Improving Precision and Power in Randomized Trials for COVID-19 Treatments Using Covariate Adjustment, for Binary, Ordinal, and Time-to-Event Outcomes

Michael Rosenblum
Associate Professor of Biostatistics
Johns Hopkins Bloomberg School of Public Health
Slides and paper at https://mrosenblumbiostat.wordpress.com

Co-authors: David Benkeser*, Ivan Diaz*, Alex Luedtke*,
Jodi Segal and Daniel Scharfstein
Recently accepted at *Biometrics* (* = co-first authors)
https://doi.org/10.1111/biom.13377

Motivation

- Over 800 randomized clinical trials (phase 2 and 3) of COVID-19 treatments registered on clinicaltrials.gov.
- March 2020: Request by the U.S. Food and Drug Administration (FDA) for statistical analysis recommendations for COVID-19 treatment trials.
- Primary outcomes in these trials often: binary, ordinal, time-to-event.
- We assessed potential value added by covariate adjustment by simulating two-arm trials with 1:1 randomization comparing a hypothetical COVID-19 treatment versus standard of care.
- Simulated distributions derived from data on over 500 patients hospitalized at New York Presbyterian Hospital, and a Centers for Disease Control and Prevention (CDC) preliminary description of 2449 cases.
- Submitted report in April, 2020, to FDA.

Our Problem and Goals

- Covariate adjustment in randomized trial:
  - Preplanned adjustment for baseline variables when estimating average treatment effect in primary efficacy analysis.
  - Can improve precision and reduce required sample size to achieve desired power.
- Problem: Covariate adjustment often misunderstood and underutilized, potentially wasting substantial resources, particularly for trials with binary, ordinal, or time-to-event outcome (common in COVID-19 treatment trials).
- Our goals:
  - Describe estimands, covariate-adjusted estimators, and implementation in R packages for these outcome types.
  - Use simulations based on real data to demonstrate impact of covariate adjustment in hypothetical COVID-19 trials.
  - Practical recommendations for implementation

Main Results

- Substantial precision gains from using covariate adjustment–equivalent to 4-18% reductions in required sample size to achieve a desired power–for a variety of estimands (targets of inference) for simulated trials with sample sizes 100, 200, 500, 1000.
- We provide an R package and practical recommendations for implementing covariate adjustment.
- The estimators that we consider are robust to model misspecification.

Benkeser, D., Diaz, I., Luedtke, A., Segal, J., Scharfstein, D., and Rosenblum, M. (In Press) Improving Precision and Power in Randomized Trials for COVID-19 Treatments Using Covariate Adjustment, for Binary, Ordinal, or Time to Event Outcomes. *Biometrics*. This paper was selected to be a discussion paper. https://doi.org/10.1111/biom.13377

Use of Covariate Adjustment in Randomized Trials: Two Surveys

Pocock et al. (2002) surveyed 50 randomized clinical trial reports.
Findings: "The statistical emphasis on covariate adjustment is quite complex and often poorly understood, and there remains confusion as to what is an appropriate statistical strategy."

Austin et al. (2010) surveyed 114 randomized trial articles.
Findings: only 39 presented an adjusted analysis.
Paper title: "A substantial and confusing variation exists in handling of baseline covariates in randomized controlled trials: a review of trials published in leading medical journals."

FDA Guidance Documents on Covariate Adjustment

1. ICH E9 Statistical Principles for Clinical Trials (1998): "Pretrial deliberations should identify those covariates and factors expected to have an important influence on the primary variable(s), and should consider how to account for these in the analysis to improve precision..."
2. FDA (2019): "Sponsors can use ANCOVA to adjust for differences between treatment groups in relevant baseline variables to improve the power of significance tests and the precision of estimates of treatment effect"
   (FDA draft guidance for continuous outcomes.)
3. FDA (2020) "To improve the precision of treatment effect estimation and inference, sponsors should consider adjusting for prespecified prognostic baseline covariates (e.g., age, baseline severity, comorbidities) in the primary efficacy analysis and should propose methods of covariate adjustment." (FDA Guidance on COVID-19 treatment and prevention trials)

Goal of Covariate Adjustment

- Population Average Treatment Effect is a contrast between (marginal) outcome distributions if all were assigned to treatment versus all assigned to control. (Intention To Treat)
- **Goal: Estimation of Average Treatment Effect in a Randomized Trial**.
  If baseline variables prognostic for outcome, can improve precision compared to unadjusted estimator.
- Related work on covariate adjustment, e.g., Yang and Tsiatis, 2001, Zhang et al. 2008; Tsiatis et al. 2008, Rubin and van der Laan, 2008, Zhang and Gilbert 2010, Moore et al. 2011, Tian et al. 2012, Zheng et al. 2015, Vermuelen et al. 2015, Wager et al. 2016, Zhang and Ma, 2019, Jiang et al. 2019.

Population, Baseline Variables, and Outcomes in COVID-19 context

- Population: hospitalized, COVID-19 positive patients
- Outcomes: intensive care unit (ICU) admission, intubation with ventilation, and death.
- Baseline variables: age, sex, required supplemental oxygen at ED presentation, dyspnea, hypertension, bilateral infiltrates on the chest x-ray

Estimands (Targets of Inference)

Estimands (contrasts between marginal distributions under treatment and control):

- For binary outcomes: risk difference, relative risk, odds ratio.
- For ordinal outcomes: difference in means, the Mann-Whitney estimand=P(random individual assigned to treatment has better outcome than random individual assigned to control with ties broken at random), and average of cumulative log odds ratios over outcome levels.
- For time-to-event outcomes: difference in restricted mean survival times, the difference in survival probabilities, and the ratio of survival probabilities.

Estimators:

- For each estimand, present a covariate adjusted estimator that leverages information in baseline variables to improve precision and reduce required sample size to achieve desired power.
- For ordinal outcomes, novel covariate adjusted estimators.

Data Generating Distributions for Simulations (Survival Outcomes)

- Patient data re-sampled with replacement from 500 patients hospitalized at Weill Cornell Medicine New York Presbyterian Hospital prior to March 28, 2020.
- Simulated sample sizes $n = 100$, 200, 500, and 1000.
- Hypothetical treatment variable drawn independent of all other data
- To simulate positive treatment effect: add independent draw from a χ^2 with 4 d.f. to each treatment arm participant's outcome
- Censoring: 5% censored completely at random; censoring time from uniform distribution on $\{1, \ldots, 14\}$.

Results: difference in restricted mean survival times (RMST) 14 days after hospitalization

Table: **Results when treatment effect is 1 day.** n=sample size; RE=relative efficiency (ratio of adjusted vs. unadj. MSE).

n	Estimator	Power	MSE	RE
100	Unadjusted	0.09	53.7	1.00
100	Adjusted	0.15	51.0	0.95
200	Unadjusted	0.33	62.7	1.00
200	Adjusted	0.40	56.4	0.90
500	Unadjusted	0.74	72.9	1.00
500	Adjusted	0.82	62.2	0.85
1000	Unadjusted	0.96	76.5	1.00
1000	Adjusted	0.98	63.5	0.83

R Packages

- Ordinal Outcomes: `R` package, `drord`, https://github.com/benkeser/drord.
- Survival Outcomes: `R` package `survtmlerct` https://github.com/idiazst/survtmlerct

Related work: Stratified Randomization and Covariate Adjustment

Wang, B., Susukida, R., Mojtabai, R., Masoumeh, A.-E.; and Rosenblum, M. (2019) Model-Robust Inference for Clinical Trials that Improve Precision by Stratified Randomization and Adjustment for Additional Baseline Variables. https://arxiv.org/abs/1910.13954

Recommendations for Primary Efficacy Analysis

1. **Estimand when the outcome is ordinal.** Recommend: difference between means or the Mann-Whitney estimand. Don't recommend log odds ratios.
2. **Covariate adjustment.** Adjust for prognostic baseline variables to improve precision and power.
    1. Baseline variables should be specified before the trial is started (or selected using prespecified algorithm, e.g., with cross-validation).
3. **Confidence intervals (CI) and hypothesis testing.** Nonparametric bootstrap (BCa), 10000 replicates for CI.
    1. Entire estimation procedure repeated in each replicate data set.
    2. Hypothesis tests: invert confidence interval or use permutation methods (latter especially for smaller sample sizes)

Recommendations for Primary Efficacy Analysis (con't)

1. **Use Information Monitoring**
   1. Final analysis time based on the information accrued (1/estimator variance).
   2. Precision gains from covariate adjustment translate into faster information accrual and shorter trial duration.
2. **Plotting the CDF and the probability mass function (PMF) when the outcome is ordinal.**
   1. Covariate adjusted estimate of the PMF and/or CDF of primary outcome plotted for each study arm.
   2. Pointwise and simultaneous confidence intervals displayed
3. **Missing covariates.** Impute based only on data from those covariates that were observed.
4. **Missing outcomes.** Use doubly robust methods and sensitivity analyses of robustness to assumptions.

Austin, P. C., A. Manca, M. Zwarenstein, D. N. Juurlink, and M. B. Stanbrook (2010). A substantial and confusing variation exists in handling of baseline covariates in randomized controlled trials: a review of trials published in leading medical journals. *Journal of Clinical Epidemiology 63*(2), 142–153.

Benkeser, D., M. Carone, and P. B. Gilbert (2018). Improved estimation of the cumulative incidence of rare outcomes. *Statistics in Medicine 37*(2), 280–293.

Benkeser, D., I. Diaz, A. Luedtke, J. Segal, D. Scharfstein, and M. Rosenblum (2020). Improving Precision and Power in Randomized Trials for COVID-19 Treatments Using Covariate Adjustment, for Ordinal or Time to Event Outcomes https://www.medrxiv.org/content/10.1101/2020.04.19.20069922v1?ve . *medRxiv*.

Benkeser, D., P. B. Gilbert, and M. Carone (2019). Estimating and testing vaccine sieve effects using machine learning. *Journal of the American Statistical Association 114*(527), 1038–1049.

Brooks, J. C., M. J. van der Laan, D. E. Singer, and A. S. Go (2013). Targeted minimum loss-based estimation of causal effects in right-censored survival data with time-dependent covariates: Warfarin, stroke, and death in atrial fibrillation. *Journal of Causal Inference 1*(2), 235–254.

CDC COVID-19 Response Team (2020). Severe Outcomes Among Patients with Coronavirus Disease 2019 (COVID-19) United States, February 12March 16, 2020. https://www.cdc.gov/mmwr/volumes/69/wr/mm6912e2.htm . *MMWR Morb Mortal Wkly Rep 69*(12), 343–346.

Chaisinanunkul, N., O. Adeoye, R. J. Lewis, J. C. Grotta, J. Broderick, T. G. Jovin, R. G. Nogueira, J. J. Elm, T. Graves, S. Berry, K. R. Lees, A. D. Barreto, J. L. Saver, null null, A. Furlan, B. Baxter, H. L. Lutsep, M. Ribo, O. Jansen, R. Gupta, and V. M. Pereira (2015). Adopting a patient-centered approach to primary outcome analysis of acute stroke trials using a utility-weighted modified rankin scale. *Stroke 46*(8), 2238–2243.

Chen, P.-Y. and A. A. Tsiatis (2001). Causal inference on the difference of the restricted mean lifetime between two groups. *Biometrics 57*(4), 1030–1038.

Díaz, I., E. Colantuoni, D. F. Hanley, and M. Rosenblum (2019). Improved precision in the analysis of randomized trials with survival outcomes, without assuming proportional hazards. *Lifetime Data Analysis 25*(3), 439–468.

Díaz, I., E. Colantuoni, and M. Rosenblum (2016). Enhanced precision in the analysis of randomized trials with ordinal outcomes. *Biometrics 72*(2), 422–431.

Efron, B. and R. J. Tibshirani (1994). *An introduction to the bootstrap*. CRC press.

EMA (2015). Guideline on adjustment for baseline covariates in clinical trials. Reference number EMA/CHMP/295050/2013. Committee for Medicinal Products for Human Use (CHMP).

FDA (2019). Adjusting for Covariates in Randomized Clinical Trials for Drugs and Biologics with Continuous Outcomes. Draft

Guidance for Industry. https://www.fda.gov/media/123801/download.

FDA (2020, May). COVID-19: Developing Drugs and Biological Products for Treatment or Prevention. Guidance for Industry. https://www.fda.gov/media/137926/download.

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

Ge, M., L. K. Durham, R. D. Meyer, W. Xie, and N. Thomas (2011). Covariate-adjusted difference in proportions from clinical trials using logistic regression and weighted risk differences. *Drug Information Journal 45*(4), 481–493.

Jiang, F., L. Tian, H. Fu, T. Hasegawa, and L. J. Wei (2019). Robust alternatives to ancova for estimating the treatment effect via a randomized comparative study. *Journal of the American Statistical Association 114*(528), 1854–1864.

Kaplan, E. L. and P. Meier (1958). Nonparametric estimation from

incomplete observations. *Journal of the American Statistical Association 53*(282), 457–481.

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.

McCullagh, P. (1980). Regression models for ordinal data. *Journal of the Royal Statistical Society: Series B (Methodological) 42*(2), 109–127.

Moore, K. L., R. Neugebauer, T. Valappil, and M. J. van der Laan (2011). Robust extraction of covariate information to improve estimation efficiency in randomized trials. *Statistics in Medicine 30*(19), 2389–2408.

Moore, K. L. and M. J. van der Laan (2009). Covariate adjustment in randomized trials with binary outcomes: Targeted maximum likelihood estimation. *Statistics in Medicine 28*(1), 39.

Moore, K. L. and M. J. van der Laan (2009). Increasing power in randomized trials with right censored outcomes through

covariate adjustment. *Journal of Biopharmaceutical Statistics 19*(6), 1099–1131. PMID: 20183467.

Nogueira, R. G., A. P. Jadhav, D. C. Haussen, A. Bonafe, R. F. Budzik, P. Bhuva, D. R. Yavagal, M. Ribo, C. Cognard, R. A. Hanel, C. A. Sila, A. E. Hassan, M. Millan, E. I. Levy, P. Mitchell, M. Chen, J. D. English, Q. A. Shah, F. L. Silver, V. M. Pereira, B. P. Mehta, B. W. Baxter, M. G. Abraham, P. Cardona, E. Veznedaroglu, F. R. Hellinger, L. Feng, J. F. Kirmani, D. K. Lopes, B. T. Jankowitz, M. R. Frankel, V. Costalat, N. A. Vora, A. J. Yoo, A. M. Malik, A. J. Furlan, M. Rubiera, A. Aghaebrahim, J.-M. Olivot, W. G. Tekle, R. Shields, T. Graves, R. J. Lewis, W. S. Smith, D. S. Liebeskind, J. L. Saver, and T. G. Jovin (2018). Thrombectomy 6 to 24 hours after stroke with a mismatch between deficit and infarct. *New England Journal of Medicine 378*(1), 11–21. PMID: 29129157.

Parast, L., L. Tian, and T. Cai (2014). Landmark estimation of survival and treatment effect in a randomized clinical trial.

Journal of the American Statistical Association 109(505), 384–394.

Royston, P. and M. K. Parmar (2011). The use of restricted mean survival time to estimate the treatment effect in randomized clinical trials when the proportional hazards assumption is in doubt. *Statistics in Medicine 30*(19), 2409–2421.

Scharfstein, D. O., A. Rotnitzky, and J. M. Robins (1999). Adjusting for nonignorable drop-out using semiparametric nonresponse models. *Journal of the American Statistical Association 94*(448), 1096–1120.

Stitelman, O. M., V. De Gruttola, and M. J. van der Laan (2011). A general implementation of tmle for longitudinal data applied to causal inference in survival analysis. *The International Journal of Biostatistics 8*(1).

Tsiatis, A. A., M. Davidian, M. Zhang, and X. Lu (2008). Covariate adjustment for two-sample treatment comparisons in randomized clinical trials: A principled yet flexible approach. *Statistics in Medicine 27*(23), 4658–4677.

van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge: Cambridge University Press.

Vermeulen, K., O. Thas, and S. Vansteelandt (2015). Increasing the power of the mann-whitney test in randomized experiments through flexible covariate adjustment. *Statistics in Medicine 34*(6), 1012–1030.

Yang, L. and A. A. Tsiatis (2001). Efficiency study of estimators for a treatment effect in a pretest–posttest trial. *The American Statistician 55*(4), 314–321.

Zhang, M. (2014). Robust methods to improve efficiency and reduce bias in estimating survival curves in randomized clinical trials. *Lifetime Data Analysis 21*(1), 119–137.

Zhang, M. and P. B. Gilbert (2010). Increasing the efficiency of prevention trials by incorporating baseline covariates. *Statistical Communications in Infectious Diseases 2*(1).

Zhang, M., A. A. Tsiatis, and M. Davidian (2008). Improving efficiency of inferences in randomized clinical trials using auxiliary covariates. *Biometrics 64*(3), 707–715.

- Austin, Peter C., et al. A substantial and confusing variation exists in handling of baseline covariates in randomized controlled trials: a review of trials published in leading medical journals. J. Clin Epi 63.2 (2010): 142-153.
- Colantuoni, E. and Rosenblum, M. (2015) Leveraging Prognostic Baseline Variables to Gain Precision in Randomized Trials. Statistics in Medicine. 34(18), 2602-2617. http://goo.gl/evGHF6
- Diaz, I., Colantuoni, E., Rosenblum, M. Enhanced Precision in the Analysis of Randomized Trials with Ordinal Outcomes. Biometrics. (2016)
- Jarskog, L. F., Hamer, R. M., Catellier, D. J., Stewart, D. D., LaVange, L., Ray,N., Golden, L. H., Lieberman, J. A. and Stroup, T. S.(2013). Metformin for weightloss and metabolic control in overweight outpatients with schizophrenia and schizoaffectivedisorder.American Journal of Psychiatry170(9), 1032–1040.

- Jiang, F., L. Tian, H. Fu, T. Hasegawa, M. A. Pfeffer, and L. J. Wei (2016). Robust alternatives to ANCOVA for estimating the treatment effect via a randomized comparative study, Harvard University Biostatistics Working Paper. Series Working paper 209.
- Moore K, van der Laan MJ. Covariate adjustment in randomized trials with binary outcomes: targeted maximum likelihood estimation. Statistics in Medicine 2009; 28(1):39-64.
- Moore, K. L., R. Neugebauer, T. Valappil, and M. J. van der Laan (2011). Robust extraction of covariate information to improve estimation efficiency in randomized trials. Statistics in Medicine 30 (19), 2389-2408.
- Petersen, R. C., Thomas, R. G., Grundman, M., Bennett, D., Doody, R., Ferris, S.,Galasko, D., Jin, S., Kaye, J., Levey, A., Pfeiffer, E., Sano, M., van Dyck, C. H.and others. (2005). Vitamin E and Donepezil for the Treatment of Mild Cognitive Impairment.New England Journal of Medicine 352(23),

- Pocock, S. J., Assmann, S. E., Enos, L. E., and Kasten, L. E. (2002). Subgroup analysis, covariate adjustment and baseline comparisons in clinical trial reporting: current practice and problems. Statistics in medicine, 21(19), 2917-2930.
- Robins JM, Sued M, Lei-Gomez Q, Rotnitzky A. Double-robustness with improved efficiency in missing and causal inference models. Technical Report, Harvard School of Public Health, 2007
- Rotnitzky A, Lei Q, Sued M, Robins JM. Improved double-robust estimation in missing data and causal inference models. Biometrika 2012; 99(2):439-456.
- Rubin D, van der Laan MJ. Empirical efficiency maximization: improved locally efficient covariate adjustment in randomized experiments and survival analysis. International Journal of Biostatistics 2008; 4(1):Article 5.

- Scharfstein, D.O., Rotnitzky, A., and Robins, J.M. Adjusting for nonignorable drop-out using semiparametric nonresponse models. Journal of the American Statistical Association 94, 1096–1120 (with Rejoinder, 1135 1146), 1999.
- Steingrimsson, Jon Arni; Hanley, Daniel F.; and Rosenblum, Michael. Improving precision by adjusting for prognostic baseline variables in randomized trials with binary outcomes, without regression modelassumptions, Contemporary Clinical Trials, 2017, http://doi.org/10.1016/j.cct.2016.12.026
- Treatment for Adolescents With Depression Study (TADS) Team. (2004). Fluoxe-tine, cognitive-behavioral therapy, and their combination for adolescents with depression: Treat-ment for adolescents with depression study (TADS) randomized controlled trial.JAMA292(7),807–820.
- Tan Z. Bounded, efficient and doubly robust estimating equations for marginal and nested structural models. Biometrika 2010; 97:661-682.

- Lu Tian, Tianxi Cai, Lihui Zhao, Lee-Jen Wei, On the covariate-adjusted estimation for an overall treatment difference with data from a randomized comparative clinical trial, Biostatistics, Volume 13, Issue 2, April 2012, Pages 256–273, https://doi.org/10.1093/biostatistics/kxr050
- Stefan Wager, Wenfei Du, Jonathan Taylor, Robert J. Tibshirani. High-dimensional randomized experiments. Proceedings of the National Academy of Sciences Nov 2016, 113 (45) 12673-12678; DOI: 10.1073/pnas.1614732113
- ZHANG, M., TSIATIS, A. AND DAVIDIAN, M. (2008). Improving efficiency of inferences in randomized clinical trials using auxiliary covariates. Biometrics 64, 707?715.
- Zhang, Z, Ma, S. Machine learning methods for leveraging baseline covariate information to improve the efficiency of clinical trials. Statistics in Medicine. 2019; 38: 1703–1714. https://doi.org/10.1002/sim.8054
- Zheng, Wenjing; Chambaz, Antoine; and van der Laan, Mark J., "Drawing Valid Targeted Inference When

Acknowledgments

David Benkeser, Iván Díaz, and Alex Luedtke are co-first authors and contributed equally to this manuscript. AL was supported by the National Institutes of Health under award number DP2-LM013340. MR was supported by the Johns Hopkins Center of Excellence in Regulatory Science and Innovation, which is funded by the Food and Drug Administration (FDA) of the U.S. Department of Health and Human Services (HHS) as part of a financial assistance award (U01FD005942). The contents are those of the author(s) and do not necessarily represent the official views of, nor an endorsement, by any of the aforementioned organizations, the FDA/HHS, nor the U.S. Government.