatment: were offered GATE services; control: were not offered GATE services.

- $p = 105$ (covariates obtained from the application packages, 13 continuous and 92 categorical)

- Sample size $n = 3,448$ (out of 4,198 participants from who answered the evaluation survey 6 months after the assignment)

- Original allocation $M = 75.27$, moderate covariate imbalance.
- We repeat the allocation 1,000 times for these participants using CAM, complete randomization and rerandomization.


Comparison of Mahalanobis Distance
Complete Randomization
CAM
Rerandomization
Density
0.00
0.05
0.10
0.15
0.20
0.25
0
20
40
60
80
100
120
140
Mahalanobis Distance

CAM vs Rerandomization

The Maximum of Malahanobis distances obtained from CAM is 12. If we set the balance criterion for rerandomization to $M < 12$, the probability of acceptance $P_a = P(\chi^2_{df=105} < 12) = 3.4 \times 10^{-31}$, which means nearly impossible for rerandomization to achieve a similar balance level as CAM.

We set $P_a = 2 \times 10^{-5}$ for Re-randomization to have similar computational time with CAM.

Estimation.

- The outcome variable (0/1): has owned a business within 6 months after assignment or not.

- After the allocation, we simulate the outcome variable according to

$$\text{logit}(P(y_i^{\text{sim}} = 1)) = \hat{\mu}_1 T_i^{\text{sim}} + \hat{\mu}_2 (1 - T_i^{\text{sim}}) + x_i^T \hat{\boldsymbol{\beta}} + \epsilon^{\text{sim}},$$

  where $\hat{\mu}_1$, $\hat{\mu}_2$ and $\hat{\boldsymbol{\beta}}$ are obtained from fitting regression to original data. ϵ^{sim} is drawn from the residuals of that regression.

Compare the estimation performance (PRIV) of CAM and rerandomization.

Method	PRIV	u_n or v_a
CAM	17.7%	0.081
Rerandomization	10.5%	0.505

3 Other new procedures

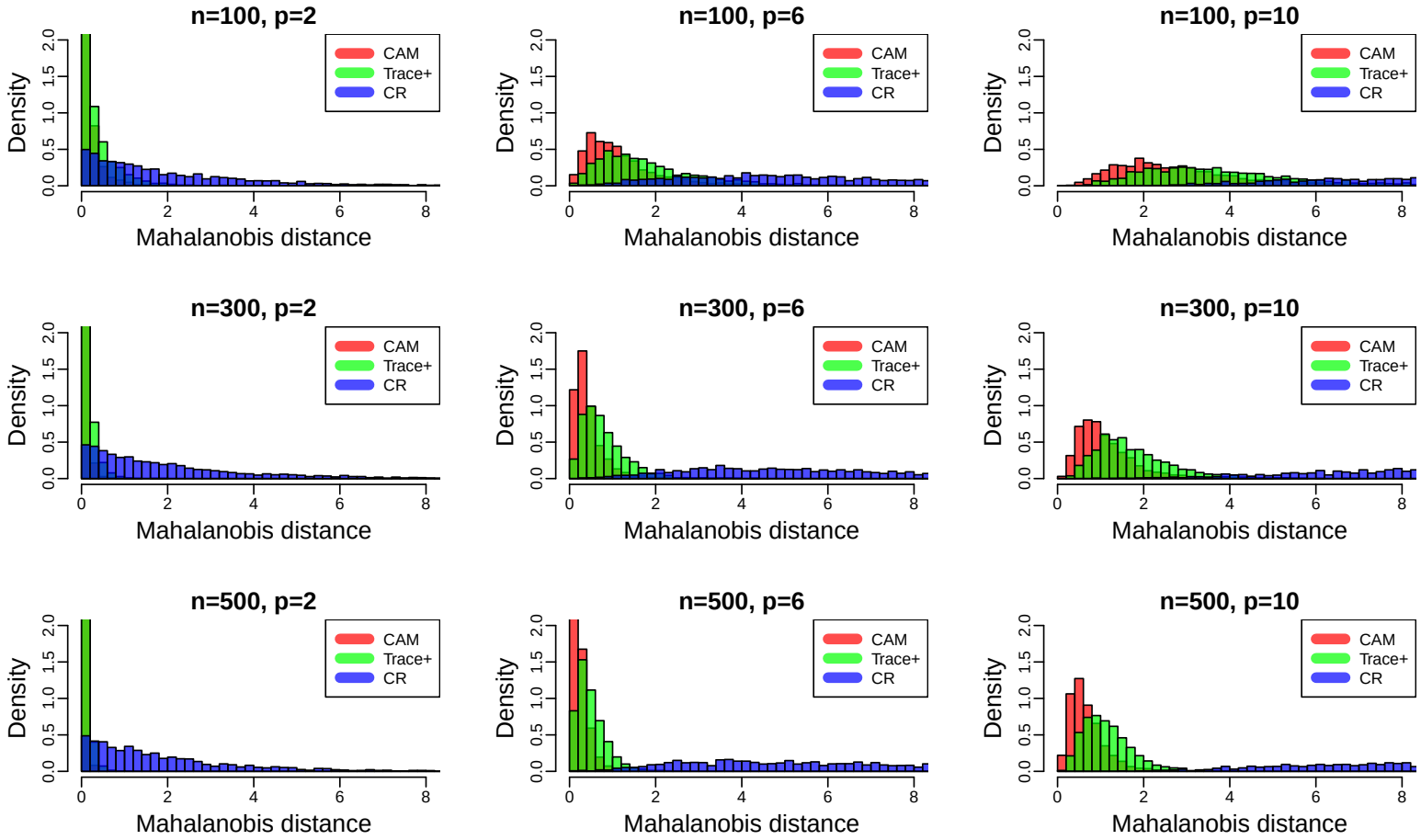
3.1 Balancing both mean and covariance matrix

The CAM only considers the mean of two groups. Covariance structure is also important in statistical analysis. We (with Ping Li, Lixin Zhang and Wei Ma, 2021) proposed a new and unified family of covariate-adaptive designs and obtained important theoretical properties. Here is an important special case:

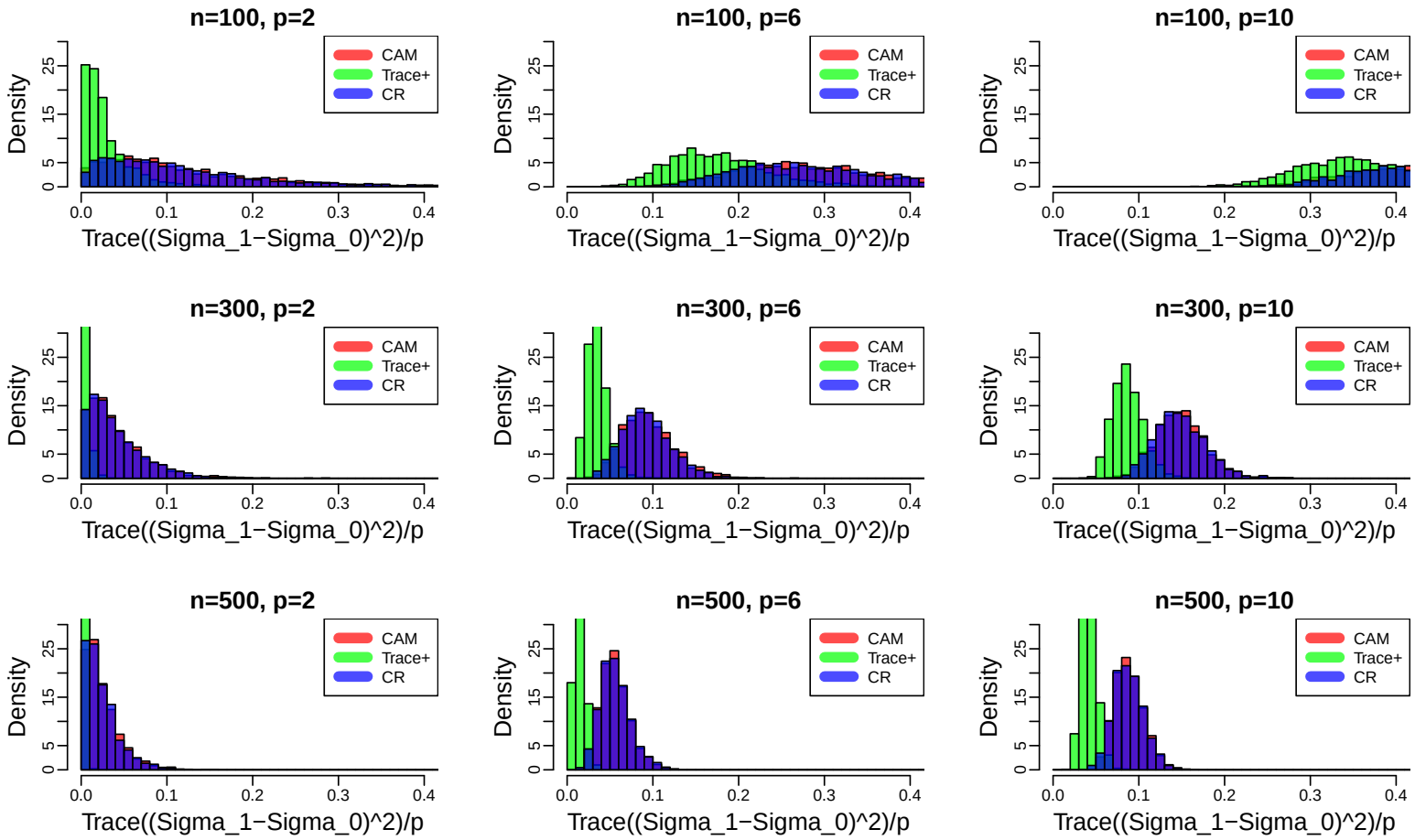
$$IB_T(n) = (\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2)^T \mathrm{cov}(\boldsymbol{x})^{-1} (\bar{\boldsymbol{x}}_1 - \bar{\boldsymbol{x}}_2) + trace\Big\{ \big(\hat{\Sigma}_1 - \hat{\Sigma}_2\big)^2 \Big\}/p$$

where $\hat{\Sigma}_1$ and $\hat{\Sigma}_2$ are the sample covariance matrices for two treatment groups.

New method vs CAM vs Complete randomization



New method vs CAM vs Complete randomization



Some other procedures:

(i) both discrete and continuous covariates (Zhou and Hu, 2021).

(ii) multi-treatments (Hu, Ye and Zhang, 2021).

3.2 Network Adaptive designs

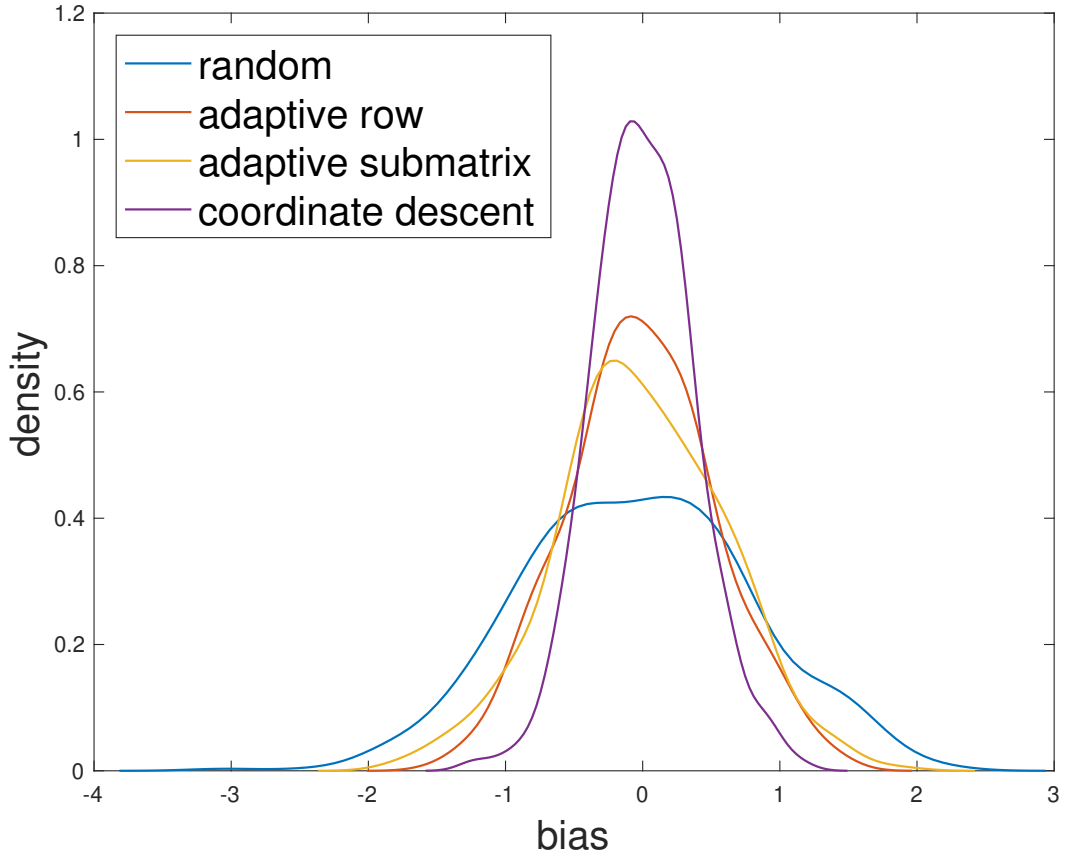
Network adaptive designs for A/B testing (Zhou, Li and Hu, 2020), Cluster-adaptive designs (Zhou, Liu, Li and Hu, 2021; Liu, Zhou, Li and Hu, 2021):

Let a graph G be represented by a $n \times n$ symmetric adjacency matrix $A = [A_{ij}]$.

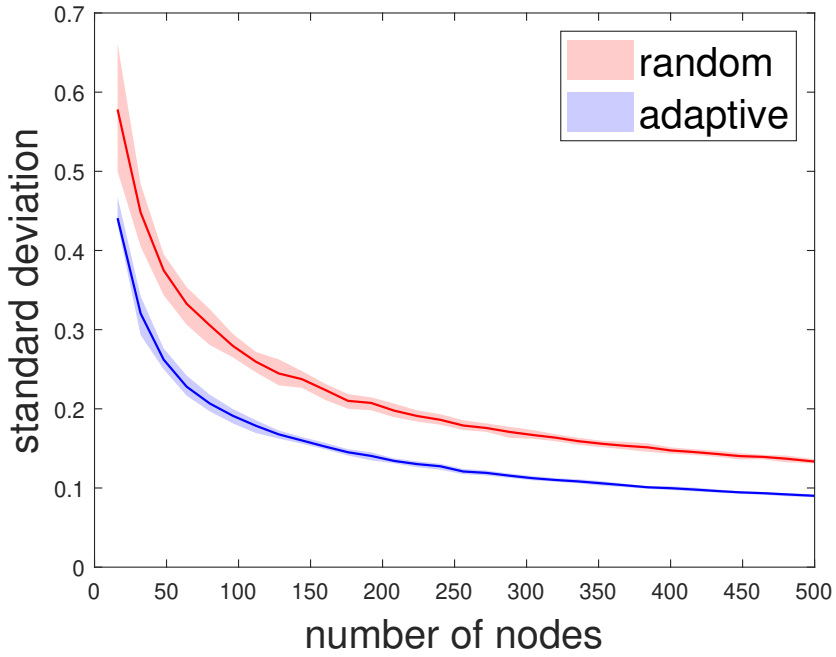
Balancing n-dimensional binary vectors, the network , is hard.

Zhou, Li and Hu (2020) proposed several methods and discussed their theoretical properties.

New methods vs Complete randomization (ATE)



New methods vs Complete randomization (MSE)



4 Statistical Inference: A review and a look forward

- Covariate-adaptive randomization is extensively used in clinical trials to enhance balance over covariates.
- There have been concerns over the validity of associated inference.
- In the past decade, significant theoretical progress has been made on statistical inference for covariate-adaptive randomization procedures.
- In this talk, we review some recent works in this area and also discuss some potential topics for future research.

4.1 Brief history (prior to 2010)

- Most early studies of inference for CAR were based on simulations.
- The validity of classical tests was also discussed from philosophical and regulatory perspectives.
- Very little work however had been done to tackle the theoretical properties of inference for CAR.

Simulation based results

- Forsythe (1987) suggests that 'minimization should be considered for group assignment only if all variables used in minimization are also to be used as covariate' to achieve valid inference.
- Several numerical studies indicate conservativeness of unadjusted analysis in covariate-adaptive randomized clinical trials (see, for example, Birkett, 1985; Forsythe, 1987).

Philosophical view

Senn (2004) gives a philosophical view of why the treatment effect should be adjusted for the covariates used in the design.

" If an investigator uses [covariate-adaptive randomization], she or he is honour bound, in my opinion, as a very minimum, to adjust for the factors used to balance, since the fact that they are being used to balance is an implicit declaration that they have prognostic value. In the case of a linear model the standard error quoted will generally be too small if this is not done ..."

Regulatory point of view

The applications of CAR were restricted in part due to the concerns from regulatory agencies. For example, EMA expressed skepticism about the use of minimization in 2003.

"...Dynamic allocation [minimization] is strongly discouraged. However, if it is used, then it is imperative that all factors used in the allocation scheme be included as covariates in the analysis. Even with this requirement, it remains controversial whether the analysis adequately reflects the randomization scheme."

Conservative Hypothesis Testing for Treatment Effect

Underlying model:

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

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 1: 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 2: 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

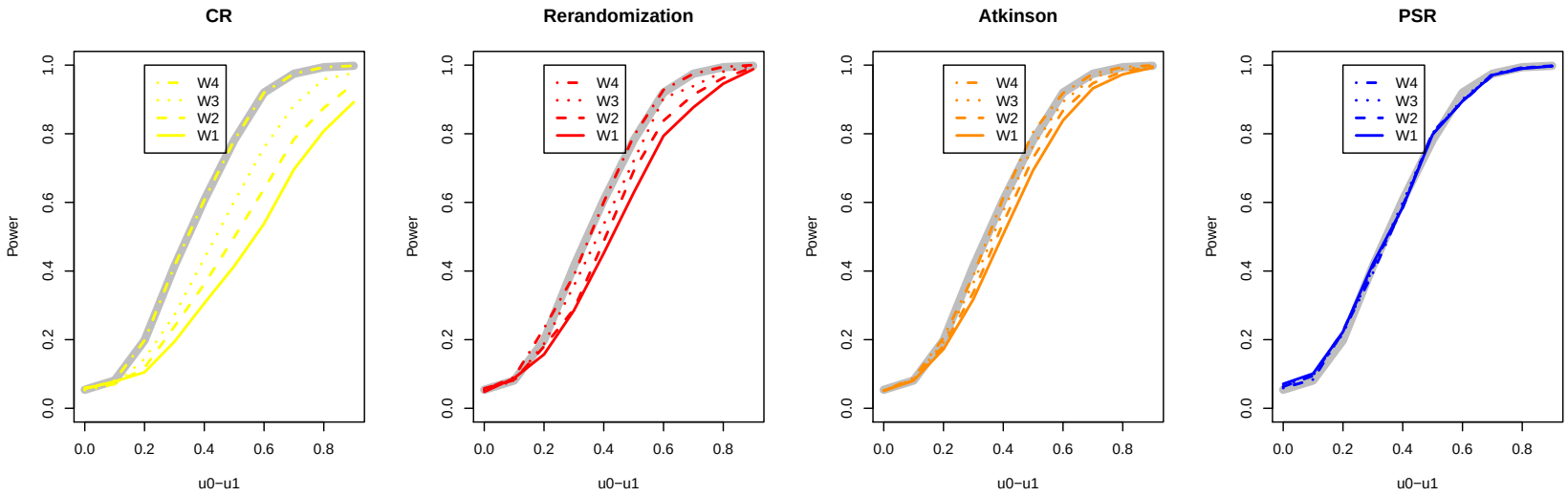


Figure 1: Power against $\mu_1 - \mu_2$ using estimated asymptotic distribution's critical values and p-values. Sample size $n = 500$. Note that we plot the power of W4 under CR in bold gray curves in all the panels for a better comparison among different randomizations.

Recent advances

To increase the understanding and acceptance of CAR in clinical trials, it is crucial to address the following questions:

- How to achieve valid inference when CAR is used?
- What are the properties of classical test procedures under different data-generating models?
- What is the relationship between covariate balance induced by CAR and associated inference?
- Is there a way to correct classical tests if they are invalid?

These questions are tackled in several recent papers, such as Shao et al. (2010, Biometrika), Ma et al. (2015, JASA), Bugni et al. (2018, JASA), Ma et al. (2020,JASA), etc.

4.2 Recent advances

Sufficient conditions for valid test

Shao et al. (2010) prove sufficient conditions for a valid test under CAR:

1. the covariates used in the randomization must be a function of the covariates used in the test procedure
2. the model for analysis must be correctly specified

The conclusion justifies the simulation based result stated above that all the covariates used in CAR should also be adjusted in the analysis, provided that the model is correct!

Underlying model vs. analysis model

- The validity requires a correctly specified model, but the true underlying data-generating model is usually unknown.
- Unadjusted test, such as the two sample t-test, is commonly used in practice due to many reasons (e.g., simplicity, robustness to model misspecification).
- It is important to investigate the inference properties when the model is misspecified, i.e., when the model for analysis is different from the true underlying model.

The two sample t-test

Shao et al. (2010) study the two sample t-test under CAR based on the following assumptions:

- **Underlying model**: A simple linear model with univariate covariate,

$$Y_{ij} = \mu_j + \beta X_i + \varepsilon_{ij},$$

  where Y_{ij} be the outcome of patient i under treatment j, $j = 1, 0$.

- **Analysis model**: The two sample t-test with test statistic

$$T_S = (\bar{Y}_1 - \bar{Y}_0)/(S_1^2/n_1 + S_0^2/n_0)^{1/2}.$$

- **Randomization**: Stratified biased coin design.

Conservativeness of t-test – Shao et al. (2010)

- It is shown that, unless trivial cases, the t-test is conservative in terms of smaller Type I error than the nominal levels, i.e.,

$$\lim_{N\to\infty} \mathcal{Y},\mathcal{T}(|T_S| > c_\alpha|\mathcal{X}) \leq \alpha_0 < \alpha,$$

  where $(\mathcal{T}, \mathcal{Y}, \mathcal{X})$ is the set of treatment assignments, outcomes, and covariates, respectively.

- An intuitive explanation of the conservativeness is that $\bar{Y}_1 - \bar{Y}_0$ is less variable than the variance estimator due to the covariance between the sample means induced by the CAR.

- The asymptotic normality of T_S is established so the extent of conservativeness and power can be assessed analytically.

More general randomization procedures

Ma et al. (2015) study testing hypothesis of covariate-adaptive randomized clinical trials in the linear model framework.

- **Randomization**: A large class of CAR procedures:
  1. the overall imbalance is $O_P(1)$
  2. the marginal imbalances for all covariates $O_P(1)$
- The class includes minimization, stratified permuted block design, stratified biased coin design, the methods in Hu and Hu (2012), etc.

Testing hypothesis for linear models

- **Underlying model**: A linear model with multivariate covariates,

$$Y_i = \mu_1 T_i + \mu_0(1 - T_i) + \sum_{k=1}^{p+q} \beta_k X_{i,k} + \varepsilon_i,$$

- **Analysis model**: A linear model with a subset of covariates,

$$(Y_i) = \mu_1 T_i + \mu_0(1 - T_i) + \sum_{k=1}^{p} \beta_k X_{i,k},$$

with the usual analysis of covariance (ANCOVA) test statistic T_C. Two extreme cases: When $q = 0$, all of the covariates are included in the analysis model, and when $p = 0$, no covariates are included.

Main results of Ma et al. (2015)

Under a large class of covariate-adaptive designs,

- The hypothesis testing to compare treatment effects is usually conservative in terms of small Type I error.
- The hypothesis testing to compare treatment effects is usually more powerful than complete randomization.
- The hypothesis testing for significance of covariates is still valid.

Some remarks for Ma et al. (2015)

- The results are more general than those in Shao et al. (2010) and provide more insight into the effect of adjusting for several covariates.
- The condition of $O_P(1)$ marginal imbalances is proved for the minimization method, and hence all the derived results apply to this most commonly used CAR method in practice.

More general randomization procedures

Ma et al. (2020) further extend the scope of CAR procedures to study associated inference.

- The "strong" balance property does not hold for some CAR, e.g., the $O_p(1)$ imbalances are not satisfied for the Atkinson's optimal design.
- So far all the CAR methods are based on only discrete covariates. It is desirable to have a framework that applies for both continuous and discrete covariates.
- **Randomization**: CAR with two widely satisfied conditions:
  1. Global balance: $n^{-1}\sum_{i=1}^{n}(2T_i-1) \xrightarrow{p} 0$.
  2. Covariate balance: $n^{-1/2}\sum_{i=1}^{n}(2T_i-1)X_i \xrightarrow{d} \xi$, where ξ is a $(p+q)$-dimensional random vector with $\mathbb{E}[\xi]=0$.

Key insights of Ma et al. (2020)

- Under $H_0 : \mu_1 = \mu_0$,

$$T_C \xrightarrow{d} \lambda_1 Z + \lambda_2 \beta_{\text{ex}}^t \xi_{\text{ex}},$$

  where β_{ex} and ξ_{ex} correspond to the q covariates of β and ξ that are excluded from the analysis model, Z is a standard normal random variable that is independent of ξ_{ex}.

- The first component is due to random errors and remains invariant under different CAR methods, while the second component reflects how well the covariates are balanced under a specific CAR.

- This result explicitly unveils the relationship between covariate balance induced by CAR and associated inference.

More general models

Bugni et al. (2018) take an different approach to generalize the results in Shao et al. (2010) by assuming more general underlying models.

- **Underlying model**: A fully nonparametric model

$$(Y_{ij}|X_i) = \mu_j + m_j(X_i),$$

  where $m_j(X_i)$, $j = 1, 0$, is an unknown function of X_i with finite second moment.

- **Analysis model**: The two sample t-test and the test based on ANCOVA.

- **Randomization**: Stratified randomization for which the fraction of units within each stratum tends to be normally distributed.

Summary

Paper	Covariate-adaptive randomization	Underlying model	Analysis model
Shao et al. (2010)	Stratified biased coin design	Linear model with univariate covariate	two sample t-test, ANCOVA
Ma et al. (2015)	Designs with $O_P(1)$ overall and marginal imbalances	Linear model with multivariate covariate	two sample t-test, linear regression with all or a subset of covariates
Ma et al. (2019)	Designs with known property of the imbalance vector	Linear model with multivariate covariate	two sample t-test, linear regression with all or a subset of covariates
Bugni et al. (2018)	A class of stratified randomization	Nonparametric model	two sample t-test, linear regression with all covariates
Zheng and Zhang (2019)	A class of stratified randomization	Nonparametric model	Quantile regression

Corrected test

Several methods have been proposed to correct the conservativeness of the classical tests.

- To more accurately estimate the variance of $\bar{Y}_1 - \bar{Y}_0$, using either bootstrap (Shao et al., 2010) or model-based approach (Ma et al., 2020)
- A nonparametric method proposed for stratified randomization (Bugni et al., 2018)
- Randomization/permutation test (Rosenberger and Lachin, 2015; Bugni et al., 2018; Ma et al., 2020; Rosenberger et al., 2019)

All these tests have certain robustness against model misspecification, but a comprehensive comparison still need to be performed.

4.3 Additional issues

Extension to non-continuous outcomes

- Logistic regression and other generalized linear models: Shao and Yu, (2013), Li et al. (2021)
- Time-to-event outcomes: Xu et al. (2016), Ye and Shao (2020), Qiu and Hu (2021)
- Longitudinal analysis: Weng et al. (2017)

Combination with other adaptive designs

- Sequential monitoring: Zhu and Hu (2019)
- Interim analysis: Li, Ma and Hu (2021)
- Seamless phase II/III trials: Ma et al. (2020)

Other topics

- Multiple treatments: Bugni et al. (2019) on regression, Ma et al. (Working paper) on the Simes test or the Dunnett test
- Measurement error: Wang and Ma (2020)
- Statistical software (carat) (Ma, Ye, Tu and Hu, 2021): A new R package available.
- Balancing unobserved covariates (Liu and Hu, 2021a).
- The Impacts of unobserved covariates of CAR clinical trials (Liu and Hu, 2021b).

5 Conclusion

- Some new covariate-adaptive designs
- 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!