



## Rejoinder on 'Inference after covariate-adaptive randomization: aspects of methodology and theory'

Jun Shao

To cite this article: Jun Shao (2021) Rejoinder on 'Inference after covariate-adaptive randomization: aspects of methodology and theory', *Statistical Theory and Related Fields*, 5:3, 196-199, DOI: [10.1080/24754269.2021.1905357](https://doi.org/10.1080/24754269.2021.1905357)

To link to this article: <https://doi.org/10.1080/24754269.2021.1905357>



Published online: 24 Mar 2021.



Submit your article to this journal [↗](#)



Article views: 31



View related articles [↗](#)



View Crossmark data [↗](#)

# Rejoinder on ‘Inference after covariate-adaptive randomization: aspects of methodology and theory’

Jun Shao<sup>a,b</sup>

<sup>a</sup>KLATASDS-MOE, School of Statistics, East China Normal University, Shanghai, People’s Republic of China; <sup>b</sup>Department of Statistics, University of Wisconsin-Madison, Madison, WI, USA

**ARTICLE HISTORY** Received 11 March 2021; Accepted 16 March 2021

I would like to thank all discussants for their insightful discussions on the topic of statistical inference after covariate-adaptive randomisation, especially for including reviews of some new results and references that are not in my review written more than a year ago. I hope these discussions together with my review will stimulate further studies in this important area having many applications particularly in clinical trials.

My rejoinder focuses on some main points from four separate groups of discussants.

## 1. The discussion by Drs. Ma, Zhang and Hu

Drs. Ma, Zhang, and Hu’s discussion brings some new results or interesting directions for new studies. These include, but are not limited to, robust inference with the randomisation scheme considered in Hu and Hu (2012), inference after covariate-adaptive randomisation with covariate misclassification, unobserved covariates, missing data or non-compliance, and high dimensional covariates. Covariate-adaptive randomisation can also be combined with other adaptive designs such as sequential monitoring (Zhu & Hu, 2019), sample size re-estimation (Li et al., *in press*), and seamless phase II/III clinical trials (Ma et al., *in press*).

As I discussed in Section 7 of my review, the theoretical study of covariate-adaptive randomisation schemes such as Pocock and Simon’s minimisation is not completed and still important, although some asymptotically valid inference procedures after these randomisation schemes have been derived. Let  $\mathbf{Z}$  be the discrete covariate used in covariate-adaptive randomisation,  $z_1, \dots, z_c$  be all possible categories of  $\mathbf{Z}$ ,  $\pi_1, \dots, \pi_k$  be the target assignment proportions in the trial,  $n$  be the total sample size of the trial,  $n_a(l)$  be the number of patients in treatment  $a$  with  $\mathbf{Z} = z_l$ ,  $a = 1, \dots, k$ ,  $l = 1, \dots, c$ , and  $n(l) = n_1(l) + \dots + n_k(l)$ . A key result for studying the asymptotic validity of inference procedures after covariate-adaptive randomisation is

$$(R1) \quad \sqrt{n} \left( \frac{n_a(l)}{n(l)} - \pi_a, \quad a = 1, \dots, k, l = 1, \dots, c \right) | \mathbf{Z}_1, \dots, \mathbf{Z}_n \rightarrow N(0, D) \text{ in distribution,}$$

i.e. conditioned on  $n$  observed values of  $\mathbf{Z}$ ,  $\mathbf{Z}_1, \dots, \mathbf{Z}_n$ , the  $k \times c$  dimensional vector whose  $(a, l)$ th component is  $\frac{n_a(l)}{n(l)} - \pi_a$  converges in distribution to a multivariate normal with mean 0 and some covariance matrix  $D$ . Result (R1) holds for type 1 or 2 covariate-adaptive randomisation schemes as described in Section 3 of my review. For Pocock and Simon’s minimisation, however, it has not been rigorously shown in general whether or not (R1) holds, although (Ma et al., 2015) showed (R1) under restrictive conditions (a correct linear model between the response and  $\mathbf{Z}$  whose components are independent). Hu and Zhang (2020) derived the asymptotic normality for  $\frac{n_a(l)}{n(l)} - \pi_a$ ,  $a = 1, \dots, k$  with a single fixed  $l$ , but (R1) requires joint asymptotic normality of the entire vector over all  $l = 1, \dots, c$ .

Without (R1), asymptotic validity of inference procedures for most covariate-adaptive randomisation schemes including Pocock and Simon’s minimisation can still be established under some conditions, which will be further explained in Section 4.

To construct asymptotic valid inference procedures, sometimes the explicit form of the covariance matrix  $D$  in (R1) is required.

I am less excited about balancing continuous covariates with covariate-adaptive randomisation. The reason is that covariate-adaptive randomisation is mainly used to enhance the credibility of the results of the trial (EMA, 2015) in terms of the balancedness of treatments across levels of some common discrete baseline prognostic factors such as institution, disease stage, prior treatment, gender, and age group. In fact, balancedness of marginal levels of these discrete prognostic factors is the main concern of agencies such as the EMA or FDA, and that is why Pocock and Simon’s minimisation is popular. I do not see a clear motivation of balancing a continuous baseline covariate at the design stage. If efficiency is the concern, we may simply adjust for this continuous covariate in the inference procedure. It is

much easier to construct a valid and efficient inference procedure, compared with to derive a valid inference procedure after balancing a continuous baseline covariate. So far, valid inference procedures after balancing a continuous baseline covariate are mostly model-based.

## 2. The discussion by Drs. Wang, Susukida, Mojtabai, Amin-Esmaeili and Rosenblum

My review mostly focuses on differences of sample means (or quantiles) and tests in survival analysis. The discussion by Drs. Wang, Susukida, Mojtabai, Amin-Esmaeili, and Rosenblum and their article (Wang et al., 2020) open up a wide range of robust inference methods to handle nonlinearity, various outcome types, repeated measures, missing outcomes, etc. Wang et al. (2020) also contain three examples of analyses of trial data, illustrating that the gain due to stratified permuted block randomisation and covariate adjustment could be as high as 36%.

In Section 5.3 of my review, empirical distribution estimators and related quantile estimators valid under covariate-adaptive randomisation are considered. Wang et al. (2020) established the asymptotic normality of Kaplan–Meier estimator under stratified permuted block or biased coin randomisation, which is important for survival analysis. Specifically, they showed that, in a trial with two arms ( $k = 2$ ),  $\{\sqrt{n}(\widehat{S}_n^{(a)} - S^{(a)}), t \in [0, \tau]\}$ , where  $\widehat{S}_n^{(a)}$  is the Kaplan–Meier estimator of the survival function  $S^{(a)}$ ,  $a$  is a fixed treatment, and  $\tau > 0$  is a fixed constant, converges weakly to a mean 0, tight Gaussian process with a covariance function  $V^{(a)}(t, t')$  explicitly given in the Supplementary Material of Wang et al. (2020). They also showed that

$$V^{(a)}(t, t) = \widetilde{V}^{(a)}(t, t) - \frac{U(t)}{\pi_1 \pi_2},$$

where  $\widetilde{V}^{(a)}(t, t')$  is the covariance function under simple randomisation and  $U(t) > 0$  under stratified permuted block or biased coin randomisation. Again, we see the common phenomenon of reducing variance by applying covariate-adaptive randomisation compared with simple randomisation.

As commented by Wang et al. (2020), the result can be extended to the estimation of survival function with adjusted baseline covariates. Alternatively, one may consider a stratified version of Kaplan–Meier estimator along with the idea in formula (9) of my review.

## 3. The discussion by Dr. Liu

Dr. Liu's discussion provides useful details and references about asymptotic validity and efficiency of

model-assisted inference procedures after covariate-adaptive randomisation and adjustment for covariates. The discussion about the efficiency gain in using ANOVA versus ANCOVA or ANCOVA with treatment-by-covariate interactions started as early as Yang and Tsiatis (2001), continued later by Freedman (2008), Lin (2013) and Wang et al. (2019), and studied under covariate-adaptive randomisation recently by Bugni et al. (2018), Bugni et al. (2019), Liu and Yang (2020), Ma et al. (2020b), Wang et al. (2020) and Ye et al. (in press).

I would like to emphasise two points here. The first one is, as pointed out by Dr. Liu, when there are only two treatment arms and equal allocation is used ( $k = 2$  and  $\pi_1 = \pi_2 = 1/2$ ), the use of ANCOVA with or without treatment-by-covariate interaction has the same asymptotic efficiency and is guaranteed to be more efficient than the use of ANOVA without adjusting for covariates. However, this phenomenon no longer exists once there are more than two treatment arms even if equal allocation is applied.

The second point is that, asymptotically, the most efficient estimator of the treatment difference  $\theta = E(Y^{(a)} - Y^{(b)})$  defined in the beginning of Section 5.2 of my review is the  $\widehat{\theta}_A$  defined in Section 6.1 of my review, which adjusts for covariates by using a working linear model and the ordinary least squares estimator

$$\widehat{\beta}_a(z) = \left[ \sum_{i \in L_a(z)} \{U_i - \bar{U}_a(z)\} \{U_i - \bar{U}_a(z)\}^T \right]^{-1} \times \sum_{i \in L_a(z)} \{U_i - \bar{U}_a(z)\} Y_i$$

of covariate effect within each stratum  $L_a(z)$  under treatment  $a$  and  $\mathbf{Z} = z$ . As Dr. Liu pointed out, however, when there are small strata formed by levels of  $\mathbf{Z}$ , the stratum-specific least squares estimator  $\widehat{\beta}_a(z)$  might lead to inferior performance due to over-fitting (Liu & Yang, 2020). One modification is to combine  $\widehat{\beta}_a(z)$  and  $\widehat{\beta}_b(z)$  within each stratum level  $z$ , although they may estimate different quantities. Alternatively, utilising the fact that baseline covariates  $U_i$ 's have the same distribution over all treatment arms, Ye et al. (in press) recommended to replace the matrix inverse in  $\widehat{\beta}_a(z)$  by an average over all treatment arms to remedy the issue of small strata, which leads to replace  $\widehat{\beta}_a(z)$  by

$$\widetilde{\beta}_a(z) = \left[ \frac{1}{n_a(z)} \sum_{a=1}^k \sum_{i \in L_a(z)} \{U_i - \bar{U}_a(z)\} \{U_i - \bar{U}_a(z)\}^T \right]^{-1} \times \frac{1}{n_a(z)} \sum_{i \in L_a(z)} \{U_i - \bar{U}_a(z)\} Y_i,$$

where  $n_a(z)$  is the number of units in  $L_a(z)$  and  $n(z) = n_1(z) + \dots + n_k(z)$ . Note that stability issues related

with dimensionality for not very large data sets are mainly in the inverses of estimated covariance matrices. Hence, using the inverse of an average may largely remedy the issue of small strata. Some simulation results in Ye et al. (in press) show that using  $\hat{\beta}_a(z)$  in the estimation of  $\theta$  leads to better finite-sample performance compared with using  $\hat{\beta}_a(z)$  or combining  $\hat{\beta}_a(z)$  and  $\hat{\beta}_b(z)$  when they actually estimate different quantities.

Finally, another way to handle many covariates is to apply high-dimensional technique as Dr. Liu commented (Ma et al., 2020a), or to use variable selection.

#### 4. The discussion by Drs. Ye and Yi

In their discussion, Drs. Ye and Yi clearly described the working models behind estimators  $\hat{\theta}_S$ ,  $\hat{\theta}_A$  and  $\hat{\theta}_B$  in Sections 5.2 and 6.1 of my review. This not only provides explanations about the asymptotic relative efficiencies among  $\hat{\theta}_S$ ,  $\hat{\theta}_A$  and  $\hat{\theta}_B$ , but also leads to a general working model (formula (1) in the discussion) that produces a class of model-assisted estimators of  $\theta$  (formula (2) in the discussion) including  $\hat{\theta}_S$ ,  $\hat{\theta}_A$  and  $\hat{\theta}_B$  as special cases.

In the beginning of Section 6.1 of my review,  $\mathbf{X}$  is considered to be the vector of all available baseline covariates,  $\mathbf{Z}$  is the discrete baseline covariate vector (part of  $\mathbf{X}$ ) used in covariate-adaptive randomisation, and  $\mathbf{U}$  is the vector of covariates not in  $\mathbf{Z}$  but in  $\mathbf{X}$  to be adjusted for efficiency in the analysis stage. I would like to point out that  $\mathbf{U}$  may contain some components which are interactions between  $\mathbf{Z}$  and covariates not in  $\mathbf{Z}$ . Drs. Ye and Yi's discussion classifies  $(\mathbf{Z}, \mathbf{U})$  into two categories or vectors,  $\mathbf{W}$  and  $\mathbf{V}$ , where  $\mathbf{W}$  contains covariates having treatment-by-covariate interaction in working model (1) in their discussion and  $\mathbf{V}$  has no treatment-by-covariate interaction. Note that either  $\mathbf{W}$  or  $\mathbf{V}$  could be empty. For example, for ANOVA without using any covariate, both  $\mathbf{W}$  and  $\mathbf{V}$  are empty; for classical ANCOVA without considering any treatment-by-covariate interaction,  $\mathbf{W}$  is empty but  $\mathbf{V}$  is not; as discussed by Drs. Ye and Yi,  $\hat{\theta}_S$  in Section 5.2 of my review corresponds to  $\mathbf{W} = \mathbf{Z}$  and empty  $\mathbf{V}$ ,  $\hat{\theta}_A$  in Section 6.1 corresponds to  $\mathbf{W} = (\mathbf{Z}, \mathbf{U})$  and empty  $\mathbf{V}$ , and  $\hat{\theta}_B$  in Section 6.1 corresponds to  $\mathbf{W} = \mathbf{Z}$  and  $\mathbf{V} = \mathbf{U}$ .

In applications, a crucial question is, what is the minimum set of covariates to be included in  $\mathbf{W}$  or  $\mathbf{V}$  to ensure that the resulting model-assisted estimator of  $\theta$  is asymptotically normal with mean  $\theta$  and variance invariant to the covariate-adaptive randomisation schemes (including Pocock and Simon's minimisation)? As pointed out by Drs. Ye and Yi, a simple answer is that  $\mathbf{W}$  should contain the dummy variables for all joint levels of  $\mathbf{Z}$ , and there is no requirement on  $\mathbf{V}$ . In fact,  $\mathbf{V}$  is used to not let the dimension of  $\mathbf{W}$  become too high. Asymptotically, the estimator with  $\mathbf{V}$  being empty is most efficient, unless some components of  $\mathbf{W}$

are actually not related with the response. We must balance between adjusting for covariates and over-fitting, for which variable selection may be a useful solution.

#### Disclosure statement

No potential conflict of interest was reported by the author(s).

#### Notes on contributor

*Dr. Jun Shao* holds a PhD in statistics from the University of Wisconsin-Madison. He is a Professor of Statistics at the University of Wisconsin-Madison. His research interests include variable selection and inference with high dimensional data, sample surveys, and missing data problems.

#### References

- Bugni, F. A., Canay, I. A., & Shaikh, A. M. (2018). Inference under covariate-adaptive randomization. *Journal of the American Statistical Association*, 113, 1784–1796. <https://doi.org/10.1080/01621459.2017.1375934>
- Bugni, F. A., Canay, I. A., & Shaikh, A. M. (2019). Inference under covariate-adaptive randomization with multiple treatments. *Quantitative Economics*, 10, 1747–1785. <https://doi.org/10.3982/QE1150>
- EMA (2015). *Guideline on adjustment for baseline covariates in clinical trials*. Committee for Medicinal Products for Human Use, European Medicines Agency.
- Freedman, D. A. (2008). On regression adjustments in experiments with several treatments. *Annals of Applied Statistics*, 2, 176–196. <https://doi.org/10.1214/07-AOAS143>
- Hu, F., & Zhang, L.-X. (2020). On the theory of covariate-adaptive designs. arXiv:2004.02994.
- Hu, Y., & Hu, F. (2012). Asymptotic properties of covariate-adaptive randomization. *Annals of Statistics*, 40, 1794–1815. <https://doi.org/10.1214/12-AOS983>
- Li, X., Ma, W., & Hu, F. (in press). Sample size re-estimation for covariate-adaptive randomized clinical trials. *Statistics in Medicine*, 28.
- Lin, W. (2013). Agnostic notes on regression adjustments to experimental data: Reexamining freedman's critique. *Annals of Applied Statistics*, 7, 295–318. <https://doi.org/10.1214/12-AOAS583>
- Liu, H., & Yang, Y. (2020). Regression-adjusted average treatment effect estimators in stratified randomized experiments. *Biometrika*, 107, 935–948. <https://doi.org/10.1093/biomet/asaa038>
- Ma, W., Hu, F., & Zhang, L. (2015). Testing hypotheses of covariate-adaptive randomized clinical trials. *Journal of the American Statistical Association*, 110, 669–680. <https://doi.org/10.1080/01621459.2014.922469>
- Ma, W., Tu, F., & Liu, H. (2020a). A general theory of regression adjustment for covariate-adaptive randomization: Ols, lasso, and beyond. arXiv:2011.09734.
- Ma, W., Tu, F., & Liu, H. (2020b). Regression analysis for covariate-adaptive randomization: A robust and efficient inference perspective. arXiv:2009.02287.
- Ma, W., Wang, M., & Zhu, H. (in press). Regression analysis for covariate-adaptive randomization: A robust and efficient inference perspective. *Statistica Sinica*, 30.
- Wang, B., Ogburn, E., & Rosenblum, M. (2019). Analysis of covariance in randomized trials: More precision and valid confidence intervals, without model assumptions.

- Biometrics*, 75, 1391–1400. <https://doi.org/10.1111/biom.v75.4>
- Wang, B., Susukida, R., Mojtabai, R., Amin-Esmaili, M., & Rosenblum, M. (2020). Model-robust inference for clinical trials that improve precision by stratified randomization and adjustment for additional baseline variables. arXiv:1910.13954v3.
- Yang, L., & Tsiatis, A. (2001). Efficiency study of estimators for a treatment effect in a pretest–posttest trial. *The American Statistician*, 55, 314–321. <https://doi.org/10.1198/000313001753272466>
- Ye, T., Yi, Y., & Shao, J. (in press). Inference on average treatment effect under minimization and other covariate-adaptive randomization methods. *Biometrika*, 108.
- Zhu, H., & Hu, F. (2019). Sequential monitoring of covariate-adaptive randomized clinical trials. *Statistica Sinica*, 29, 265–282.