REJOINDER

Biometrics WILEY

Rejoinder: Improving precision and power in randomized trials for COVID-19 treatments using covariate adjustment, for binary, ordinal, and time-to-event outcomes

David Benkeser[1] | **Iván Díaz[2]** | **Alex Luedtke[3,4]** | **Jodi Segal[5]** |
Daniel Scharfstein[6] | **Michael Rosenblum[7]**

[1] Department of Biostatistics and Bioinformatics, Emory University, Atlanta, Georgia, USA

[2] Division of Biostatistics, Department of Population Health Sciences, Weill Cornell Medicine, New York, New York, USA

[3] Department of Statistics, University of Washington, Seattle, Washington, USA

[4] Vaccine and Infectious Disease Division, Fred Hutchinson Cancer Research Center, Seattle, Washington, USA

[5] Department of Medicine, School of Medicine, Johns Hopkins University, Baltimore, Maryland, USA

[6] Division of Biostatistics, Department of Population Health Sciences, University of Utah School of Medicine, Salt Lake City, Utah, USA

[7] Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Johns Hopkins University, Baltimore, Maryland, USA

Correspondence
Michael Rosenblum, Department of Biostatistics, Johns Hopkins Bloomberg School of Public Health, Johns Hopkins University.
Email: mrosen@jhu.edu

We thank Drs. L. LaVange, M. Zhang, B. Zhang, and M. Proschan for their insightful commentaries on our manuscript. We respond to each in turn.

Dr. LaVange

Dr. LaVange highlighted the importance of covariate adjustment in large randomized trials. We could not agree more! Related to this, it is important to avoid the common mistake of selecting baseline variables to adjust for based on which ones have statistically significant imbalances across study arms (Pocock *et al.*, 2002). The variables should either be selected before the trial starts (selecting those that are most prognostic for the outcome based on prior data), or selected using the trial data based on a completely prespecified algorithm that aims to select the most prognostic variables.

Furthermore, Dr. LaVange emphasizes the importance of effect estimates that are valid "even in the presence of model misspecification." We entirely agree with this point which was the impetus for the model-robust, covariate

adjusted estimators in our paper. Related to this point, we agree with Uno *et al.* (2015); Pak *et al.* (2017), who advocate for using estimands (i.e., targets of inference) such as the restricted mean survival time that have a model-free interpretation. This is in contrast to commonly used estimands such as the hazard ratio (for time-to-event outcomes under a proportional hazards model) or odds ratio (for ordinal outcomes under a proportional odds model), which are not model-free estimands, and therefore may be difficult or impossible to interpret under model misspecification (i.e., if the proportional hazards model or proportional odds model is misspecified).

Drs. M. Zhang and B. Zhang

Drs. M. Zhang and B. Zhang describe a general approach for constructing covariate adjusted estimators that applies to many estimands. In particular, one of their estimators has improved precision compared to the methods that we proposed, when applied to our simulation distributions. This is impressive and makes a good argument for using the corresponding estimator in practice. One difference between their general approach and the approach in our paper is that the latter produces substitution estimators.

Substitution estimators have the potential advantage of always being in the parameter space, for example, being between 0 and 1 when estimating a probability. Though in many cases this may not matter, it may be important when the true parameter values are close to the boundary. It is an area of future research to compare the general approach of Drs. M. Zhang and B. Zhang versus our general approach (using substitution estimators) across a variety of simulation studies that mimic features of completed trial data sets.

Drs. M. Zhang and B. Zhang state that "there is a dilemma in that covariate adjustment is more useful in improving efficiency of inferences when sample size is large, in which case efficiency is of less a concern." We respectfully disagree, and think that efficiency can be a major concern in large trials (as well as small trials). For example, a relative efficiency of 0.83 at sample size 1000 (as in the last row of Table 6) is approximately equivalent to a 17% reduction in the required sample size to achieve a desired power; this is approximately equivalent to a sample size reduction of 170 participants, which we consider to be important. More generally, at a fixed relative efficiency that is less than 1, the sample size reduction due to covariate adjustment is approximately proportional to the trial's sample size; this means that the impact can be substantial at large sample sizes. The importance of improving precision by covariate adjustment in large trials was highlighted by Dr. LaVange in her commentary.

Dr. Proschan

Dr. Proschan correctly pointed out that our methods were presented in the context of simple randomization. Stratified randomization is often used in phase 2 and 3 clinical trials (Lin *et al.*, 2015). The methods that we presented for binary and ordinal outcomes can be directly applied to this case, except that the variance estimator should be modified to account for the stratified randomization procedure (thereby potentially increasing power in a corresponding hypothesis test). This modification can be done using the method in Wang *et al.* (2020), which gives a general formula for the asymptotic variance of M-estimators in randomized trials that use stratified randomization. For time-to-event outcomes, it is currently an open problem to determine the asymptotic variance for the covariate adjusted estimator that we used from Díaz *et al.* (2019), under stratified randomization. We conjecture that variance formulas (4-5) in Wang *et al.* (2020) can be used to consistently estimate the asymptotic variance in this case. An alternative estimator to consider for time-to-event outcomes is the augmented, inverse probability weighted estimator from Díaz *et al.* (2019), for which the variance estimation method in Wang *et al.* (2020) can be directly applied.

Dr. Proschan clarified the contrast between conditional and marginal treatment effects. We much appreciate this, since it has been a common source of confusion in the context of covariate adjustment. Dr. Proschan posed the important question of which is more efficient in the context of logistic regression for binary outcomes: conditional or marginal tests (where both are covariate adjusted, and where the marginal test is a Wald test based on the estimator from Section 3.1 of our paper)? It was shown by Rosenblum and Steingrimsson (2016) that these are equally efficient, asymptotically, when the logistic regression model includes an intercept and main terms for treatment and baseline variables; this result holds under arbitrary model misspecification. Since the marginal treatment effect is interpretable without requiring model assumptions (unlike the conditional effect), we recommend to estimate and test marginal treatment effects in practice (as we did in our manuscript).

Dr. Proschan suggested the use of randomization inference, which can also easily incorporate covariate adjustment. It has the advantage of not requiring distributional assumptions, and may be especially useful at smaller sample sizes where asymptotic arguments may be less applicable. A potential downside is that the confidence intervals produced by randomization inference are typically based on inverting tests of the null hypothesis that the treatment effect is identical to a fixed value for all participants. If treatment effects differ across individuals, then it may be difficult to interpret these confidence intervals. Lastly, we thank Dr. Proschan for pointing out our mistaken reference to the primary outcome in Beigel *et al.* (2020) as time to death; it was time to recovery.

ORCID

David Benkeser https://orcid.org/0000-0002-1019-8343
Daniel Scharfstein https://orcid.org/0000-0001-7482-9653
Michael Rosenblum https://orcid.org/0000-0001-7411-4172

REFERENCES

Beigel, J.H., Tomashek, K.M., Dodd, L.E., Mehta, A.K., Zingman, B.S., Kalil, A.C. *et al.* (2020) Remdesivir for the treatment of Covid-19 – Preliminary report. *New England Journal of Medicine*, 383, 1813–1826. https://doi.org/10.1056/NEJMoa2007764

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

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

Pak, K., Uno, H., Kim, D.H., Tian, L., Kane, R.C., Takeuchi, M. et al. (2017) Interpretability of cancer clinical trial results using Restricted mean survival time as an alternative to the hazard ratio. *JAMA Oncology*, 3, 1692–1696.

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, 2917–2930.

Rosenblum, M. and Steingrimsson, J.A. (2016) Matching the efficiency gains of the logistic regression estimator while avoiding its interpretability problems, in randomized trials. Johns Hopkins University, Department of Biostatistics Working Papers. Working Paper 281. https://biostats.bepress.com/jhubiostat/paper281

Uno, H., Wittes, J., Fu, H., Solomon, S.D., Claggett, B., Tian, L. et al. (2015) Alternatives to hazard ratios for comparing the efficacy or safety of therapies in noninferiority studies. *Annals of Internal Medicine*, 163, 127–134.

Wang, B., Susukida, R., Mojtabai, R., Amin-Esmaeili, M. and Rosenblum, M. (2020) Model-robust inference for clinical trials that improve precision by stratified randomization and covariate adjustment. Available at https://arxiv.org/abs/1910.13954. Accessed May 10, 2021.