Covariate Adjusted Randomization (CAR)

Analyzed With

Randomization-Based Inference (RBI)

Jonathan J. Chipman, PhD
May 22, 2024

University of Utah Intermountain Healthcare
Department of Population Health Sciences
Division of Biostatistics

Disclosures

No relevant disclosures.

REACH R01 Funding: NIDDK R01 DK100694

Chipman, J. J., Mayberry, L., and Greevy, R. A. J. (2023). Rematching on-the-fly: Sequential matched randomization and a case for covariate-adjusted randomization. *Statistics in medicine*, 42(22):3981–3995

FDA 2023 Guidance for covariate-adjustment in RCTs (FDA, 2023)

Estimand of interest: Population-level treatment effect

An unadjusted estimator is internally valid

- Unbiased
- Correct Type I error when modeling assumptions are met
- Could be more efficient by adjusting for covariates

Podcast Q and A: What are a couple of key items that you especially want listeners to remember?

I really want listeners to remember that **the FDA encourages covariate adjustment** because we believe that it is a **low hanging fruit** that can be used **to improve the efficiency** of a clinical trial analysis without creating additional burdens for sponsors. We encourage sponsors to discuss covariate adjustment with the FDA during the development of the protocol, particularly for situations not explicitly covered in the guidance. - Dan Rubin

CAR+RBI

Proactive covariate adjustment

Benefits

- Could increase efficiency
- Could reduce covariate imbalances
- Non-parametric, exact test
- Unadjusted estimator
  (Simple summary measure)

Considerations

- Finite population estimand
- Real-time implementation
- See also slides 20-22

Can CAR+RBI be as powerful as regression adjustment?

Randomization-Based Inference (RBI)

The randomization test minimally assumes

1. The treatment assignment is randomized
2. The potential outcomes values under each arm

Choose a Randomization Model

- Complete Randomization
- Restricted Randomization

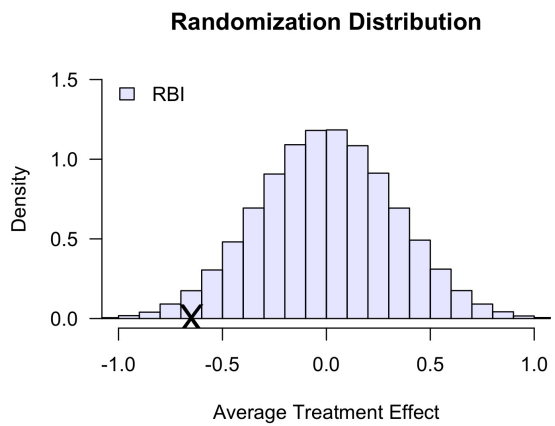
Obtain summary estimate

Compare to null distribution

- Summary measure across all possible randomization sequences

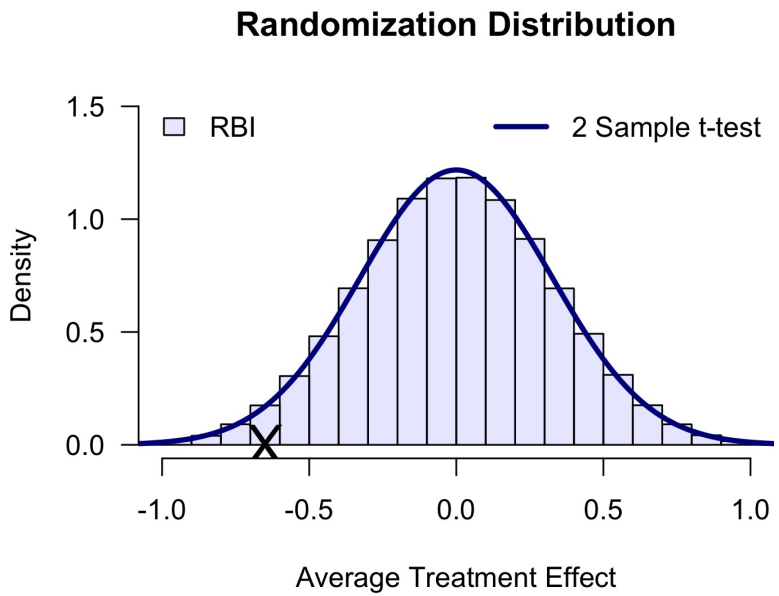
Pedagogical example

- 50 participants, equal allocation
- Observed outcomes $\sim$ N(0,1)
- Sharp null (i.e., $Y_i(0) = Y_i(1)$)



Randomization-Based Inference (RBI)

- 50 participants, equal allocation
- Observed outcomes ~ N(0,1)
- Sharp null (i.e., $Y_i(0) = Y_i(1)$)

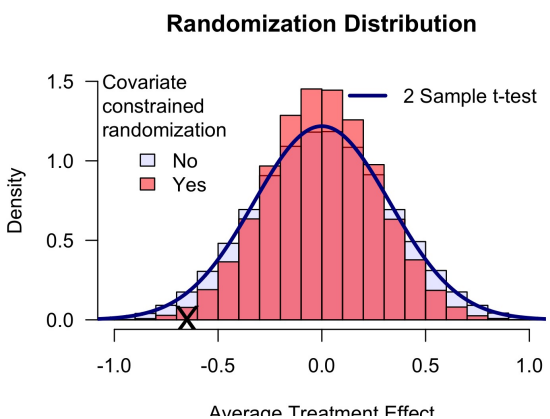


- T-Test: Super-population (no treatment assignment variability)
- RBI: Finite population (no sampling variability)

CAR+RBI

- 50 participants, equal allocation
- Observed outcomes ~ N(0,1)
- Sharp null (i.e., $Y_i(0) = Y_i(1)$)

- $X \sim N(0,1)$ baseline health status, $\rho_{X,Y} = 0.67$
- Constrain allowable randomization sequences
  - Standardized Mean Difference (SMD): $|(\overline{X}_T - \overline{X}_C)/SD(X)|$
  - Exclude randomizations resulting in SMD > 0.2 SD
  - Re-randomization (Morgan and Rubin, 2012)



After CAR

- 2-Sample T-Test is conservative
- Parametric inference requires correction/adjustment
- Or ... use RBI

Can CAR+RBI be as powerful as regression adjustment?

Context of linear models, power increases when ...

- Model specification: Closely approximating the relationship between the outcomes, treatment, and covariates
  - Y ~ TX + COV + TX*COV (Lin, 2013)
  - Can be complex to specify

- Treatment assignment: TX $\perp$ COV (no covariate imbalance) (Atkinson, 1982; Senn et al., 2010)

Conjecture: A good CAR with RBI can competitively capture regression adjustment efficiency

1. Can reduce chance imbalances
2. Averts model complexity

Selected CAR strategies

When X is not known for all participants before randomization

1. Stratified randomization
2. Minimization with a biased coin (Pocock, 1977)
3. Sequential Matched (and Rematched) randomization (Kapelner and Krieger, 2014; Chipman et al., 2023)
4. Sequential Re-randomization (Zhou et al., 2018)

Stratified Randomization

- Most commonly implemented covariate-adjusted randomization scheme (Sverdlov et al., 2023; McPherson et al., 2012)

- Randomize within categorized patient profiles
  - Continuous covariates must be categorized
  - Quickly limited by the number of adjusting covariates

Minimization with a biased coin

Weighted randomization to arm that reduces imbalance (Pocock and Simon, 1975)

Examples:

1. D_A Biased Coin Design (D_A-BCD) (Atkinson, 1982)

   Minimize standard error of ATE from pre-specified model

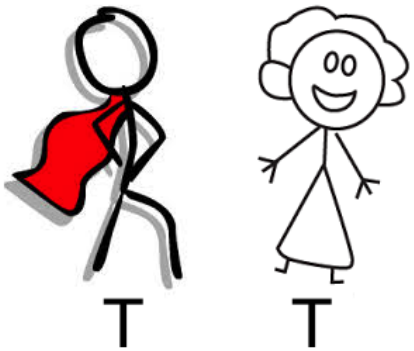
2. Pairwise Sequential Randomization (PSR) (Ma et al., 2020):

$$M = (\overline{x}_T - \overline{x}_C)' cov(\overline{x}_T - \overline{x}_C)^{-1} (\overline{x}_T - \overline{x}_C)$$

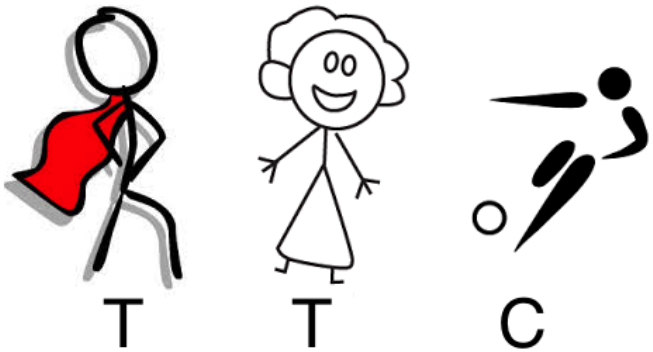
Sequential Matching (Kapelner and Krieger, 2014)

Sequential Matching (Kapelner and Krieger, 2014)

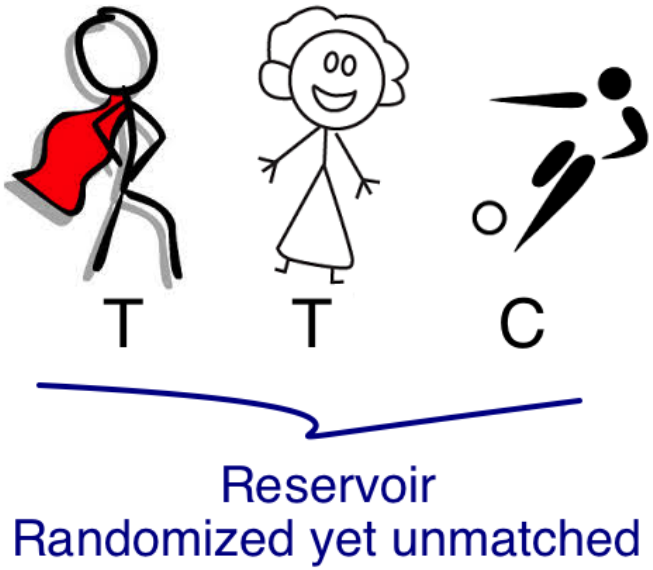
Sequential Matching (Kapelner and Krieger, 2014)



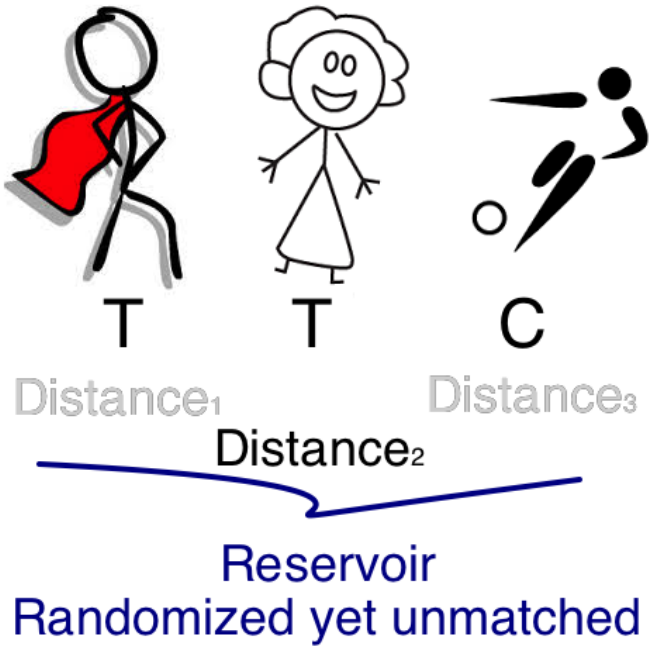
Sequential Matching (Kapelner and Krieger, 2014)



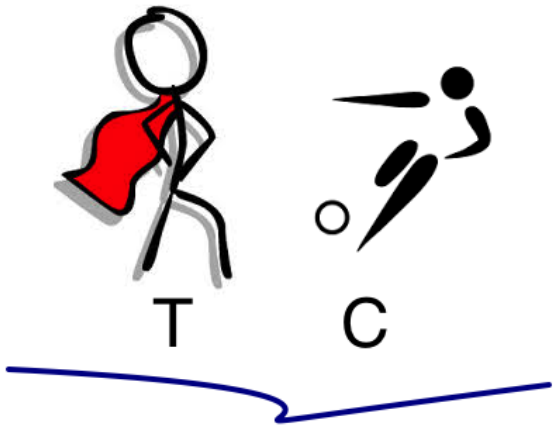
Sequential Matching (Kapelner and Krieger, 2014)



Sequential Matching (Kapelner and Krieger, 2014)



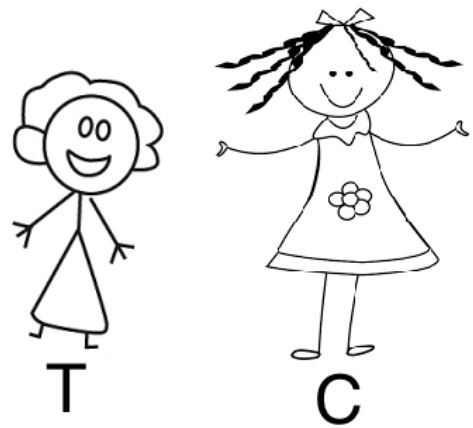
Sequential Matching (Kapelner and Krieger, 2014)



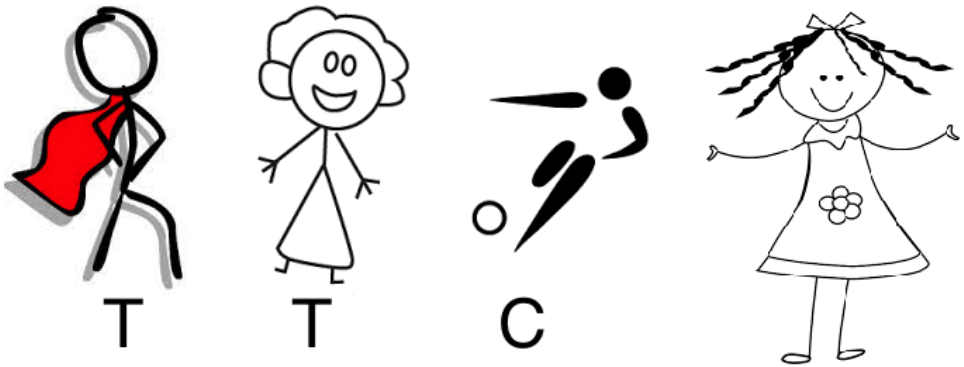
Reservoir
Randomized yet unmatched

If $\text{Distance}_2 < \text{Threshold}$

Matched Randomized



Sequential Matching (Kapelner and Krieger, 2014)



Reservoir
Randomized yet unmatched

If $\text{Distance}_2 > \text{Threshold}$

Sequential Matching, additional notes (Kapelner and Krieger, 2014)

Implementation details ....

1. Enrollment: One individual at a time (greedy matching)
2. Similarity: Prespecified fixed threshold*
3. Reservoir: Not required to deplete

* Often quantile of $F_{p,n-p}$ distribution though can be of an empirically estimated distribution

SMR(20, E): Sequential Matched Randomization with a 20^{th} percentile of empirically estimated null distribution

3 Extensions (Chipman et al., 2023)

Batch Enrollment: MR when single-batch (Greevy et al., 2004)

- Identify pairs based upon set of pairs with minimum total distance
- Reduces greedy matching

Dynamic Threshold: SMR(D,E)

- Chance of matching now versus later
- Can achieve equal allocation

Re-matching: SRR(D,E) Rematching with dynamic threshold

- Allow matches to break and re-match
- Patients keep original treatment assignment
- Must match to opposite treatment arm

Performance of extensions (Chipman et al., 2023)

In simplified settings and a real-world example, extensions improved:

1. Average balance in covariates between arms
2. Precision estimating average treatment effect
3. Average distance among matched pairs (i.e., "match quality")

CAR + RBI in real-world case study

REACH Trial: Causal questions, estimands, and design

Rapid **E**ducation/Encouragement **A**nd **C**ommunications for **H**ealth

Population	Adults with Type 2 Diabetes (DM)
Purpose	Increase glycemic control and adherence to medications
Main Intervention	Text message-delivered diabetes support for 12 months
Outcome	12 month glycemic control (A1c) compared to control
Multi-site enrollment	512 patients from Vanderbilt and Non-Vanderbilt Clinics

Key Baseline Covariates

Biological Factors

- Baseline A1c*
- Age at baseline
- Time since DM dx*
- DM type*
- Race / Ethnicity

Socio-economic Factors

- Yrs of education
- Income level
- Insurance type

* Greater priority for balancing ($R^2 = 0.26$ vs 0.32 for all covariates)

Questions of interest

1. Can CAR+RBI be as powerful as regression adjustment?

2. How sensitive are CAR schemes to the number of covariates?

Randomization models:

- Complete Randomization (CR)
- Covariate-Adjusted Randomization (CAR)
  - Stratified randomization
  - D_A-BCD (3/4 biased coin)
  - PSR (3/4 biased coin)
  - Sequential Matched Randomization and extensions

OLS-Adjustment: Linear adjustment for covariates (no interactions)

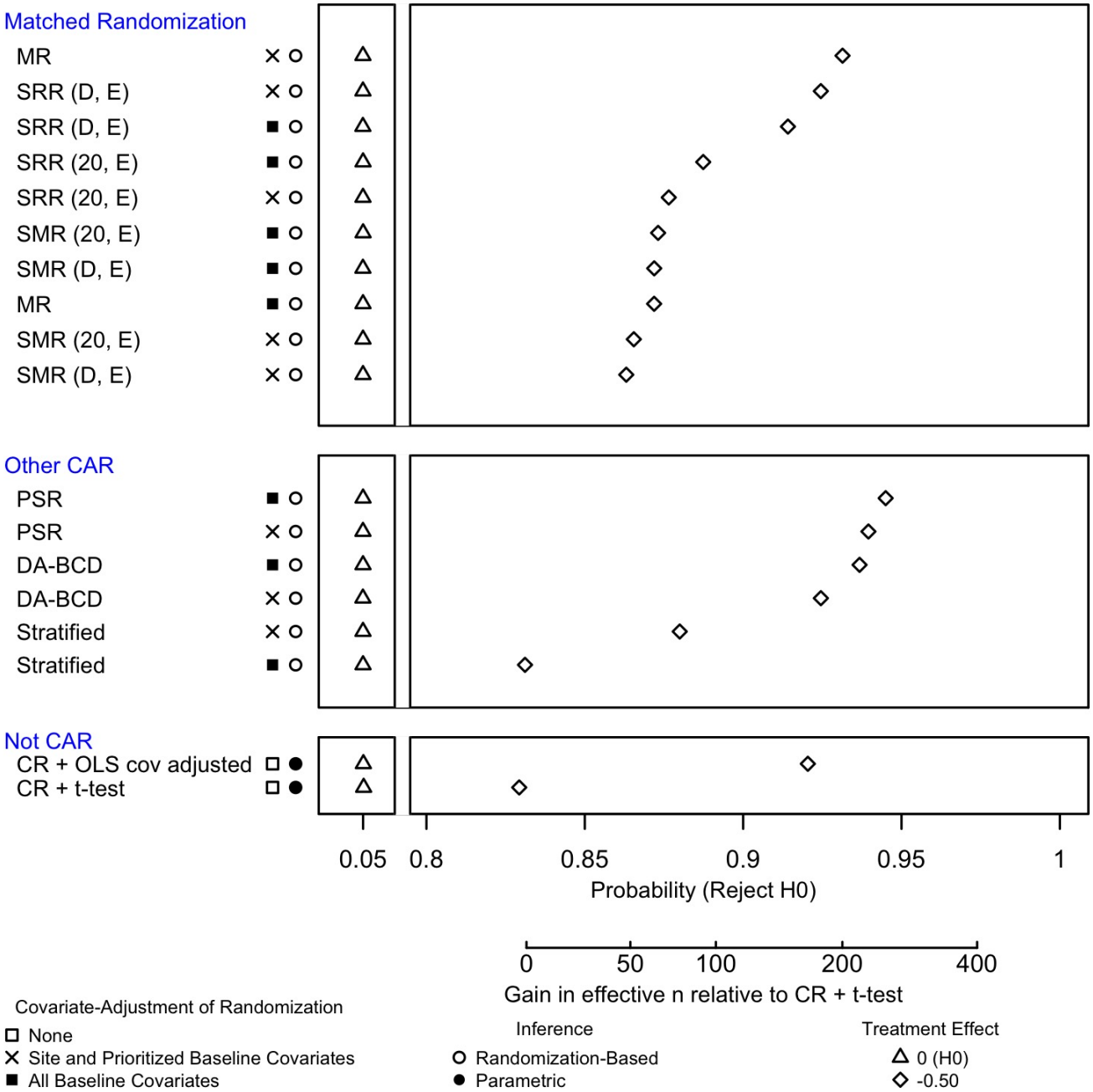
Simulation of REACH outcomes

For sharp treatment effects of 0 (null) and -0.5 (beneficial)

1. Conditioning on enrollment (i.e., randomness via trt assignment)
2. 100K Randomization sequences for each randomization procedure
3. For each sequence, record
   - Covariate balance: Average SMD across covariates
   - Study efficiency: Whether to reject H_0: ATE = 0
   - Match quality: Distance of matched pairs

Case-study caveats:

- Assumes variability due to observed outcomes
- Results may differ for each trial


Matched Randomization
MR
SRR (D, E)
SRR (D, E)
SRR (20, E)
SRR (20, E)
SMR (20, E)
SMR (D, E)
MR
SMR (20, E)
SMR (D, E)
Other CAR
PSR
PSR
DA-BCD
DA-BCD
Stratified
Stratified
Not CAR
CR + OLS cov adjusted
CR + t-test
0.05
0.8
0.85
0.9
0.95
1
Probability (Reject H0)
0
50
100
200
400
Gain in effective n relative to CR + t-test
Covariate-Adjustment of Randomization
None
Site and Prioritized Baseline Covariates
All Baseline Covariates
Inference
Randomization-Based
Parametric
Treatment Effect
0 (H0)
-0.50

REACH Trial: Case study

1. Can CAR+RBI be as powerful as regression adjustment?
   - SRR, PSR, and D_A-BCD were more powerful
   - SRR, PSR, and D_A-BCD with RBI increased effective n by ≥ 170 compared to CR with unadjusted test
2. How sensitive are CAR schemes to the number of covariates?
   - Stratification: Worsened when over-adjusted
   - Sequential Matching: Mixed impact
   - PSR and DA-BCD: improved with more covariates

Generalized code to reassess questions for any trial (Chipman et al., 2023)

Questions to consider

Covariate-adjustment with RBI:

- Can increase efficiency
- Can reduce covariate imbalances
- Can use an unadjusted estimator (simple summary measure)
- Is a non-parametric, exact test

What barriers/considerations remain for adopting this analytic strategy?

1. Reconciling with finite population estimand
2. Assumption of a 'sharp'-null hypothesis
3. Adoption into databases for real-time randomization
4. Modelling may be desirable for more complex questions
5. Data may have missing outcomes or covariates

Reconciling with finite population estimands

FDA Estimand of interest: Population-level treatment effect

RCT's carry internal validity but are generally limited in external validity:

- Participants are not a random sample of broader population

Viewpoint: A finite population estimand is consistent with the generalizability of the trial's participants.

Additional considerations

- FDA supports RBI and has approved treatments based upon RBI(FDA, 2023; FDA, 2017)
- A super-population model can be used Imbens and Menzel (2021a)

Questions to consider

What barriers/considerations remain for adopting this analytic strategy?

1. Reconciling with finite population estimands
   - FDA supports RBI and has approved treatments based upon RBI(FDA, 2023; FDA, 2017)
   - A super-population model can be used Imbens and Menzel (2021a)
   - Trial participants are often unlike the population of interest
2. Assumption of a 'sharp'-null hypothesis
   - A sharp-null can be relaxed (Imbens and Menzel, 2021b)
3. Adoption into databases for real-time randomization
   - Proof of principle in ECOG (Lange and MacIntyre, 1985)
4. Modelling may be desirable for more complex questions
   - (Shao et al., 2010; Kapelner and Krieger, 2014; Ma et al., 2020; Bannick et al., 2023)
5. Data may have missing outcomes or covariates
   - (Rubin, 1998; Ivanova et al., 2022; Heussen et al., 2023)

References I

Atkinson, A. C. (1982). Optimum Biased Coin Designs for Sequential Clinical Trials with Prognostic factors. *Biometrika*, 69(1):61.

Bannick, M. S., Shao, J., Liu, J., Du, Y., Yi, Y., and Ye, T. (2023). A General Form of Covariate Adjustment in Randomized Clinical Trials. *arXiv preprint arXiv:2306.10213*.

Chipman, J. J., Mayberry, L., and Greevy, R. A. J. (2023). Rematching on-the-fly: Sequential matched randomization and a case for covariate-adjusted randomization. *Statistics in medicine*, 42(22):3981–3995.

Greevy, R., Lu, B., Silber, J. H., and Rosenbaum, P. (2004). Optimal multivariate matching before randomization. *Biostatistics (Oxford, England)*.

References II

Heussen, N., Hilgers, R.-D., Rosenberger, W. F., Tan, X., and Uschner, D. (2023). Randomization-based inference for clinical trials with missing outcome data. *Statistics in Biopharmaceutical Research*, pages 1–12.

Imbens, G. and Menzel, K. (2021a). A causal bootstrap. *Annals of Statistics*, 49(3):1460–1488.

Imbens, G. and Menzel, K. (2021b). A causal bootstrap. *Annals of Statistics*, 49(3):1460–1488.

Ivanova, A., Lederman, S., Stark, P. B., Sullivan, G., and Vaughn, B. (2022). Randomization tests in clinical trials with multiple imputation for handling missing data. *Journal of Biopharmaceutical Statistics*, 32(3):441–449.

Kapelner, A. and Krieger, A. (2014). Matching on-the-fly: Sequential allocation with higher power and efficiency. *Biometrics*, 70(2):378–388.

References III

Lange, N. and MacIntyre, J. (1985). A computerized patient registration and treatment randomization system for multi-institutional clinical trials. *Controlled Clinical Trials*, 6(1):38–50.

Lin, W. (2013). Agnostic notes on regression adjustments to experimental data: Reexamining Freedman' s critique. *The Annals of Applied Statistics*, 7(1):295–318.

Ma, W., Qin, Y., Li, Y., and Hu, F. (2020). Statistical Inference for Covariate-Adaptive Randomization Procedures. *Journal of the American Statistical Association*, 115(531):1488–1497.

McPherson, G. C., Campbell, M. K., and Elbourne, D. R. (2012). Use of randomisation in clinical trials: a survey of UK practice. *Trials*, 13(1):1–7.

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

References IV

Pocock, S. J. (1977). Group sequential methods in the design and analysis of clinical trials. *Biometrika*, 64(2):191–199.

Pocock, S. J. and Simon, R. (1975). Sequential Treatment Assignment with Balancing for Prognostic Factors in the Controlled Clinical Trial. *Biometrics*, 31(1):103–115.

Rubin, D. B. (1998). More powerful randomization-based p-values in double-blind trials with non-compliance. *Statistics in Medicine*, 17(3):371–385.

Senn, S., Anisimov, V. V., and Fedorov, V. V. (2010). Comparisons of minimization and Atkinson's algorithm. *Statistics in medicine*, 29(7-8):721–730.

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

References V

Sverdlov, O., Carter, K., Hilgers, R.-D., Everett, C. C., Berger, V. W., Luo, Y. A., Chipman, J. J., Ryeznik, Y., Ross, J., Knight, R., and Yamada, K. (2023). Which Randomization Methods Are Used Most Frequently in Clinical Trials? Results of a Survey by the Randomization Working Group. *Statistics in Biopharmaceutical Research*, 0(ja):1–21.

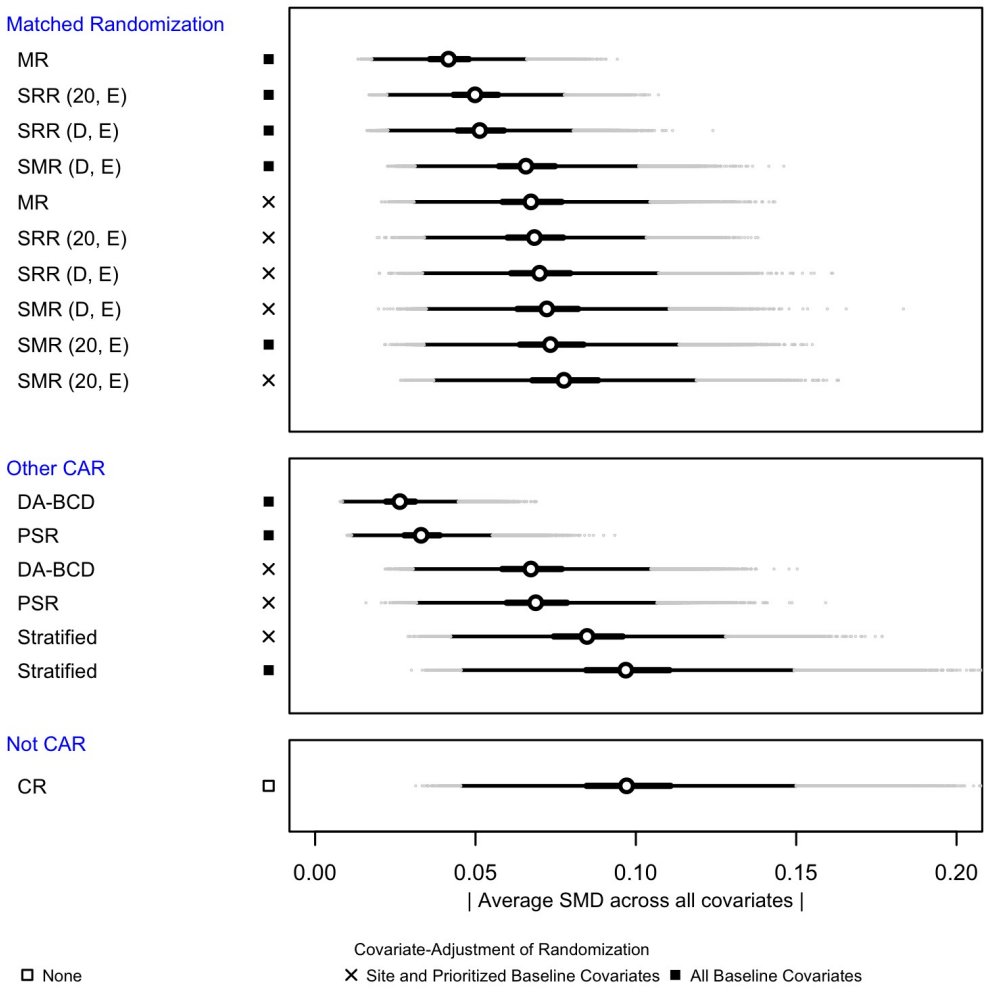
U.S. Department of Health and Human Services Food and Drug Administration (2023). Adjusting for Covariates in Randomized Clinical Trials for Drugs and Biological Products. *Guidance Document*, (May).

U.S. Food and Drug Administration (2017). Statistical Review and Evaluation of NDA/Serial Number: 125261 / 138. https://www.fda.gov/media/108953/download.

Zhou, Q., Ernst, P. A., Morgan, K. L., Rubin, D. B., and Zhang, A. (2018). Sequential rerandomization. *Biometrika*, 105(3):745–752.

Bonus

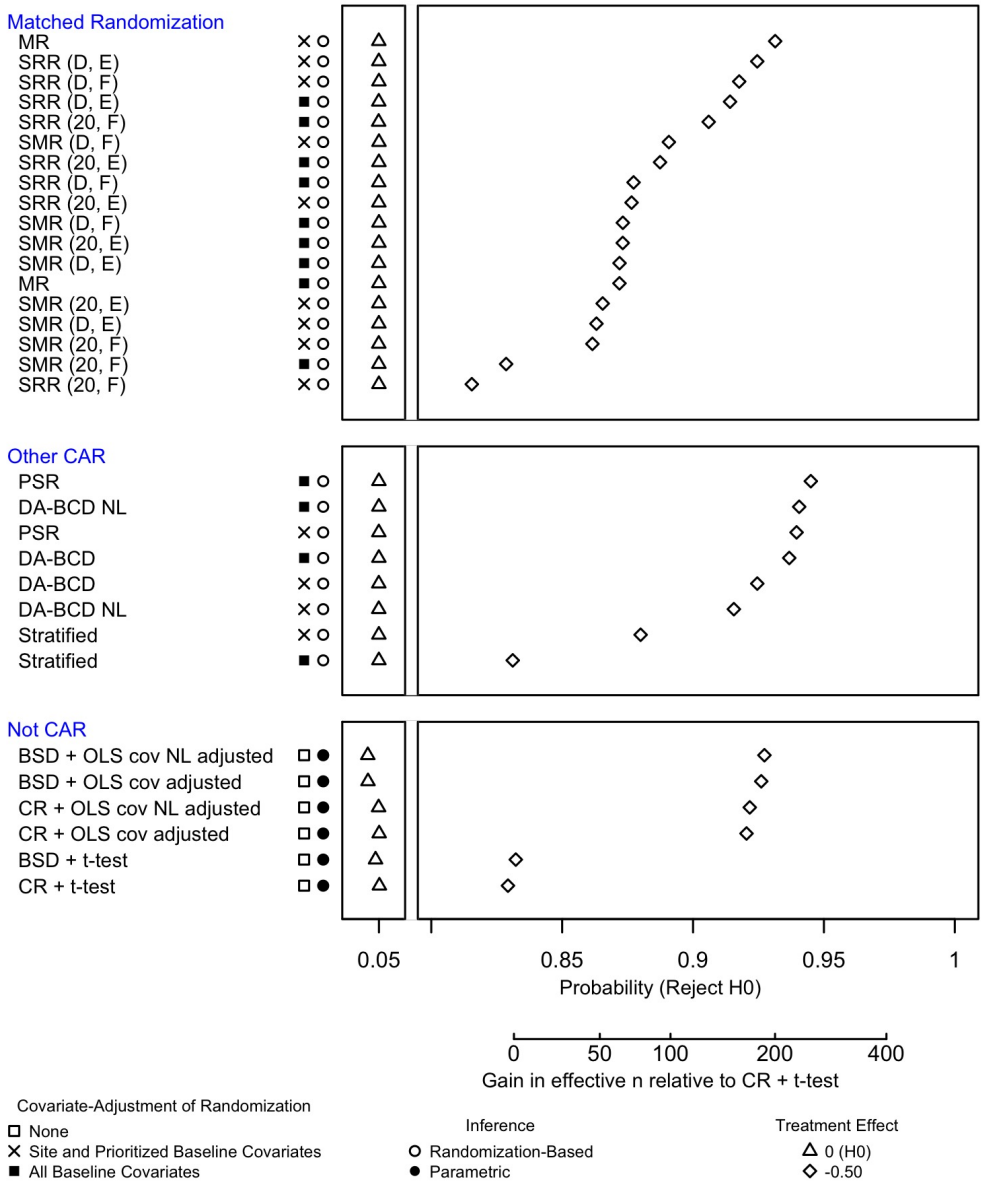
CAR to reduce chance covariate imbalances



Non-linear covariate adjustment

Notes on upcoming slide:

1. The OLS model was fit using splines on continuous covariates
2. The "Big Stick Design" (BSD) was used similar to random permuted blocks
3. The BSD with a finite population also requires adjustment to get the correct Type I error. (Here the design is not adjusted).


Matched Randomization
MR
SRR (D, E)
SRR (D, F)
SRR (D, E)
SRR (20, F)
SMR (D, F)
SRR (20, E)
SRR (D, F)
SRR (20, E)
SMR (D, F)
SMR (20, E)
SMR (D, E)
MR
SMR (20, E)
SMR (D, E)
SMR (20, F)
SMR (20, F)
SRR (20, F)
Other CAR
PSR
DA-BCD NL
PSR
DA-BCD
DA-BCD
DA-BCD NL
Stratified
Stratified
Not CAR
BSD + OLS cov NL adjusted
BSD + OLS cov adjusted
CR + OLS cov NL adjusted
CR + OLS cov adjusted
BSD + t-test
CR + t-test
0.05
0.85
0.9
0.95
1
Probability (Reject H0)
0
50
100
200
400
Gain in effective n relative to CR + t-test
Covariate-Adjustment of Randomization
None
Site and Prioritized Baseline Covariates
All Baseline Covariates
Inference
Randomization-Based
Parametric
Treatment Effect
0 (H0)
-0.50

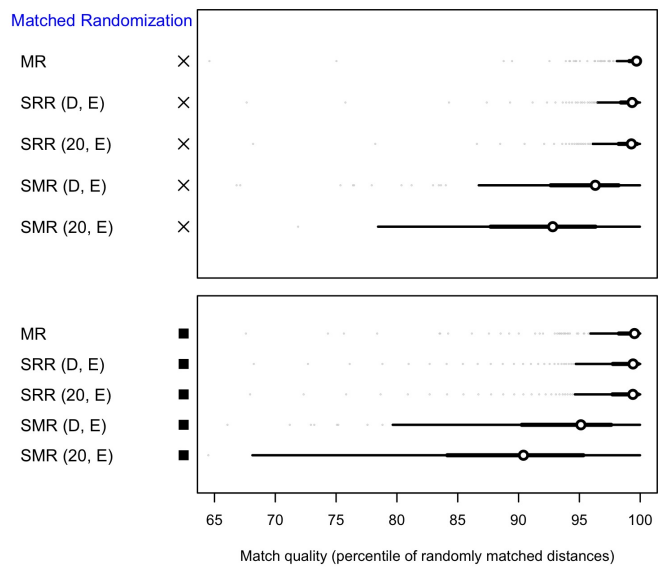
Additional case study observations

1. Sequential matching extensions improved overall covariate balance, power, and match quality
2. Sequential matching performed better in power when adjusting for priority covariates
3. Stratified randomization on priority covariates meaningful improved power

Where does sequentially matched randomization fit given alternatives?

In this case study, other CAR methods were superior in balancing covariates and increasing power. Note that PSR uses a block-two randomization scheme which reduces the risk of chronological imbalance yet can be predictable.

A unique aspect of matching is that the objective function (best match) is measured at the individual-level rather than group-level. It could be an avenue for personalized medicine trial designs. Sequential matching extensions improved the quality of matches.



Model-Robust Inference for Clinical Trials: Improving Precision by Stratified Randomization and Covariate Adjustment

Bingkai Wang, Assistant Professor, Department of Biostatistics

SCT 2024

Disclosure

No relevant disclosures.

Two aspects for statistical analysis of RCTs

Robustness

Valid inference when the model is wrong.

Precision

Increased power given a fixed sample size.

A motivating example: CTN44

Treatment:
Internet-delivered therapy for substance abuse

Stratified randomization by baseline substance use and abstinence status

Outcome:
Number of negative urine tests

Baseline variables: age, sex, etc.

Goal: Average treatment effect.

A motivating example: CTN44

Current practice for Inference: the ANCOVA estimator

- Regressing the outcome on treatment, randomization strata and other baseline variables.
- Model-based inference on the regression coefficients.

 Great under simple randomization!  How about stratified randomization?

Questions

1. (**Robustness**) Is the ANCOVA estimator robust under stratified randomization?
2. (**Precision**) Does stratified randomization make ANCOVA more precise?

Background

- Stratified randomization is used by 70% RCTs. (Lin et al. 2015).
- Less than 50% of trials that used stratified randomization adjusted for strata in their analyses (Kahan and Morris, 2012).
- For most commonly-used methods, little is known about their property under stratified randomization:
  - Binary outcomes: GLM
  - Missing data: doubly-robust estimation
  - longitudinal outcomes: MMRM (mixed models for repeated measures)
  - Time-to-event outcomes: Kaplan-Meier estimator

Our main results

For all commonly used estimators used in the primary analysis of RCTs, compared to simple randomization,

1) they remain valid/robust under stratified randomization;

2) they become equally or more precise under stratified randomization;

3) we showed how to fully leverage precision gain from stratified randomization.

The above results also hold for biased-coin randomization

Recent attention to this work

"There are several methods for computing standard errors when combining stratification with covariate adjustment and possible model misspecification (e.g., Wang et al. 2021)."

Examples

- **Continuous outcomes**: ANCOVA
- **Binary outcomes**: G-computation with logistic regression
- **Time-to-event outcomes**: Kaplan-Meier

The ANCOVA estimator

The ANCOVA estimator for ATE is $\widehat{\beta_A}$ in model

$$E[Y|A,X] = \beta_0 + \beta_A A + \beta_X^{\top} X.$$

Outcome — Treatment — Covariates

We assume covariates X include stratification variables.

The ANCOVA estimator

Asymptotic variance reduction by stratified randomization:

$$\frac{(1-2\pi)^2}{\pi(1-\pi)} E[Var\{Y(1)-Y(0)|S\}]$$

Anti-conservative inference procedure:

1. Compute point estimates and variance estimators as usual.
2. Correct variance by the above formula.
3. Hypothesis testing with the corrected variance.

The ANCOVA estimator

Variance correction is unnecessary under any of the following cases:

- 1:1 randomization ($\pi = 0.5$),
- the model is correctly specified,
- we add the treatment-by-covariate interaction terms into the model,

In the above cases, ANCOVA analysis based on simple randomization is robust, non-conservative, and can improve precision.

The ANCOVA estimator

What if ANCOVA does not adjust for stratification variables?

- Standard model-based inference is conservative.

- One needs to perform variance correction by

$$\frac{1}{\pi(1-\pi)} E[E\{(A-\pi)IF|S\}^2].$$

G-computation with logistic regression

Step 1: fit outcome model:

$$P(Y = 1|A, X) = logit^{-1}(\beta_0 + \beta_A A + \beta_X^{\top} X)$$

and get model-fit $\hat{P}(Y = 1|A, X)$.

Step 2: Estimate ATE by

$$\frac{1}{n}\sum_{i}^{n} \hat{P}(Y = 1|A = 1, X_i) - \hat{P}(Y = 1|A = 0, X_i)$$

Step 3: Sandwich variance estimation.

G-computation with logistic regression

Steps for non-conservative inference:

1. Compute point estimates and variance estimators as usual.
2. Correct asymptotic variance by $\frac{1}{\pi(1-\pi)} E[E\{(A-\pi)IF|S\}^2]$.
3. Hypothesis testing with the corrected variance.

G-computation with logistic regression

Variance correction is unnecessary under any of the following cases:

- 1:1 randomization ($\pi = 0.5$),
- the model is correctly specified,
- we add the treatment-by-covariate interaction terms into the model.

In the above cases, inference based on simple randomization is robust, non-conservative, and can improve precision.

The Kaplan-Meier estimator

The Kaplan-Meier estimator targets the survival curve for time-to-event outcomes.

It remains valid under stratified randomization.

Variance reduction by stratified randomization:

$$\frac{1}{\pi(1-\pi)} E\left[E\{(A-\pi)IF^a(t)|S\}^2\right]$$

How much variance reduction by stratified randomization?

Data example 1: variance reduction due to stratified randomization

- Trial: CTN44
- Time-to-event outcome: time to abstinence
- Group: 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	12%	12%	11%	8%	7%	4%

Data example 2: variance reduction due to covariate adjustment and stratified randomization

Study	Outcome type	Unadjusted estimator	Adjusted estimator	Proportional variance reduction
CTN03	Binary	-0.11(-0.21, -0.01)	-0.10(-0.19, -0.02)	35%
CTN30	Continuous	0.02(-0.02, 0.05)	0.01(-0.02, 0.04)	17%
CTN44	Continuous	-0.09(-0.14, -0.03)	-0.09(-0.14, -0.03)	2%

Take-away message

1. Stratified randomization retains valid and can improve precision.

2. When using stratified randomization, doing statistical inference with the correct variance can avoid being conservative.

3. Adjusting for a set of preplanned baseline variables can lead to substantial variance reduction.

A big shout

ASA-BIOP Covariate Adjustment working group

- Statisticians from academia, industry, and government;
- https://carswg.github.io/
- Monthly journal clubs, education materials, softwares …

Thank you!

The slides are available at

https://bingkaiwang.com.

The paper is available at

https://www.tandfonline.com/doi/full/10.1080/01621459.2021.1981338.

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

Email:

bingkai.w@gmail.com

References

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.A merican Journal ofPsychiatry 171(6), 683–690.

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 randomized using stratified blocks or minimisation. Stat. Med. 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.

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 ofcovariate-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.

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.

Covariate Adjustment in Randomized Clinical Trials with Missing Outcomes and Covariates

Rui Wang

Department of Population Medicine
Harvard Pilgrim Health Care Institute and Harvard Medical School

May 22, 2024

SCT Annual Meeting

No relevant disclosures

Chance imbalance in RCT

- Randomized controlled trials (RCT) provide high quality evidence for evaluating the efficacy and safety of new treatments
- Randomization ensures balance in both measured and unmeasured covariates in large samples
- Chance imbalance in baseline characteristics can happen in RCT, particularly in small studies
- Using regression to model outcome with covariates reduces variance when estimating treatment effect (Fisher, 1935)

Example: BestAIR trial

- Best Apnea Interventions for Research (BestAIR) trial: individual randomized, parallel-group trial
- Target population: patients with high cardiovascular disease risk and obstructive sleep apnea but without severe sleepiness
- Treatment: continuous positive airway pressure (CPAP)
- Goal: evaluate whether CPAP leads to better health outcomes
- Outcome: 24-hour systolic blood pressure (SBP) measured at six months
- 169 patients: 83 in treatment and 86 in control
- 45 patients missing outcomes; small amount of missingness in baseline SBP and other covariates

Chance imbalance in BestAIR trial

Zeng et al., 2021 SIM

	All patients $N = 169$	CPAP group $N_1 = 83$	Control group $N_0 = 86$	ASD^{UNADJ}
Baseline categorical covariates and number of units in each group				
Gender (male)	107	54	53	0.046
Race and ethnicity				
White	152	75	77	0.051
Black	11	5	6	0.060
Other	5	2	3	0.086
Recruiting center				
Site 1	54	26	28	0.046
Site 2	10	5	5	0.065
Site 3	105	52	53	0.073
Baseline continuous covariates, mean and standard deviation (in parenthesis)				
Age (years)	64.4 (7.4)	64.4 (8.0)	64.3 (6.8)	0.020
BMI (kg/m^2)	31.7 (6.0)	31.0 (5.3)	32.4 (6.5)	0.261
Baseline SBP (mm Hg)	124.3 (13.2)	121.6 (11.1)	127.0 (14.6)	0.477
Baseline SDP (beats/min)	63.1 (10.7)	63.0 (10.4)	63.2 (10.9)	0.020
Baseline AHI (events/h)	28.8 (15.4)	26.5 (13.0)	31.1 (17.2)	0.348
Baseline ESS	8.3 (4.5)	8.0 (4.5)	8.5 (4.6)	0.092

Covariate adjustment

- Two commonly-used methods
  - Outcome regression (Fisher 1935; Tsiatis et al. 2008; Lin 2013)
  - Propensity score weighting (Williamson et al. 2014; Shen et al. 2014; Zeng et al. 2021)
- Barrier: missing data in both outcomes and covariates

Setup

- A RCT of N units with two arms: N_1, N_0 patients in the treatment and control arm
- Z_i: binary treatment indicator ($z = 1$ treated; $z = 0$ control)
- Assuming the Stable Unit Treatment Values Assumption (SUTVA), each unit has two potential outcomes $\{Y_i(1), Y_i(0)\}$
- Observed outcome: $Y_i = Y_i(Z_i)$
- Baseline (pre-randomization) covariates: X_i (dimension J)
- Causal estimand: average treatment effect (ATE)

$$\tau = \mathbb{E}\{Y_i(1) - Y_i(0)\}$$

Unadjusted Estimator

- Under randomization: $\tau = \mathbb{E}(Y_i|Z_i = 1) - \mathbb{E}(Y_i|Z_i = 0)$
- Unadjusted (diff-in-means) estimator $\hat{\tau}_{\text{unadj}}$, unbiased for τ:

$$\hat{\tau}_{\text{unadj}} = \bar{Y}_1 - \bar{Y}_0 = \frac{\sum_{i=1}^{N} Z_i Y_i}{\sum_{i=1}^{N} Z_i} - \frac{\sum_{i=1}^{N} (1 - Z_i) Y_i}{\sum_{i=1}^{N} (1 - Z_i)}$$

- $\hat{\tau}_{\text{unadj}}$ is equivalent to coefficient of Z of linear regression

$$\texttt{lm}(Y_i \sim 1 + Z_i)$$

- Adjusting for prognostic covariates can reduce variance of $\hat{\tau}_{\text{unadj}}$ (Fisher, 1935)

Covariate adjustment method 1: outcome regression

- Interacted outcome regression (OLS) adjustment (X is centered):

$$\texttt{lm}(Y_i \sim 1 + Z_i + X_i + Z_i X_i)$$

  - OLS estimate of Z coefficient $\hat{\tau}_{\text{reg}}$: consistent for τ and asymptotically more efficient than $\hat{\tau}_{\text{unadj}}$ (Tsiatis et al., 2008; Lin, 2013)

Covariate adjustment method 2: prop score weighting

- Propensity score: $e(x) = \Pr(Z_i = 1 | X_i = x)$
- In RCT, propensity score is known and fixed for all units
- Covariate adjustment uses estimated propensity score $\widehat{e}_i = e(X_i; \widehat{\theta})$, obtained from a *working* model, e.g.

$$e_i = e(X_i; \theta) = \text{expit}(\theta_0 + X_i^T \theta_1)$$

- Estimated-propensity-score weighting estimator of τ

$$\hat{\tau}_{\text{ps}} = \frac{\sum_{i=1}^N \hat{\pi}_i Z_i Y_i}{\sum_{i=1}^N \hat{\pi}_i Z_i} - \frac{\sum_{i=1}^N \hat{\pi}_i (1 - Z_i) Y_i}{\sum_{i=1}^N \hat{\pi}_i (1 - Z_i)}$$

with inverse probability weights (IPW): $\hat{\pi}_i = \frac{Z_i}{\hat{e}_i} + \frac{1 - Z_i}{1 - \hat{e}_i}$

Covariate adjustment

- $\hat{\tau}_{\text{ps}}$: also a regression estimator, equivalent to coefficient of Z from a weighted least square (WLS) fit of

$$\texttt{lm}(Y_i \sim 1 + Z_i, \texttt{weights} = \hat{\pi}_i)$$

- OLS and IPW are asymptotically equivalent for covariate adjustment with fully observed data (Shen et al. 2014)
- Other weights, e.g. overlap weights can also be used, good finite-sample performance (Zeng et al. 2021; Chang et al. 2023)
- Weighting has a few finite-sample advantages
  - Avoid outcome model
  - Stable regardless of outcome types (e.g. rare outcomes)
  - OW leads to exact mean balance of covariates (Zeng et al. 2021)

Covariate adjustment with missing data

- Theory is clean with fully observed data, but...
- What to do with missing data in RCT?
- How to implement covariate adjustment?
- What is the added value by adjusting for partially missing covariates?

Covariate adjustment with missing outcomes

- Weight observed outcome by inverse of probability of being observed (Seaman and White, 2013)
  - R_i^y: indicator of outcome being observed, 1 observed, 0 missing
  - Estimate $p_i = \Pr(R_i^y = 1 \mid -)$, e.g. from a logistic model
  - Regression $\longrightarrow$ weighted regression (weights $1/\hat{p}_i$)

Covariate adjustment with or without missing data: three estimators

Table. Summary of $\{\hat{\tau}_{\text{unadj}}, \hat{\tau}_{\text{reg}}, \hat{\tau}_{\text{ps}}\}$ when data are fully observed (column 3) and when outcomes are partially observed (column 4).

	Regression specification	Weight over $i = 1, ..., n$	Weight over $\{i : R_i^Y = 1\}$
$\hat{\tau}_{\text{unadj}}$	$\texttt{lm}(Y_i \sim 1 + Z_i)$	1	$\hat{p}_i^{-1}$
$\hat{\tau}_{\text{reg}}$	$\texttt{lm}(Y_i \sim 1 + Z_i + X_i + Z_i X_i)$	1	$\hat{p}_i^{-1}$
$\hat{\tau}_{\text{ps}}$	$\texttt{lm}(Y_i \sim 1 + Z_i)$	$\hat{\pi}_i$	$\hat{p}_i^{-1}\hat{\pi}_i$

- $\hat{e}_i$: estimated PS from $\texttt{glm}(Z_i \sim 1 + x_i)$; $\hat{\pi}_i = \frac{Z_i}{\hat{e}_i} + \frac{1-Z_i}{1-\hat{e}_i}$
- $\hat{p}_i$: estimated prob of being observed (inversely weight $\hat{p}_i$ to adjust for missing outcome)

Missing data mechanism of outcome

- Assumption 1 (Complete randomization): $Z_i \perp\!\!\!\perp \{Y_i(1), Y_i(0), X_i\}$
- Assumption 2
  - i (Outcome is missing-at-random):
    $\Pr\{R_i^Y = 1 \mid Y_i(1), Y_i(0), X_i, Z_i\} = \Pr(R_i^Y = 1 \mid X_i, Z_i)$
  - ii (Logistic outcome missingness model):
    $p_i = \Pr(R_i^Y = 1 \mid X_i, Z_i) = \{1 + \exp(\{X_i, Z_i\}^{\mathrm{T}}\beta)\}^{-1}$,
- Assumption 2(ii) is NOT crucial, the authors conjecture main results also apply to other functional forms of the missingness model
- Also assume the outcome missingness model is correctly specified and p_i is consistently estimated

Asymptotic properties with missing Y

Zhao et al. (2024, Theorem 1) shows

- Consistency and asymptotic normality of $\{\hat{\tau}_{\text{unadj}}, \hat{\tau}_{\text{reg}}, \hat{\tau}_{\text{ps}}\}$ with closed-form variance
- In the presence of missing outcomes:
  - PSW guarantees efficiency gain compared to the unadjusted estimator ($\nu_{ps} \leq \nu_{unadj}$)
  - Interacted OLS no longer guarantees efficiency gain compared to the unadjusted estimator except for two special cases—MCAR or linear outcome model—where it is no worse than PSW ($\nu_{reg} \leq \nu_{ps} \leq \nu_{unadj}$)

Covariate adjustment with missing covariates

Zhao and Ding (2024)

- What if there is missingness in X?
- X: fully observed covariates
- $W_1, ..., W_K$: K partially observed covariates
- R^w: corresponding indicators of covariate being observed
- $W_i R^w$: effectively impute all missing values in W_i by zero, denoted as W_i^0
- Key insights of Zhao and Ding (2024): crucial to add R^w and W_i^0

Covariate adjustment with missing covariates

Zhao and Ding (2024)

- Vector of augmented fully observed covariates under the missingness indicator method: $X_i^{\text{mim}} = (X_i, R_i^w, W_i^0)$
- A modified missingness indicator method:

$$\texttt{lm}(Y_i \sim 1 + (X_i^{\text{mim}} - \bar{X}^{\text{mim}}) + Z_i + (X_i^{\text{mim}} - \bar{X}^{\text{mim}})Z_i) \qquad (1)$$

- OLS estimate of coef of Z is asymptotically more efficient than $\widehat{\tau}^{\text{unadj}}$, regardless of missing data mechanism of X
- Choice of imputation method and validity of imputation model does not matter using Model (1)

Cov adj with missing covariates and outcomes

- In practice, commonly missing data in both X and Y
- Question: can we combine the missing indicator method for missing X with the IPW estimators for missing Y?
- Answer: Yes. Zhao et al. (2024, Theorem 2) shows
  - the asymptotic properties with missing Y still holds when X is replaced by X^{mim}
  - Adding more covariates does not harm efficiency of PSW estimators

Cov adj with missing covariates and outcomes

- Assumptions
  - Assumption 1' (Complete randomization):
    $Z_i \perp\!\!\!\perp \{Y_i(1), Y_i(0), X_i, W_i, R_i^w\}$
  - Assumption 2' (Outcome MAR with logistic missingness model).
    Assumption 2 with X_i replaced by X_i^{mim}
- Renew
  $e_i = \Pr(Z_i = 1 \mid X_i^{X\text{mim}}) = e$; $p_i = \Pr(R_i^Y = 1 \mid X_i^{\text{mim}}, Z_i)$
- PSW can use IPW or overlap weighting (OW)
  - OW has an exact mean balance property, advantagous in small samples

Unified algorithm for cov adj with missing X, Y

Chang et al. (2023); Zhao et al. (2024)

S1. Impute all missing X values by zeros and obtain X^{mim}

S2. Estimate the outcome response probability for each unit $\widehat{p}_i$, e.g. via $\texttt{glm}(R_i^Y \sim 1 + X_i^{\text{mim}} + Z_i + X_i^{\text{mim}} Z_i)$

S3 When the linear outcome model or outcome MCAR is

S3.1 regression adjustment, i.e. WLS fit of $\texttt{lm}\{Y_i \sim 1 + Z_i + (X_i^{\text{mim}} - \bar{X}^{\text{mim}}) + (X_i^{\text{mim}} - \bar{X}^{\text{mim}})Z_i\}$ over units with observed outcome, with weights $\widehat{p}_i^{-1}$

S3.2 weighting adjustment, i.e, WLS fit of $\texttt{lm}(Y_i \sim 1 + Z_i)$ over units with observed outcome, with weights $\widehat{p}_i^{-1}\{Z_i/\widehat{e}_i + (1 - Z_i)/(1 - \widehat{e}_i)\}$, and $\widehat{e}_i$ estimated, e.g. via $\texttt{glm}(Z_i \sim 1 + X_i^{\text{mim}})$

- Coefficient of Z_i in the WLS in Step 3 is the covariate adjusted estimator of τ

Application: BestAIR Trial

- Adjust for four covariates: (i) fully observed: age, gender, baseline Apnea-Hypopnea Index, (ii) partially observed: baseline SBP

	$\hat{\tau}_{\text{unadj}}(x_i^{\text{mim}})$	$\hat{\tau}_{\text{reg}}(x_i^{\text{mim}})$	$\hat{\tau}_{\text{ps}}(x_i^{\text{mim}})$	$\hat{\tau}_{\text{unadj}}(x_i)$	$\hat{\tau}_{\text{reg}}(x_i)$	$\hat{\tau}_{\text{ps}}(x_i)$
$\widehat{\tau}$	-7.03	-6.07	-5.47	-4.78	-4.66	-4.60
$\text{var}(\widehat{\tau})$	6.26	7.50	4.72	5.69	5.85	5.63

Summary

- Provide a simple unified algorithm for covariate adjustment in RCT with missing outcomes and covariates
- Avoid the need for complex imputation model or multiple imputation
- Key: adding missingness indicators of partially observed covariates
- Cluster randomized trials (Wang et al., 2024)
  - Doubly-robust estimator when outcomes are missing at random
  - Considered settings where the nuisance parameters are estimated by machine learning algorithms
  - Sensitivity analyses for missing data assumptions
- Software packages (Stata and R)

Acknowledgements

- National Institute of Allergy and Infectious Diseases R01 AI136947
- Patient-Centered Outcomes Research Institute (PCORI) contract ME-2019C1-16146
- Fan Li, Anqi Zhao (Duke)
- Peng Ding (UC Berkeley)
- Jerry Chang, Yue Song (Harvard)
- Fan Li (Yale)
- Bingkai Wang (Umich)

Key References

- C-R Chang, Y Song, F Li, R Wang. Covariate adjustment in randomized experiments with incomplete covariate and outcome data. Statistics in Medicine. 42(22), 3919-3935. 2023.
- A Zhao, P Ding, F Li. Covariate adjustment of randomized experiments with missing outcomes and covariates. 2024. Biometrika, forthcoming.
- Wang, B., Li, F., Wang, R. (2024). Handling incomplete outcomes and covariates in cluster-randomized trials: doubly-robust estimation, efficiency considerations, and sensitivity analysis. arXiv preprint arXiv:2401.11278.
- W Lin. Agnostic notes on regression adjustments to experimental data: Reexamining Freedman's critique. Ann Applied Stat, 7(1):295-318, 2013.
- EJ Williamson, A Forbes, IR White. Variance reduction in randomised trials by inverse probability weighting using the propensity score. Statistics in Medicine, 33(5):721-737, 2014.
- SR Seaman, IR White. Review of inverse probability weighting for dealing with missing data. Statistical methods in medical research, 22(3), 278-295, 2013
- C Shen, X Li, L Li. Inverse probability weighting for covariate adjustment in randomized studies. Statistics in Medicine, 33, 555-568, 2014.
- S Zeng, F Li, R Wang, F Li.Propensity score weighting for covariate adjustment in randomized clinical trials. Statistics in medicine, 40(4):842-858, 2021.
- A Zhao, P Ding. To adjust or not to adjust? estimating the average treatment effect in randomized experiments with missing covariates. JASA, 119, 450-460, 2024

Covariate adjustment in randomized trials: discussion

Robin Dunn

Society for Clinical Trials 2024, Boston, MA

May 22, 2024

NOVARTIS | Reimagining Medicine

Disclaimer

No relevant disclosures

Key questions

Covariate adjustment has received much recent regulatory, pharmaceutical, and academic interest as a way to increase the efficiency of clinical trials.

The papers in this session cover important questions on covariate adjustment in randomized trials:

- How can we use matching to adjust for covariates in the trial design, and how well do matching methods perform relative to other covariate-adjusted randomization approaches? (Chipman et al., 2023)

- In stratified randomized trials, how can we estimate the variance of covariate-adjusted binary, continuous, and time-to-event treatment effects in a way that is consistent and model-robust? (Wang et al., 2023)

- When outcome and/or covariate data are partially missing, how can we estimate covariate-adjusted treatment effects while guaranteeing efficiency gains over an unadjusted analysis? (Chang et al., 2023; Zhao et al., 2023)

Presentation 1

Covariate adjusted randomization with randomization-based inference

Presenter: Dr. Jonathan Chipman

Chipman, J.J., Mayberry, L. and Greevy Jr, R.A., 2023. Rematching on-the-fly: Sequential matched randomization and a case for covariate-adjusted randomization. *Statistics in Medicine,* 42(22):3981–3995.

Covariate adjusted randomization

Main idea

Setting: In a trial with sequential enrollment, we would like to randomize within pairs of matched participants to target covariate balance across treatment groups.

Impact: These covariate-adaptive randomization methods can lead to more efficient treatment effect estimation, relative to complete randomization.

Key points:

- Sequential (re)matched randomization can account for covariates at the randomization stage.
- Dynamic similarity threshold to determine cutoffs for similarity of matches.
- Randomization-based inference to calculate p-values.

Covariate adjusted randomization
Connections to clinical trial practice

What do regulatory guidances say about covariate-adjusted randomization?

- In general, there is much discussion of stratified randomization but less discussion/recommendation of other covariate-dependent randomization methods.
- FDA guidance on adaptive designs (2019): "Minimization" (Pocock and Simon, 1975) discussed as a way to assign treatment to consecutive enrollees while balancing baseline covariate distributions.
  Control Type I error with appropriate methodologies such as randomization or permutation tests.
- EMA guidance on covariate adjustment (2015): Recognizes limitation of stratification for more than a few variables. Dynamic allocation methods try to minimize covariate imbalance.
  Some concerns: no guarantees of balance, standard statistical approaches may not control Type I error.
- FDA guidance on covariate adjustment (2023): "Sponsors should discuss proposals for complex covariate-adaptive randomization, data-adaptive covariate selection, or use of covariate adjustment in an adaptive design with the relevant review division."

Covariate adjusted randomization
Connections to clinical trial practice

What are some examples of randomized trials that used matched randomization?

- Matching in cluster-randomized trials (Balzer et al., 2015):
  - Effect of community support program on maternal depression six months after birth (Watson et al., 2004).
    16 local government areas matched 1:1 on birth numbers, geographic size, and ratings of community infrastructure.
  - Effect of clinic access and health education on HIV incidence (Grosskurth et al., 1995).
    12 communities matched 1:1 on variables such as geography and health center utilization.
  - Effect of multicomponent interventions (e.g., outdoor areas, transportation) on average daily physical activity of 11- to 15-year-olds (Toftager et al., 2011).
    14 school districts matched 1:1 on variables such as school district urbanity and condition of school outdoor areas.

- Matching individuals in RCTs (Welsh et al., 2023):
  - Effect of sanocrysin as a treatment for pulmonary tuberculosis (Amberson et al., 1931).
    24 people with tuberculosis matched 1:1 on clinical X-ray and laboratory variables.
  - Effect of continuous insulin infusion as a treatment for insulin-dependent diabetes (Feldt-Rasmussen et al., 1986).
    36 people with diabetes matched 1:1 on urinary albumin level, blood glycosylated hemoglobin level, and sex.

Covariate adjusted randomization
Remaining questions

- How could you adjust for covariates in analysis on top of sequential rematched randomization?
  - Would you adjust for all covariates that you used in matched randomization?
  - Would you recommend using randomization based inference?
- If you incorporate a large number of covariates into the distance metric used for matching, can the treatment effect estimates become unstable?
- Could we apply the supercovariate ideas from the covariate adjustment literature to sequential matched/rematched randomization?
  - A supercovariate is a combination of covariates that predicts patient outcomes in the control group (e.g., a flexible prognostic model $\hat{f}(\boldsymbol{x})$ fitted on external data).
  - "Supercovariates" have garnered recent attention in the statistical (Holzhauer and Adewuyi, 2023) and regulatory literature (EMA 2022 qualification opinion on PROCOVA™).
  - Could matching on supercovariates provide a benefit over matching on only "priority covariates" or matching on all baseline covariates?

Presentation 2

Model-robust inference for clinical trials: Improving precision by stratified randomization and covariate adjustment

Presenter: Dr. Bingkai Wang

Wang, B., Susukida, R., Mojtabai, R., Amin-Esmaeili, M. and Rosenblum, M., 2023. Model-robust inference for clinical trials that improve precision by stratified randomization and covariate adjustment. *Journal of the American Statistical Association*, 118(542), pp.1152-1163.

Model-robust inference for stratified randomization

Main idea

Setting: Randomized controlled trial with stratified or biased coin randomization and with binary, continuous, or time-to-event outcomes.

Impact: Under stratified or biased coin randomization, this work establishes the consistency and asymptotic normality (CAN) as well as the asymptotic variance of a large class of treatment effect estimators.

Key points:

- Prevalence of stratified randomization but inconsistency in adjusting for stratification vars.
- Models in scope:
  - M estimators consistent under simple randomization (e.g., standardized logistic regression, MMRM)
  - Kaplan-Meier estimator of survival curve.

Model-robust inference for stratified randomization

Connections to clinical trial practice

- **What do regulatory guidances say about stratified randomization?**
  - ICH E9 (1998): Stratify on centers (multicenter trials) and potentially other important baseline characteristics. Account for stratification factors in analysis.
  - EMA guidance on covariate adjustment (2015):
    - Recommend stratification for covariate balance (incl. stratifying on center for multicenter trials).
      Why? Efficiency, study credibility, may expect different treatment effects within subgroups.
    - Include stratification variables in primary outcome model (unless stratifying only for administrative reasons).
  - FDA guidance on covariate adjustment (2023): Generally adjust for strata and potentially other important non-stratification variables.
- **What do regulatory guidances say about variance estimation under stratified randomization?**
  - FDA guidance on covariate adjustment (2023): Analyses that ignore stratification will be conservative. Recommend a few methods, including Bugni et al. (2018), Ye et al. (2022), and **Wang et al. (2023)**.

Model-robust inference for stratified randomization
Remaining questions

- Would you consider adjusting your sample size based on the expected gain in precision from stratification?
- How many stratification variables would you use? It is common to stratify on 1 or 2 variables, but would you consider using more variables in a large study?
- If you stratify on a discretized version of a continuous variable, would you adjust for the strata or the continuous variable?
- In your analyses of the substance use disorder trials, proper adjustment for the stratification factors and other variables leads to …
  - Major efficiency gains in a binary outcome (36% variance reduction relative to unadjusted analysis)
  - Small to moderate efficiency gains in the continuous outcomes (3% and 17% variance reductions)
  - Small to moderate efficiency gains in the Kaplan-Meier estimator (1% to 12% at various time points).

  Would you expect a stronger effect from stratification and covariate adjustment for certain outcome types?

Presentation 3

Covariate adjustment in randomized clinical trials with missing outcomes and covariates

Presenter: Dr. Rui Wang

Chang, C.R., Song, Y., Li, F. and Wang, R., 2023. Covariate adjustment in randomized clinical trials with missing covariate and outcome data. *Statistics in Medicine*, 42(22), pp.3919-3935.

Zhao, A., Ding, P. and Li, F., 2024. Covariate adjustment in randomized experiments with missing outcomes and covariates. *Biometrika*, p.asae017.

Covariate adjustment with missing outcomes and covariates

Main idea

Setting: Randomized controlled trial with missing outcome and/or covariate data.

Impact: The proposed missing outcome and covariate model provides a consistent and asymptotically normal estimator of the average treatment effect.

Key points:

- With only missing outcomes, propensity score-adjusted approach more efficient than unadjusted approach.
- With missing outcomes and covariates, modified propensity score-adjusted approach also more efficient than unadjusted approach.

Covariate adjustment with missing outcomes and covariates

Connections to clinical trial practice

Missing outcome data is often an estimand question. Reasons for missingness and the question of interest will guide strategy. The proposed methods are relevant when the outcomes are missing at random conditional on the observables.

National Research Council's 2010 report on The Prevention and Treatment of Missing Data in Clinical Trials (referenced in FDA's 2023 covariate adjustment guidance) provides six principles to consider for addressing missing data:

1. Is missingness meaningful?
2. What is the causal estimand of interest?
3. Document reasons for missingness.
4. Be transparent about assumptions on missingness.
5. Under the stated assumptions, the analysis should be statistically valid (e.g., consistent, with SE that takes into account uncertainty from sampling randomness and from missing outcomes).
6. Assess robustness (e.g., to assumptions from (4)) via sensitivity analysis.

ICH E9 (1998) and ICH E9 R1 (2021):

- Handling of missing data depends on the reason for missingness and the estimand strategy. E.g., missing data due to administrative censoring versus missing data from discontinuation of treatment due to lack of efficacy.
- Sensible methods for handling missing data should be pre-specified in the protocol.

Covariate adjustment with missing outcomes and covariates

Remaining questions

- Could we modify the missing outcome methods to use repeated outcome measurements?
  For instance, if someone is missing a week 20 outcome (primary outcome), an available week 16 outcome could provide relevant information.
- If the reason for missingness depends on post-baseline covariates or outcomes (e.g., drop-out due to lack of efficacy), could we still apply the missing outcome methods? What would we need to modify?
- Multiple imputation is a common approach for addressing missing baseline covariates.
  How does multiple imputation compare to the proposed missingness indicator method?
- Missingness percentage:
  - Could the method fail to reach nominal coverage if a sufficiently high percentage of covariates are missing?
  - Even if coverage is not an issue, how might you communicate that you are confident in the answer despite a high percentage of missing covariates or outcomes?

References

Session papers

Chang, C.R., Song, Y., Li, F. and Wang, R., 2023. Covariate adjustment in randomized clinical trials with missing covariate and outcome data. *Statistics in Medicine*, 42(22), pp.3919-3935.

Chipman, J.J., Mayberry, L. and Greevy Jr, R.A., 2023. Rematching on-the-fly: Sequential matched randomization and a case for covariate-adjusted randomization. *Statistics in Medicine*, 42(22), pp.3981-3995.

Wang, B., Susukida, R., Mojtabai, R., Amin-Esmaeili, M. and Rosenblum, M., 2023. Model-robust inference for clinical trials that improve precision by stratified randomization and covariate adjustment. *Journal of the American Statistical Association*, 118(542), pp.1152-1163.

Zhao, A., Ding, P. and Li, F., 2024. Covariate adjustment in randomized experiments with missing outcomes and covariates. *Biometrika*, p.asae017.

Regulatory guidance

European Medicines Agency, 2015. Guideline on adjustment for baseline covariates in clinical trials.

European Medicines Agency, 2022. Qualification opinion for Prognostic Covariate Adjustment (PROCOVA™).

International Council for Harmonisation, 1998. ICH E9: Statistical principles for clinical trials.

International Council for Harmonisation, 2021. ICH E9(R1) Statistical principles for clinical trials. Addendum: Estimands and sensitivity analysis in clinical trials.

National Research Council, 2010. The Prevention and Treatment of Missing Data in Clinical Trials. Ch 4: Drawing Inferences from Incomplete Data. Washington, DC: The National Academies Press. https://doi.org/10.17226/12955.

U.S. Food and Drug Administration, 2019. Adaptive designs for clinical trials of drugs and biologics.

U.S. Food and Drug Administration, 2023. Adjusting for covariates in randomized clinical trials for drugs and biological products.

References

Additional papers

Amberson Jr, J.B., McMahon, B.T. and Pinner, M., 1931. A clinical trial of sanocrysin in pulmonary tuberculosis. *American Review of Tuberculosis*, 24(4), pp.401-435.

Balzer, L.B., Petersen, M.L., van der Laan, M.J. and Search Consortium, 2015. Adaptive pair-matching in randomized trials with unbiased and efficient effect estimation. *Statistics in Medicine*, 34(6), pp.999-1011.

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

Feldt-Rasmussen, B., Mathiesen, E.R., Hegedüs, L. and Deckert, T., 1986. Kidney function during 12 months of strict metabolic control in insulin-dependent diabetic patients with incipient nephropathy. *New England Journal of Medicine*, 314(11), pp.665-670.

Grosskurth, H., Todd, J., Mwijarubi, E., Mayaud, P., Nicoll, A., Newell, J., Mabey, D., Hayes, R., Mosha, F., Senkoro, K. and Changalucha, J., 1995. Impact of improved treatment of sexually transmitted diseases on HIV infection in rural Tanzania: randomised controlled trial. *The Lancet*, 346(8974), pp.530-536.

Holzhauer, B. and Adewuyi, E.T., 2023. “Super-covariates”: Using predicted control group outcome as a covariate in randomized clinical trials. *Pharmaceutical Statistics*, 22(6), pp.1062-1075.

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

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, pp.21 – 25.

Pocock, S.J. and Simon, R., 1975. Sequential treatment assignment with balancing for prognostic factors in the controlled clinical trial. *Biometrics*, 31(1), pp.103-115.

Toftager, M., Christiansen, L.B., Kristensen, P.L. and Troelsen, J., 2011. SPACE for physical activity – a multicomponent intervention study: study design and baseline findings from a cluster randomized controlled trial. *BMC Public Health*, 11, pp.1-11.

Watson, L., Small, R., Brown, S., Dawson, W. and Lumley, J., 2004. Mounting a community-randomized trial: sample size, matching, selection, and randomization issues in PRISM. *Controlled Clinical Trials*, 25(3), pp.235-250.

Welsh, B.C., Podolsky, S.H. and Zane, S.N., 2023. Pair-matching with random allocation in prospective controlled trials: the evolution of a novel design in criminology and medicine, 1926–2021. *Journal of Experimental Criminology*, 19(4), pp.1115-1130.

Ye, T., Yi, Y. and Shao, J., 2022. Inference on the average treatment effect under minimization and other covariate-adaptive randomization methods. *Biometrika*, 109(1), pp.33-47.

Robin Dunn
robin.dunn@novartis.com

Thank you