Statistical Methods for Analyzing Randomized Trials and Brain Imaging Data

Bingkai Wang, February 2021

Outline

1. Stratified Randomization in Clinical Trials
   - How to improve precision with stratified randomization and covariate adjustment?
   - 35 min
2. Analysis of Covariance in Randomized Trials
   - When and how to use analysis of covariance?
   - 5 min
3. Joint Modeling Multiple Covariance Matrices
   - How to identify brain networks that are common across people via analysis of their functional brain imaging data?
   - 5 min

Part 1

Model-Robust Inference for Clinical Trials that Improve Precision by Stratified Randomization and Covariate Adjustment

Bingkai Wang, Ryoko Susukida, Ramin Mojtabai, Masoumeh Amin-Esmaeili, Michael Rosenblum

arXiv, 2020, https://arxiv.org/abs/1910.13954

Open question

In randomized clinical trials, how can we do model-robust inference under stratified randomization?

Open question

In randomized clinical trials, how can we do model-robust inference under stratified randomization?

- **Randomized clinical trial (RCT)**: gold standard for evaluating the efficacy of new treatments.
- **Model-robust inference**: valid statistical inference even when the assumed model is wrong.
- **Stratified randomization**: treatment allocation stratified by baseline strata using permuted blocks.

Open question

In randomized clinical trials, how can we do model-robust inference under stratified randomization?

- **Randomized clinical trial (RCT)**: gold standard for evaluating the efficacy of new treatments.
- **Model-robust inference**: valid statistical inference even when the assumed model is wrong.
- **Stratified randomization**: treatment allocation stratified by baseline strata using permuted blocks.

Female	1	0	1	0	0	0	1	1	⋯
Male	0	0	1	1	0	1	0	1	⋯

Background

- Stratified randomization is widely used in practice.
  - Lin et al. (2015) surveyed 224 randomized trials, 183 (70%) used stratified randomization.

- Stratified randomization ensures treatment balance within each stratum.
  - **Simple randomization** allocates treatment by independent coin flips and does not ensure treatment balance.

Background

- The difference between simple randomization and stratified randomization is usually ignored.
  - Less than 50% of trials that used stratified randomization adjusted for strata in their analyses (Kahan and Morris, 2012).
- For many commonly used estimators, little is known about their asymptotics under stratified randomization.
  - The asymptotic results are generally needed to construct confidence intervals.

Background

- The difference between simple randomization and stratified randomization is usually ignored.
  - Less than 50% of trials that used stratified randomization adjusted for strata in their analyses (Kahan and Morris, 2012).
- For many commonly used estimators, little is known about their asymptotics under stratified randomization.
  - The asymptotic results are generally needed to construct confidence intervals.

First goal of this paper

Derive the asymptotic results for commonly used estimators in RCTs under stratified randomization.

Background

- Recent work showed that stratified randomization can lead to precision gain for some simple estimators under certain scenarios. (Shao et al., 2010; Bugni et al., 2018; Li and Ding, 2020, etc)
- Covariate adjustment is also known as a tool to improve precision. (FDA, 2019, 2020)
- Precision gain can be translated into sample size reduction.

Second goal of this paper

Improve precision by combining stratified randomization and covariate adjustment.

Example

CTN44 (Campbell et al., 2014) is an RCT evaluating internet-delivered treatment for substance abuse.

- **Treatment**: Therapeutic Education System versus Treatment as usual.
- **Outcomes**: number of negative urine tests (continuous) and time to abstinence (time-to-event).
- **Baseline variables**: age, sex and urine laboratory result.
- **Stratified randomization**: treatment allocation stratified by abstinence status at baseline (4 strata).
- **Estimands**: Average treatment effect and survival function.

Example

CTN44 (Campbell et al., 2014) is an RCT evaluating internet-delivered treatment for substance abuse.

- **Treatment**: Therapeutic Education System versus Treatment as usual.
- **Outcomes**: number of negative urine tests (continuous) and time to abstinence (time-to-event).
- **Baseline variables**: age, sex and urine laboratory result.
- **Stratified randomization**: treatment allocation stratified by abstinence status at baseline (4 strata).
- **Estimands**: Average treatment effect and survival function.

Questions

1. How to perform valid statistical inference?
2. How to improve its precision? (12%)

Our contributions

1. I derived the **model-robust** consistency and asymptotic normality for a wide class of estimators under stratified randomization, which covers essentially all estimators used in primary analyses of RCTs.

2. I showed how to improve precision by stratified randomization and covariate adjustment compared to standard practice.

3. The above results also hold for the biased-coin covariate-adaptive design.

Related work

- Shao et al. (2010); Shao and Yu (2013) proved the validity of the two-sample t-test under the biased-coin design assuming generalized linear model.
- Ma et al. (2015, 2018) assumed a linear model and derived the asymptotic distribution of the test statistic of ATE for the ANCOVA estimator and a class of covariate-adaptive designs.
- **Bugni et al. (2018) established the asymptotic theory of the unadjusted estimator and the ANCOVA estimator (adjusting for strata only) of ATE for a wide range of covariate-adaptive designs.**
- Ye and Shao (2020) derived asymptotics for log-rank and score tests in survival analysis under covariate adaptive randomization.
- Li and Ding (2020) established the asymptotic theory for the ANCOVA estimator under covariate-adaptive randomization in the randomization inference framework.
- More recent papers: Yang et al. (2020); Ma et al. (2020); Ye et al. (2020).

Definition

For participant $i = 1, \ldots, n$, we define

- Y_i is the outcome variable, which can be continuous, binary or time-to-event,
- M_i denotes whether Y_i is observed ($M_i = 1$) or missing ($M_i = 0$),
- A_i is a binary treatment indicator,
- $\boldsymbol{X}_i$ is a vector of baseline variables, which includes randomization strata S_i.

We use the Neyman-Rubin causal model and assume

$$Y_i = Y_i(A_i) = A_i Y_i(1) + (1 - A_i) Y_i(0),$$
$$M_i = M_i(A_i) = A_i M_i(1) + (1 - A_i) M_i(0),$$

where $Y_i(a)$ is the potential outcome and $M_i(a)$ is the "potential non-missing indicator" for $a = 0, 1$.

Definition

Simple randomization

$(A_1, \ldots, A_n)$ are assigned by independent Bernoulli draws with $P(A_i = 1) = \pi, \pi \in (0, 1)$.

- Large treatment imbalance may occasionally occur.
- Data vectors $(Y_i, M_i, A_i, X_i), i = 1, \ldots, n$ are independent, identically distributed.
- Classical theorems, such as the central limit theorem, can be directly applied to prove asymptotics.

Treatment	1	1	1	0	0	1	1	1
Sex	F	M	F	F	F	M	M	M

Definition

Stratified (permuted block) randomization

For each stratum, randomly permuted blocks with fraction π 1's and $1-\pi$ 0's are used for sequential allocation.

- Treatment balance is ensured within each stratum.
- Data vectors $(Y_i, M_i, A_i, X_i), i = 1, \ldots, n$ are identically distributed, but **not** independent.

$$P\left(A_1 = 1, \ldots, A_n = 1\right) = 0 \text{ instead of } \pi^n.$$

- The dependency among data leads to the main challenge to derive the asymptotics.

Female	1	0	1	0	0	0	1	1	⋯
Male	0	0	1	1	0	1	0	1	⋯

Estimands

Binary and continuous outcomes

Our goal is to estimate a population parameter, for example, the average treatment effect (ATE)

$$\Delta^* = E[Y_i(1) - Y_i(0)].$$

Time-to-event outcomes

Our goal is to estimate the survival curve $\{S_a(t) : t \in [0, \tau]\}$ for each $a = 0, 1$, where

$$S_a(t) = P(Y(a) > t)$$

and $[0, \tau]$ is the time window of interest.

Binary and continuous outcomes

Observed data

$(A_i, \boldsymbol{X}_i, Y_i M_i, M_i)$ for $i = 1, \ldots, n$.

M-estimators $\widehat{\Delta}$

M-estimators refer to a wide class of estimators that solve systems of estimating equations (van der Vaart, 1998, Ch. 5).
In general, an M-estimator of parameters $\boldsymbol{\theta}$ is the solution to the following estimating equations:

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

where $\boldsymbol{\psi}$ is a column vector of known functions.

Binary and continuous outcomes

Examples of M-estimators

- the ANCOVA estimator (for continuous outcomes).
  It is defined as the ordinary least squares estimator of β_A in the linear regression working model

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

  The corresponding estimating equations are

$$\boldsymbol{\psi}(A, \boldsymbol{X}, Y, M; \beta_0, \beta_A, \boldsymbol{\beta}_{\boldsymbol{X}}) = \{Y - (\beta_0 + \beta_A A + \boldsymbol{\beta}_{\boldsymbol{X}}^t \boldsymbol{X})\} \begin{pmatrix} 1 \\ A \\ \boldsymbol{X} \end{pmatrix}.$$

- DR-WLS estimator (handling missing outcomes),
- Mixed-effects model for repeated measures (MMRM),
- Targeted maximum likelihood estimator (TMLE).

Binary and continuous outcomes

Assumptions

(i) The full data vector $(Y_i(1), Y_i(0), M_i(1), M_i(0), \boldsymbol{X}_i)$, $i = 1, \ldots, n$ are i.i.d.

(ii) **Missing at random**. $M(a)$ is independent of $Y(a)$ given $\boldsymbol{X}$ for $a = 0, 1$.

(iii) **Regularity conditions** that are similar to the classical conditions for simple randomization (Section 5.3 of van der Vaart, 1998).

(iv) The estimating equations $\boldsymbol{\psi}$ are properly chosen.

Binary and continuous outcomes

Theorem 1

Given the assumptions and under stratified randomization, an M-estimator $\widehat{\Delta}$ is consistent to Δ^* and satisfies

$$\sqrt{n}(\widehat{\Delta} - \Delta^*) \xrightarrow{d} N(0, V_{\text{strat}}),$$

with $V_{\text{strat}} \leq V_{\text{simple}}$, where V_{simple} is the asymptotic variance of $\widehat{\Delta}$ under simple randomization.

We also derive the analytical formula of V_{strat} and provide consistent estimators for V_{strat}.

Binary and continuous outcomes

Example

Given the assumptions, the ANCOVA estimator is consistent and asymptotically normal under stratified randomization, with

$$V_{\text{strat}} = V_{\text{simple}} - \frac{(1-2\pi)^2}{\pi(1-\pi)} E[Var\{Y_i(1) - Y_i(0)|S_i\}].$$

Binary and continuous outcomes

Example

Given the assumptions, the ANCOVA estimator is consistent and asymptotically normal under stratified randomization, with

$$V_{\text{strat}} = V_{\text{simple}} - \frac{(1-2\pi)^2}{\pi(1-\pi)} E[Var\{Y_i(1) - Y_i(0)|S_i\}].$$

Cases where stratified randomization does not improve precision

- 1:1 randomization ($\pi = 0.5$),
- or no treatment effect heterogeneity among strata ($Var\{Y_i(1) - Y_i(0)|S_i\} = 0$),
- ANCOVA also adjusts for treatment-by-strata interaction terms.

Time-to-event outcomes

Observed data

$(A_i, \boldsymbol{X}_i, U_i, \delta_i)$, where

- $U_i = \min\{Y_i, M_i\}$ is the event time,
- $\delta_i = I\{Y_i \leq M_i\}$ is the indicator for not being censored.

Estimator

The Kaplan-Meier estimator for the survival curve is defined as

$$\widehat{S}_a(t) = \prod_{j:T_j \leq t} \left(1 - \frac{\sum_{i=1}^n \delta_i I\{A_i = a\} I\{U_i = T_j\}}{\sum_{i=1}^n I\{A_i = a\} I\{U_i \geq T_j\}}\right),$$

where $\{T_j, j = 1, \ldots, m_n\}$ is the list of unique observed failure times.

Time-to-event outcomes

Assumptions

(i) The full data vector $(Y_i(1), Y_i(0), M_i(1), M_i(0), \boldsymbol{X}_i)$, $i = 1, \ldots, n$ are i.i.d.

(ii) **Censoring completely at random**: $M(a)$ is independent of $Y(a)$ for $a = 0, 1$.

(iii) $P(\min\{Y(a), M(a)\} > \tau) > 0$ for each $a = 0, 1$.

Time-to-event outcomes

Theorem 2

Given the assumptions and under stratified randomization,

1. $\widehat{S}_a(t)$ is consistent to $S_a(t)$,
2. $\left\{\sqrt{n}[\widehat{S}_a(t) - S_a(t)] : t \in [0, \tau]\right\}$ weakly converges to a Gaussian process $\mathcal{GP}(0, V_{\text{strat}})$ with $V_{\text{strat}}(t,t) \leq V_{\text{simple}}(t,t)$ for all $t \in [0, \tau]$,

where V_{simple} is the asymptotic covariance function of $\widehat{S}_a(t)$ under simple randomization.

We also derive the analytical formula of V_{strat} and provide consistent estimators for V_{strat}.

Proof Sketch

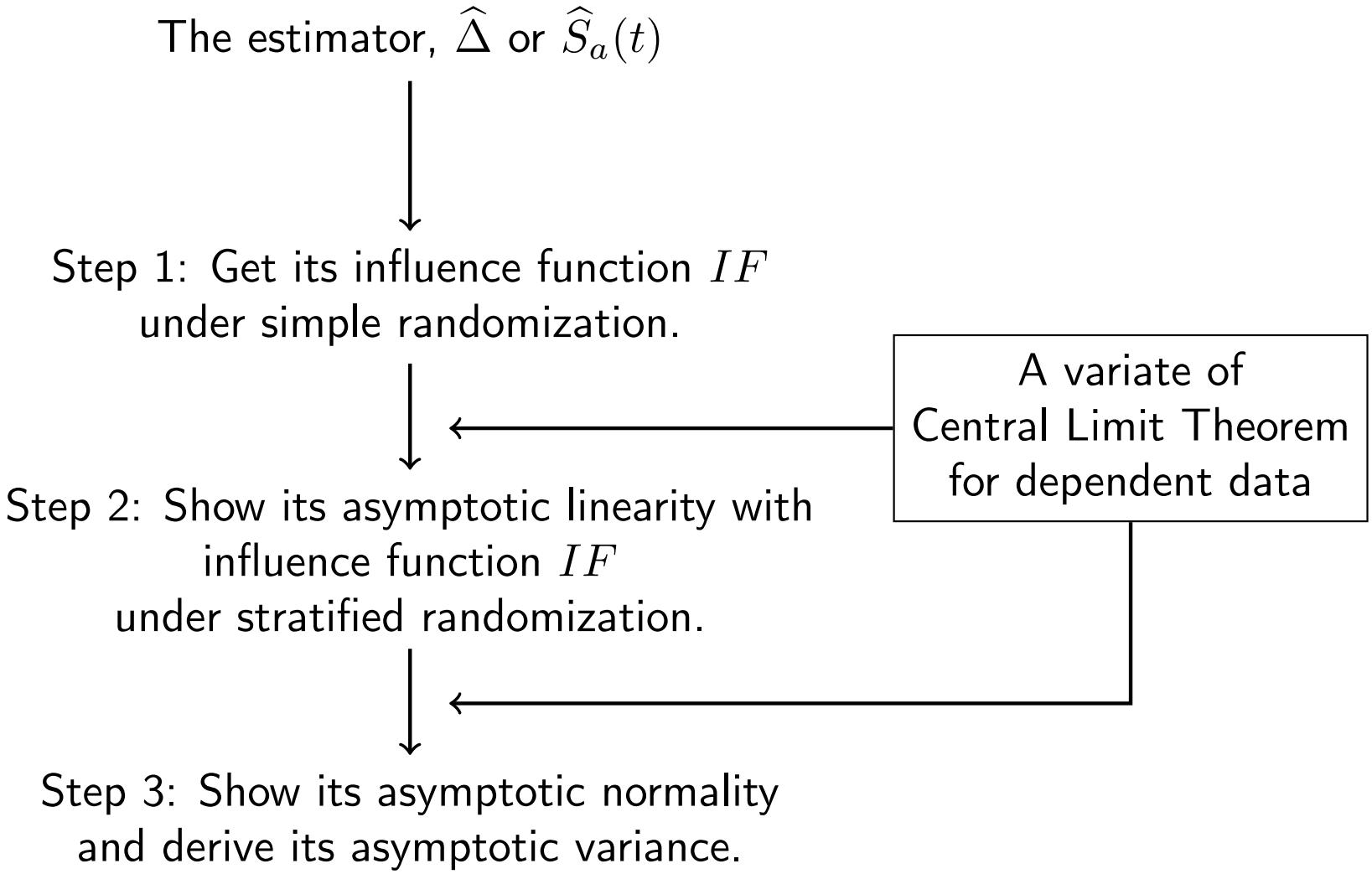
The estimator, $\widehat{\Delta}$ or $\widehat{S}_a(t)$

Step 1: Get its influence function IF
under simple randomization.

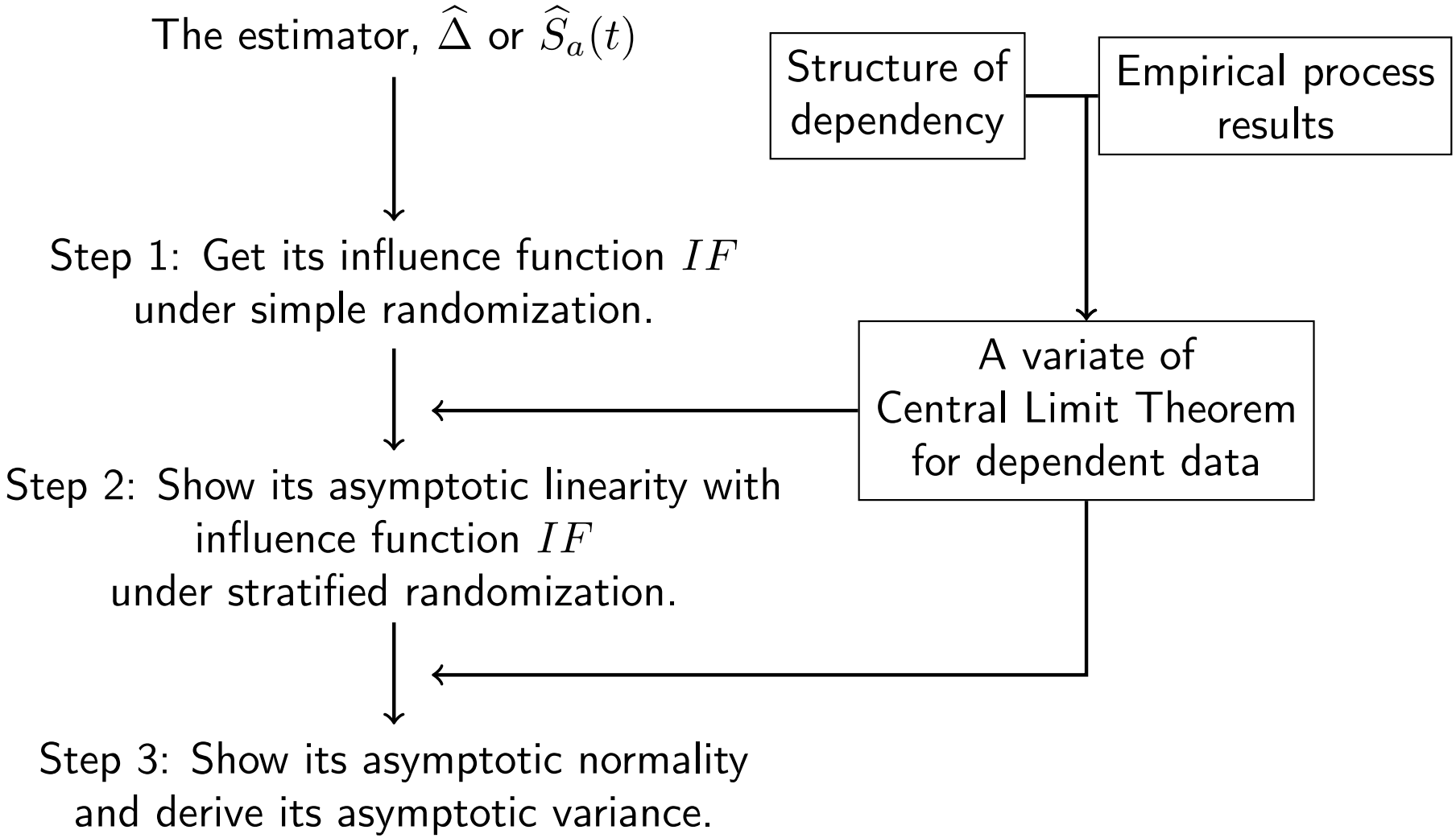
Step 2: Show its asymptotic linearity with
influence function IF
under stratified randomization.

Step 3: Show its asymptotic normality
and derive its asymptotic variance.

Proof Sketch



Proof Sketch



We can use two ways to improve precision/reduce variance of an estimator:

- Stratified randomization.
- Covariate adjustment.

How much variance reduction can be achieved?

Data example 1: variance reduction due to stratified randomization

- Trial: CTN44
- Time-to-event outcome: time to abstinence
- Group: Therapeutic Education System (treatment)
- Estimator: the Kaplan-Meier estimator

Visit	1	2	3	4	5	6
Survival probability	0.58	0.53	0.47	0.40	0.39	0.33
Proportional variance reduction $(1 - V_{\text{strat}}/V_{\text{simple}})$	11%	12%	11%	9%	7%	4%

Data example 2: variance reduction due to covariate adjustment

CTN03, CTN30 and CTN44 are RCTs of treatment of substance use disorder using stratified randomization.

- CTN03 has binary outcomes and CTN30 and CTN44 have continuous outcomes, all being measures of treatment success.
- Each study has $\sim$5 baseline variables.

Study	Number of Strata	Unadjusted estimator (95% CI)	Adjusted estimator (95% CI)	Proportional variance reduction
CTN03	3	-0.11(-0.21, -0.01)	-0.10(-0.19, -0.02)	**35%**
CTN30	4	0.02(-0.02, 0.05)	0.01(-0.02, 0.04)	**17%**
CTN44	4	-0.09(-0.14, -0.03)	-0.09(-0.14, -0.03)	**2%**

Discussion

Practical implications

1. When using stratified randomization, doing statistical inference with the correct variance (based on V_{strat} instead of V_{simple}) can avoid being conservative.
2. Adjusting for a set of preplanned baseline variables may lead to substantial variance reduction.

Limitations

1. Our results cannot handle cases where some strata have few participants.
2. For estimating the survival curves, our results only cover the Kaplan-Meier estimator.

Part 2

Analysis of Covariance (ANCOVA) in Randomized Trials: More Precision and Valid Confidence Intervals, Without Model Assumptions

Bingkai Wang, Elizabeth Ogburn, Michael Rosenblum

Biometrics, 2019, https://doi.org/10.1111/biom.13062

Motivation

In randomized trials, when and how to adjust for baseline variables are still debatable.

- Pocock et al. (2002) surveyed 50 randomized trials.
  - Only 12 reports emphasized adjusted over unadjusted analysis.
  - "The statistical emphasis on covariate adjustment is quite complex and often poorly understood."
- Austin et al. (2010) reviewed 114 randomized trials.
  - Only 39 presented adjusted estimator.
  - They suggested the need for an informed debate about the merit of adjusted estimators of treatment effect.

Goal of This Paper

We aim to clear common confusions related to covariate adjustment.

Our contributions

- I proved the variance estimator (output by standard statistical softwares) of the ANCOVA estimator is robust to model misspecification.
- I provided intuition for how covariate adjustment works by an analogy to linear regression.
- I provided recommendations on when and how to use covariate adjustment.

Data application

The table below shows data analyses of three completed trials for mild cognitive impairment, schizophrenia, and depression.

Trial Name	Unadjusted estimator (95% CI)	ANCOVA estimator (95% CI)	Variance reduction ($\hat{R}^2$)
MCI	-0.19(-0.49, 0.11)	-0.18(-0.45, 0.08)	25%
METS	-3.66(-6.83, -0.49)	-3.60(-6.71, -0.50)	4%
TADS	-1.44(-6.02, 3.15)	-4.36(-8.14, -0.58)	32%

Highlights of Results

- Variance reduction can be as high as 32%.
- For TADS, covariate adjustment leads to a significant results with p-value 0.01.

Practical Recommendations

1. The ANCOVA model can be used for inference even if it is misspecified.
2. Adjusting for the baseline score of the outcome measure is recommended.
3. For large trials, covariate adjustment is highly recommended.

Part 3

Semiparametric Partial Common Principal Component Analysis for Covariance Matrices

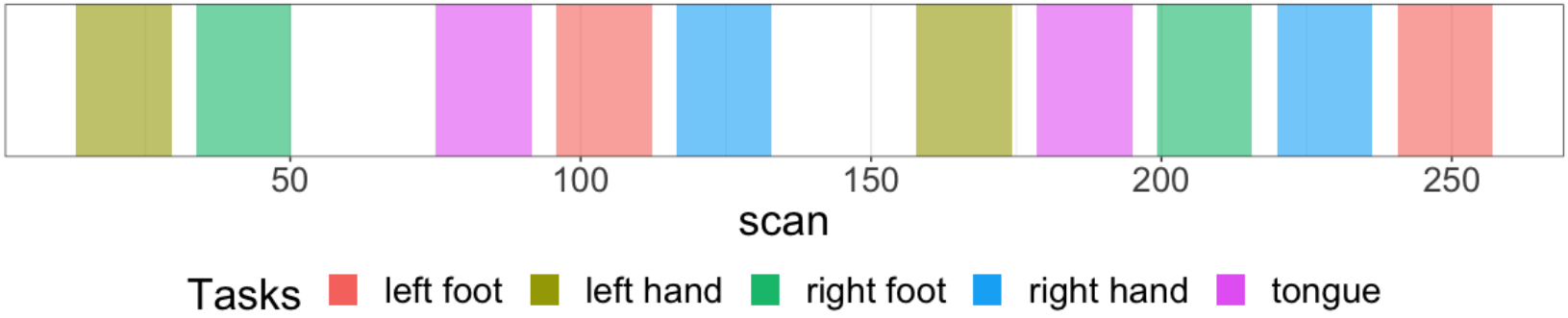
Bingkai Wang, Xi Luo, Yi Zhao, Brian Caffo

Biometrics, 2020, https://doi.org/10.1111/biom.13369

Motivating data example: HCP motor-task fMRI data

The data set includes fMRI scans of 136 healthy young adults.

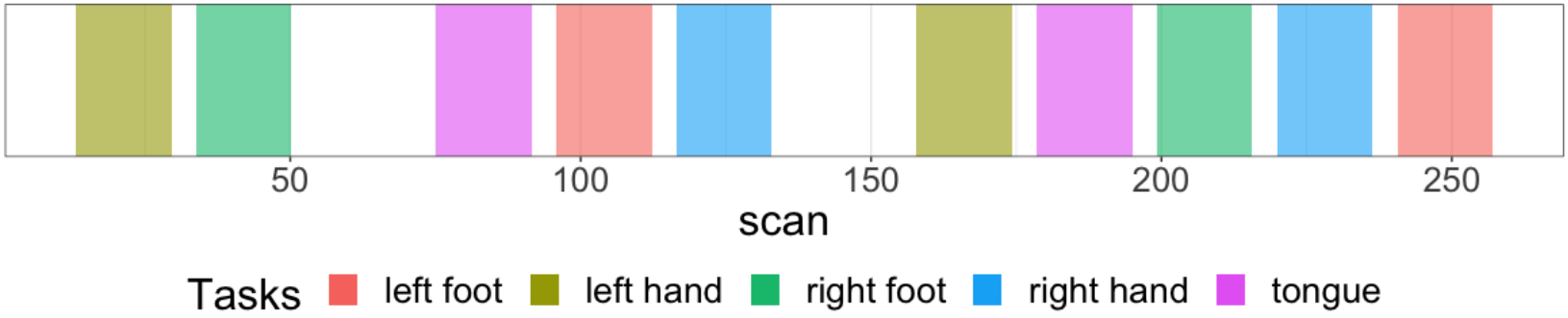
In each scan, the task fMRI consists of ten task blocks including moving tongue, hand or foot.



Motivating data example: HCP motor-task fMRI data

The data set includes fMRI scans of 136 healthy young adults.

In each scan, the task fMRI consists of ten task blocks including moving tongue, hand or foot.



Given the task fMRI data, we are interested in identifying task-related "common brain networks".

- A common brain network represents correlations in functional brain measures consistent across subjects.

Mathematical formulation

- For each subject $i, i = 1, \ldots, n$, the correlation among p brain regions forms a $p \times p$ covariance matrix $\mathbf{\Sigma}_i$.
- Each $\mathbf{\Sigma}_i$ can be decomposed by principal component analysis (PCA), resulting in n sets of principle components.
- A common brain network is a principle component that is the same across subjects.
- A common brain network may be associated with large or small eigenvalues for different subjects.

Goal of this paper

Given the sample covariance matrices $\mathbf{S}_i$, we aim to estimate:

1. The number of common brain networks.
2. Each common brain network.

Main contributions

1. I proposed consistent estimators of common brain networks and the number of common brain networks.

2. The proposed estimators relax assumptions made by existing literature and are able to handle more complex and realistic settings.

3. The proposed methods can be applied to other areas, such as economics.

Data application

Below is an example of common brain network identified by our method.

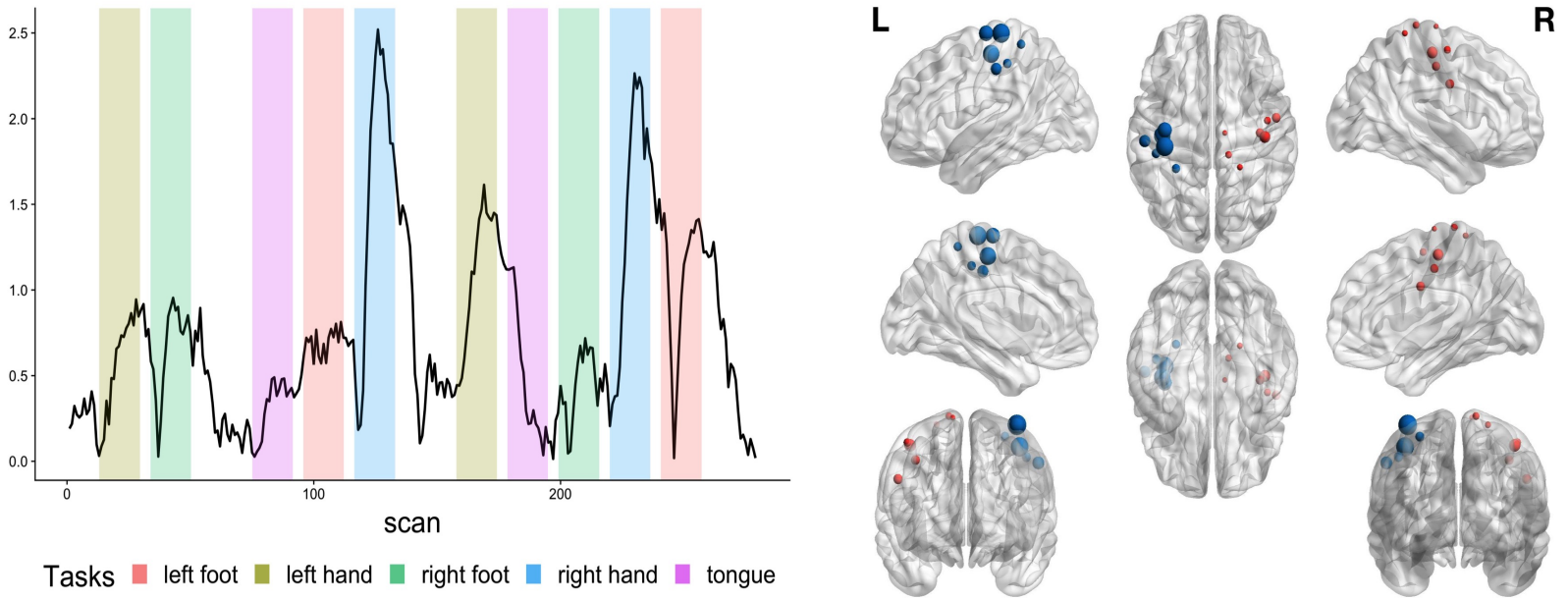


Figure 1: Average time course (left panel) and brain regions (right panel).

Future directions

1. For RCTs using stratified randomization and having time-to-event outcomes, an open question is how to derive the asymptotic distribution of covariate-adjusted estimators (Zhang, 2015, Lu and Tsiatis, 2011).

2. When identifying common brain networks, how to best handle correlation of data vectors remain future directions.

3. In analyses of RCTs, how to best pick the set of covariates to use in an adjusted estimator is a challenging problem.

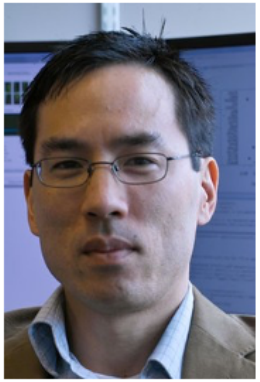
Acknowledgements

Michael Rosenblum

Brian Caffo

Acknowledgements

Ramin Mojtabai



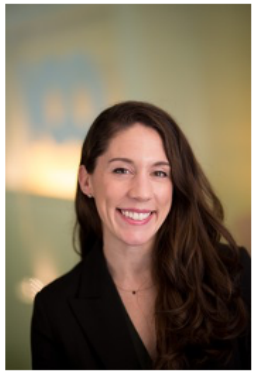
Bryan Lau

Brian Caffo

Xi (Rossi) Luo

Michael Rosenblum

Elizabeth Ogburn



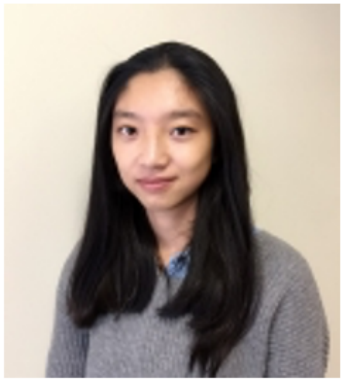
Alyssa Moran

Acknowledgements

Marcia Irene Canto

Elizabeth Ogburn

Xi (Rossi) Luo



Yi Zhao

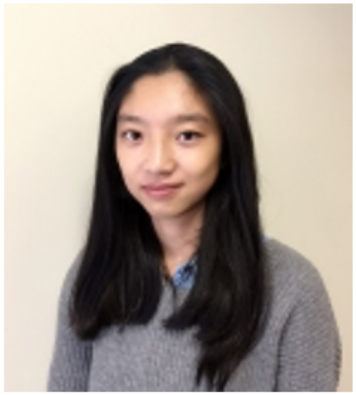
Masoumeh Amin-Esmaeili

Ryoko Susukida

Paniz Charkhchi

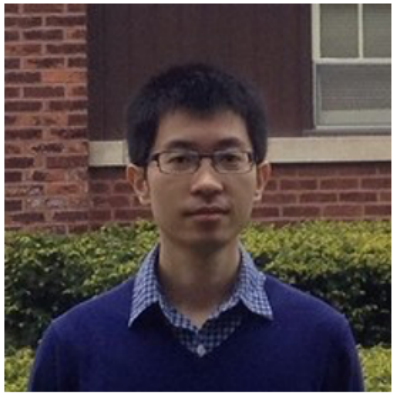
Acknowledgements

Marie Diener-West



Yi Zhao

Ning Yang



Dong Xi

Acknowledgements





Thank you!

The slides are available at
https://bingkaiwang.com.

The papers are available at
https://doi.org/10.1111/biom.13062,
https://arxiv.org/abs/1910.13954,
https://doi.org/10.1111/biom.13369.

The R code is available at
https://github.com/BingkaiWang.

Bugni, F. A., I. A. Canay, and A. M. Shaikh (2018). Inference under covariate-adaptive randomization. *Journal of the American Statistical Association 113*(524), 1784–1796.

Campbell, A. N., E. V. Nunes, A. G. Matthews, M. Stitzer, G. M. Miele, D. Polsky, E. Turrigiano, S. Walters, E. A. McClure, T. L. Kyle, A. Wahle, P. Van Veldhuisen, B. Goldman, D. Babcock, P. Q. Stabile, T. Winhusen, and U. E. Ghitza (2014). Internet-delivered treatment for substance abuse: A multisite randomized controlled trial. *American Journal of Psychiatry 171*(6), 683–690.

FDA (2019). Adjusting for Covariates in Randomized Clinical Trials for Drugs and Biologics with Continuous Outcomes. Draft Guidance for Industry. https://www.fda.gov/media/123801/download.

FDA (2020). COVID-19: Developing Drugs and Biological Products for Treatment or Prevention. *U.S. Food and Drug Administration: CDER*.

Kahan, B. C. and T. P. Morris (2012). Improper analysis of trials randomised using stratified blocks or minimisation. *Statistics in Medicine 31*(4), 328–340.

Li, X. and P. Ding (2020). Rerandomization and regression adjustment. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 82*(1), 241–268.

Lin, Y., M. Zhu, and Z. Su (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.

Lu, X. and A. A. Tsiatis (2011). Semiparametric estimation of treatment effect with time-lagged response in the presence of informative censoring. *Lifetime Data Analysis 17*(4), 566–593.

Ma, W., F. Hu, and L. Zhang (2015). Testing hypotheses of covariate-adaptive randomized clinical trials. *Journal of the American Statistical Association 110*(510), 669–680.

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

Ma, W., F. Tu, and H. Liu (2020). Regression analysis for covariate-adaptive randomization: A robust and efficient inference perspective. *arXiv https://arxiv.org/abs/2009.02287*.

Shao, J. and X. Yu (2013). Validity of tests under covariate-adaptive biased coin randomization and generalized linear models. *Biometrics 69*(4), 960–969.

Shao, J., X. Yu, and B. Zhong (2010). A theory for testing hypotheses under covariate-adaptive randomization. *Biometrika 97*(2), 347–360.

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

Yang, L., W. Ma, Y. Qin, and F. Hu (2020). Testing for treatment effect in covariate-adaptive randomized clinical trials with generalized linear models and omitted covariates. *arXiv https://arxiv.org/abs/2009.04136*.

Ye, T. and J. Shao (2020). Robust tests for treatment effect in survival analysis under covariate-adaptive randomization. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 82*(5), 1301–1323.

Ye, T., Y. Yi, and J. Shao (2020). Inference on average treatment effect under minimization and other covariate-adaptive randomization methods. *arXiv https://arxiv.org/abs/2007.09576*.

Zhang, M. (2015). Robust methods to improve efficiency and reduce bias in estimating survival curves in randomized clinical trials. *Lifetime Data Analysis 21*(1), 119–137.