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

Min Zhang

Department of Biostatistics, University of Michigan, Ann Arbor, MI, U.S.A.

email: mzhangst@umich.edu

and

Baqun Zhang

School of Statistics and Management, Shanghai University of Finance and Economics, Shanghai, P.R.China.

email: zhang.baqun@mail.shufe.edu.cn

1. Introduction

We congratulate the authors on a timely paper on covariate adjustment for COVID-19 treatment trials. It clearly demonstrates the great potential of leveraging covariates in increasing precision and power for randomized trials. Despite great advance in theory and methods, covariate adjustment is still underused in practice. This is partly due to that many practioners remain skeptical of its usefulness or have other concerns. Some people think adjusted estimator estimates the conditional, as opposed to the desired marginal, treatment effect. This thought is naturally influenced by the traditional adjustment method where one directly models outcomes as a function of treatment and covariates. Perhaps the biggest obstacle to leveraging covariate is the concern about model misspecification. People may think that the validity of inference and/or improvement in efficiency rely on the assumption of correct modeling. Benkeser et al alleviate these concerns through intuitive explanations and convincing empirical evidences. To further elucidate misunderstandings we elaborate the key points from a theoretical point of view and discuss issues from a practical point of view.

2. A Theoretical Perspective

The semiparametric framework of Zhang, Tsiatis and Davidian (2008, henceforth, ZTD) considers a trial with data on (Y, A, X). The outcome Y is general and can be continuous, binary, ordinal or other types. It starts with a relevant unadjusted estimand, β, and seeks to identify all valid estimators by studying influence functions. Then it characterizes all possible joint distribution of data without imposing any additional assumptions, except for that treatment A is independent of covariates X. The estimand is general and can be of any form, eg, difference in means, odds ratio, relative risk, or the Mann-Whiteney estimand.

ZTD showed that, for $A = 0$ or 1, the class of all unbiased estimating functions for β are

$$m^*(Y, X, A; \beta) = m(Y, A; \beta) - (A - \pi)h(X), \tag{1}$$

where $m(Y, Z; \beta)$ is any unbiased estimating function used in an unadjusted analysis, $h(X)$ is an arbitrary function of X, and $\pi = Pr(A = 1)$. Given $m(Y, A; \beta)$, the optimal $h(X)$ is

$$E\{m(Y, A; \beta) \,|\, X, A = 1\} - E\{m(Y, A; \beta) \,|\, X, A = 0\}. \tag{2}$$

According to (1), an adjusted estimator solving $\sum_{i=0}^{n} m^*(Y_i, X_i, A_i; \beta) = 0$ is consistent and asymptotically normal regardless of the form of $h(X)$ and therefore is guaranteed to be robust. Result (2) says that the estimator with the optimal $h(X)$ is always more efficient than the unadjusted one, which corresponds to $h(X) = 0$. Result (2) suggests that to improve efficiency one needs to model $E\{m(Y, A; \beta) \,|\, X, A = a\}$. However, it may seem less satisfying if efficiency gain relies on correct modeling. Suppose one postulates a model, likely misspecified, for $E\{m(Y, A; \beta) \,|\, X, A = a\} = \zeta_a^T g(X)$, where $g(X)$ is a vector of known basis functions. Restricting the class in (1) to the subclass with $h(X) = \zeta^T g(X)$ for some ζ, the optimal ζ is the limit of the ordinary least square (OLS) estimator in a regression with $m(Y, A; \beta)$ as the outcome and $(A - \pi)g(X)$ as covariates (Leon, Tsiatis, Davidian, 2003; ZTD). It is also equivalent to fitting $E\{m(Y, A; \beta) \,|\, X, A = a\} = \zeta_a^T g(X)$ using OLS. This adjusted estimator is guaranteed to be as good and often better than the unadjusted estimator. Therefore, efficiency gain also does not rely on correct modeling.

To summarize, if done properly, 1) covariate adjusted estimator estimates the marginal treatment effect; 2) the validity of adjusted analysis does not require the assumption of correct modeling; 3) efficiency improvement (smaller variance and better power) is guaranteed without the assumption of correct modeling, as long as covariates are predictive of outcomes.

3. A Unified Implementation Strategy

The theory leads to a simple unified covariate adjustment method for all estimands, where one augments the estimation equation used for an unadjusted analysis by an augmentation term, $-(A - \pi)h(X)$. It is easier to see the augmentation term will not introduce bias as

$E\{(A - \pi)h(X)\} = 0$. Obtaining estimators by solving estimating equations is perhaps the most familiar approach to statisticians, owing to the widespread use of likelihood-based methods and score equations. We think it is a huge advantage to embed covariate adjustment methods within a familiar and well-accepted framework, as it promotes understanding and use of it. Otherwise, learning how to do covariate adjustment robustly can be a daunting task for practioners, as there are so many different estimands of interest in practice and all kinds of adjustment methods. In fact, all consistent and asymptotically normal adjusted estimators are in this class or asymptotically equivalent to estimators in this class. So we do not lose by focusing on one unified augmented estimating equation framework.

In practice, one needs to replace π with $\widehat{\pi} = n_1/n$ and model $E\{m(Y, A; \beta) \,|\, X, A = a\}$, treating $m(Y_i, A_i; \beta)$ as data and replacing β in m by the unadjusted estimator. Often $m(Y, A; \beta)$ is linear in Y. Then it is equivalent to modeling for $E(Y|X, A = a)$. Examples for several common estimands are given in ZTD, including adjusted Mann-Whitney (MW) U/KruskalWallis test. To further illustrate the simplicity of this approach, we provide explicit formula for the adjusted MW estimator. By ZTD, augmented estimating equation for β is

$$\sum_{i=1}^{n} \left\{ \left[\frac{n}{n_0 n_1} \sum_{j=1}^{n} \{(1 - A_i)A_j C(Y_i, Y_j)\} - \beta \right] - (A_i - \widehat{\pi})h(X_i) \right\} = 0,$$

where $C(Y_i, Y_j) = I(Y_i < Y_j) + \frac{1}{2}I(Y_i = Y_j)$ and terms in [.] are the usual estimating function for the MW estimator. The covariate adjusted MW estimator is

$$\widehat{\beta}_{adj} = \left\{ \frac{1}{n_0 n_1} \sum_{i=1}^{n} \sum_{j=1}^{n} (1 - A_i)A_j C(Y_i, Y_j) \right\} - \frac{1}{n} \sum_{i=1}^{n} (A_i - \widehat{\pi})\widehat{h}(X_i).$$

That is, it is simply the usual MW estimator in {.} plus an augmentation term. As described previously, $\widehat{h}(X_i)$ can be obtained by $\widehat{g}_1(X_i) - \widehat{g}_0(X_i)$, where $\widehat{g}_a(X_i)$, $a = 0, 1$, is the fitted model for $M_i \equiv \frac{n}{n_0 n_1} \sum_{j=1}^{n} (1 - A_i)A_j C(Y_i, Y_j)$ as a function of covariates using data from group a. Instead of two working models, alternatively one may fit a model for M_i including $(A_i - \widehat{\pi})g(X_i)$ as covariates, where $g(X_i)$ includes 1 and other basis functions. Different strategies for fitting working models are further discussed later.

4. Practical Considerations

Theoretical results on covariate adjustment methods, including those studied in Benkeser et al, are mainly based on asymptotics. Although empirical performances have been evaluated in many simulation studies and real data analyses, questions and challenges remain in practice. Below we discuss factors affecting practical performances, attempting to address the question of how and when to use covariate adjustment. The main goal is to foster further discussions.

4.1 *Finite sample effects*

In addition to the strength of covariate associations with outcomes, an important factor affecting the degree of efficiency gain is sample size. For the same data generating scenario, ie, the same predictive strength of covariates and treatment effect, efficiency gain decreases with sample size and the effect can be quite large. We see this phenomenon in Benkeser et al as well, although treatment effects in scenarios with different sample sizes are not kept the same. Thus, 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. Therefore, one must take into account the sample size in planning covariate adjusted analysis and in anticipating realistic benefit.

A more useful perspective is to study better strategies on building working models for outcomes. In Benkeser et al adjustment was carried out by separately fitting two working models, one for each treatment group. Separate working models were used in implementing the augmentation approach of ZTD as well. Benkeser et al did not directly use working models for augmentation, but asymptotically it is equivalent to some working models in the augmentation framework. Therefore, our discussion below applies more generally. In randomized trials, estimating parameters in working models will not introduce additional variation asymptotically relative to when the limiting values are known. In finite samples it does matter, especially when there are many covariates relative to the sample size. Taking

the difference in means/risks as an example, there are three strategies for building working models. The first two strategies are to

(i) model $E(Y|A = a, X)$ separately for $a = 0, 1$;

(ii) model $E(Y|A, X)$, leaving out or partially including if needed, interactions of A and X, and make predictions for $E(Y|A = a, X)$ for each a.

Strategy (i) would be better when sample size is moderate or large. However it may lead to less efficiency gain in small samples and run into difficulty of even fitting the models. If there are many covariates and not much interactions with treatment, strategy (ii) may seem to have an advantage when sample size is small. Based on our simulations using same scenarios as in Table 2 of Benkeser et al, when $n = 100$, the improvement of strategy (ii) relative to (i) is very slight, about 1%(see below), where age was modeled as a categorical variable. Strategy (ii) does not realize the most efficiency gain when interaction does exist. We also note it barely makes any difference whether logistic or linear models were used.

Based on our experience, the strategy that has the best performance overall is to

(iii) model $E(M_i|X_i^*) = \zeta^T X_i^*$, where $M_i = A_i Y_i / \widehat{\pi} - (1 - A_i) Y_i / (1 - \widehat{\pi})$, $X_i^* = (A_i - \widehat{\pi}) g(X_i)$ and $g(X_i)$ is a vector including 1 and basis functions of X_i; e.g., $g(X_i) = (1, X_i)$.

We can fit the model by the OLS. This strategy is motivated by directly minimizing variance within a subclass. As strategy (ii) it reduces the number of estimated nuisance parameters, but it does not assume no treatment covariate interactions. Asymptotically it is equivalent to separately fitting working models but in small samples it can improve efficiency considerably. Using scenarios in Table 2 ($n = 100$, 1000 replicates), MSE of Benkeser et al and strategies (i-iii) are $(0.902, 0.932, 0.922, 0.834), (0.775, 0.820, 0.811, 0.744), (0.746, 0.789, 0.780, 0.723)$ respectively. Strategy (iii) improves about 10% relative to (i) and (ii), and improves up to 8% relative to Benkeser et al. The improvement relative to Benkeser et al is smaller because in Benkeser et al (see details in their code) the age level (1-7) was treated as a numeric variable,

but to help illustrate the point results for strategies (i-iii) are based on models where age was categorical with 6 levels (groups 1 and 2 combined). When age was modeled as numeric, results for strategies (i-ii) are similar to Benkeser et al and 2-3% better than when age was categorical. When $n = 1000$, strategy (iii) has a very slight advantage. Strategy (iii) can improve efficiency in situations when associations with outcomes are weak and n is small so that other methods even lose efficiency slightly. For simplicity our discussion focused on estimating difference in means. These strategies work for other estimands as well as long as the estimating function of an unadjusted analysis is known; see ZTD for details.

4.2 *Stratified Randomization*

The authors briefly touched the issue of stratified randomization, an area worth more future study. It is generally thought that it is important to adjust for stratifying variables in analysis. In our opinion, in terms of improving precision of estimation, adjusting for stratifying variables in the analysis actually is less useful than variables not used for stratification. The reason is that stratifying by prognostic variables at the design stage already reduces variability in estimation. As a result there would be not much room for further reducing variance by adjusting these variables in analysis. To illustrate this point, we conducted a simple simulation using the same settings as in Table 2 of Benkeser et al except that block randomization within each age group was used to assign treatment. Across all scenarios, for all sample sizes and for all methods, the relative efficiency in terms of MSE is close to 1. This is a rather extreme case as the only predictive variable is the age group and we stratify exactly by the same age groups. More realistically, within a stratum age is still predictive of outcomes and we may still expect some improvement by adjusting it in analysis. Certainly we do not intend to say that we should not adjust for stratifying variables. But this factor worth careful consideration in deciding variables to adjust in analysis, especially when sample size is small. Based on our experience and intuition, prognostic variables not used for

stratification at the design stage offer more utility in improving precision of estimation in analysis. The discussion above is from the perspective of improving estimation precision. A main argument for adjusting stratifying variables in the usual regression setting is that the unadjusted variance estimator and inference would be conservative and thus not realizing the efficiency benefit of stratified randomization, whereas adjusted inference mitigates the problem. This is a fair argument but this is from the perspective of how variance of estimation can be accurately estimated when stratified randomization is used. We think it is important to distinguish the different roles played by different variables used in adjustment.

4.3 *What Variables to Adjust*

One natural question in practice is what variables we should adjust. In principle all variables predictive of outcomes can be leveraged to improve efficiency. However, due to the finite sample effect, it is limited by the sample size. Technically the convergence rate of estimating nuisance parameters can be slower than the typical $n^{1/2}$ convergence rate. Thus one may include more variables in working models than what is judged to be appropriate in a usual regression setting. Yet, given the finite sample effect on efficiency, it is reasonable to use the general wisdom and experience on how many variables one may include in a working model given a sample size. In general, the stronger the association with outcomes, the more utility of the variable in improving efficiency. When sample size is small, one needs to prioritize variables or choose a functional form that lead to larger improvement in R square (more reduction in residual variance for outcomes) per one degree of freedom. For example, Tsiatis et al (2008) showed when n is small and nonlinear effect on outcomes exists but weak, including many higher order terms may not help much or even lose relative to including only linear terms.

Based on our previous discussion it also seems that one needs to prioritize adjusting for predictive variables that are not being stratified by at the design stage. This statement is

under the condition that variances of estimators can be correctly estimated through some way, for example, bootstrapping. More studies are much needed in this direction. Some people might think that one should only adjust for variables shown imbalance between treatment groups. We disagree with this and discourage the practice of choosing adjustment variables based on empirical evidence of imbalance. We agree with the authors that adjustment variables are better to be prespecified based on anticipated predictive power. When sample size allows, we also recommend the flexible strategy of Tsiatis et al (2008) and ZTD, which uses empirical evidence in working models to guide selection of covariates while avoiding fishing expedition.

5. Summary

As the authors, we are strong advocates of robust covariate adjustment in analyzing randomized clinical trials. Educating and encouraging pratitioners thinking beyond the traditional model-based regression for covariate adjustment is a key step. We thank the authors for their dedicated effort on this. We think a convenient and unified implementation strategy within a familiar estimating equation framework can help in this direction. Future research focusing on practical issues is greatly needed.

REFERENCES

Leon, S. Tsiatis, A. A., and Davidian, M. (2003). Semiparametric efficient estimation of treatment effect in a pretest-posttest study. *Biometrics* **59,** 1046–1055.

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

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