Estimating the efficiency gain of covariate-adjusted analyses in future clinical trials using external data

Mengqi Xu

May 10, 2024

1 Introduction

In clinical trials, observed data typically include three components: outcomes Y_i, treatment A_i, and covariates W_i, where $i = 1, ..., n$. Researchers commonly focus on estimating treatment effects such as the average treatment effect (ATE), defined as $\psi = E(Y|A = 1) - E(Y|A = 0)$. It is recommended to adjust for baseline covariates when estimating these effects, as this approach offers a more consistent estimator under less stringent assumptions and increases precision, thus reducing the sample size needed to achieve the desired statistical power. However, in practice, the use of covariate adjustment methods in the analysis of clinical trial data is often underutilized. This underutilization typically arises because, at the planning stage of a trial, investigators may have insufficient information about the potential efficiency gains—or relative efficiency—of covariate adjustment compared to unadjusted analyses.

As we plan a future clinical trial, a key consideration is determining the required sample size to achieve a desired power, which helps investigators assess the worthiness and methodology of the trial. The relative efficiency between two analytical methods can indicate their sample size ratio under identical power and type I error rates. Consequently, researchers have been developing methods to make use of external data sources, such as existing clinical trial data, pilot studies, and observational studies, to estimate the efficiency gains from covariate-adjusted estimators. One method involves estimating the precision gain by comparing the standard errors of adjusted and unadjusted estimators from previous clinical trials. This method, however, introduces uncertainty due to the sampling variability of the point estimate (Steingrimsson et al., 2017; Wang et al., 2019). Another method uses simulation studies based on data from an existing trial to estimate precision gains, allowing for more controlled data-generation processes but still affected by the variability of the existing trial data (Colantuoni & Rosenblum, 2015; Kahan et al., 2014). A third approach employs simulations using external datasets that reflect the covariate and outcome distributions of the control arm of the prospective trial, which also involves uncertainties from the simulation base dataset (Benkeser et al., 2021).

The challenge of making statistical inferences about the precision gains from covariate-adjusted estimators has not been comprehensively addressed in the literature. In (Li et al., 2023), the authors seek to fill this knowledge gap. They examine scenarios where the external data possess the same distribution of covariates and control-arm outcomes as the anticipated clinical trial, enabling the use of this external data to evaluate the efficiency gains of adjusted analyses over unadjusted ones. Specifically, under the sharp null hypothesis where the treatment effect is zero for all individuals, their approach can compute the relative efficiency, which is defined as the ratio of the asymptotic variances between the covariate-adjusted estimator and the unadjusted estimator, using only the external data. The authors also demonstrate that this relative efficiency under the sharp null hypothesis corresponds to the efficiency ratio under a local alternative hypothesis. (For more details, readers are referred to the appendix of the paper.)

In their study, Li et al. analyze both the working-model-based approach and the fully adjusted estimator for covariate adjustment. The working-model-based method involves fitting a parametric model conditioned on covariates within each treatment arm, followed by marginalizing over the pooled empirical distribution of covariates across arms (Benkeser et al., 2021; Moore et al., 2011). Although there is usually a risk of model misspecification, this approach often yields consistent and asymptotically normal estimators for the marginal effects of interest. However, a notable drawback of the working-model-based approach is its inefficiency when the model specification is incorrect. In contrast, the fully adjusted model approach, which utilizes either correctly specified parametric or non-parametric methods, consistently provides estimations that are asymptotically efficient under the right regularity conditions. An example of this is the AIPW estimator used for calculating the average treatment effect (ATE).

In the following part of the report, the first half will cover definitions, settings, Li et al.'s method, and their simulation results. In the second half, we extend the setting to an alternative hypothesis, conduct a simulation, and provide a discussion.

2 Influence Function and Relative Efficiency in Clinical Trials

Let $\mathbf{X}^t = \{X_1^t, \ldots, X_{n_t}^t\}$ denote the sample of size n^t from a future clinical trial with some distribution ν, and $\psi(v)$ denote the treatment effect measure. (The superscript 't' means they are from the future trial.) Let $\hat{\psi}$ denote an estimator of $\psi(\nu)$ constructed from $\mathbf{X}^t$, and suppose the estimator is regular and asymptotically linear (RAL), so that

$$\hat{\psi} - \psi(\nu) = \frac{1}{n^t} \sum_{i=1}^{n^t} D_\nu(X_i^t) + o_P((n^t)^{-1/2}),$$

where D_ν is called the **influence function** of $\hat{\psi}$. Letting σ_ν^2 denote the variance of $D_\nu(X^t)$ when $X^t \sim \nu$, we see that $\hat{\psi}$ is consistent and asymptotically normal:

$$\sqrt{n^t}\{\hat{\psi} - \psi(\nu)\} \xrightarrow{d} N(0, \sigma_\nu^2). \tag{1}$$

In a randomized clinical trial, we want to conduct a hypothesis test on the marginal treatment effect:

$$H_0 : \psi = 0 \text{ v.s. } H_0 : \psi \neq 0$$

Given a chosen value of the power and type I error, the required sample size is specified. Suppose a Wald test based on an asymptotically normal estimator as in Eq. (1), then it could be shown that the sample size is proportional to σ_v^2. This gives us insight that the **relative efficiency**

$$\phi := \tilde{\sigma}_v^2 / \sigma_v^2$$

can reflect the sample size ratio between two RAL estimators with asymptotic variance $\tilde{\sigma}_v^2$ and σ_v^2. For detailed discussion, please refer to the Supplementary Material of the paper.

3 Continuous Outcomes

Now let's consider a simple setting, when the outcome Y is **continuous and fully observed** in the trial.

3.1 ATE Estimators

Suppose the observation unit in a future randomized trial is $X^t = (Y^t, A, W^t)$, where we assume that

1. Treatment is independent of covariates ($A \perp W^t$).
2. Treatment probability $\Pi(A = 1) \in (0, 1)$.
3. Both Y^t and W^t have bounded support.

We are interested in the average treatment effect (ATE) $\psi = E(Y|A = 1) - E(Y|A = 0)$.

- The **unadjusted estimator** is the mean difference of the outcome between the treatment and controlled arms:

$$\hat{\psi}_u = \sum_{i=1}^{n^t} A_i Y_i^t / \sum_{i=1}^{n^t} A_i - \sum_{i=1}^{n^t} (1 - A_i) Y_i^t / \sum_{i=1}^{n^t} (1 - A_i).$$

- The **fully adjusted estimator** is the augmented inverse probability weighted (AIPW) estimator:

$$\hat{\psi}_f = \frac{1}{n^t} \sum_{i=1}^{n^t} \left[\frac{A_i\{Y_i^t - \hat{r}_1(W_i^t)\}}{\hat{\pi}(W_i^t)} - \frac{(1 - A_i)\{Y_i^t - \hat{r}_0(W_i^t)\}}{1 - \hat{\pi}(W_i^t)} + \hat{r}_1(W_i^t) - \hat{r}_0(W_i^t) \right],$$

  where $\hat{r}_a(w)$ is an estimator of $r_a(w) := E[Y^t \mid A = a, W^t = w]$ and $\hat{\pi}$ is an estimator of the treatment mechanism $\pi(w) := P(A = 1 \mid W^t = w)$.

- As for the **working-model-based** adjusted estimator, we fit arm-specific linear models on the outcome Y and covariate W, i.e. $E[Y^t \mid A = a, W^t = w] = \alpha_a + \beta_a^\top w$. Then the estimator of ATE is:

$$\hat{\psi}_m = \hat{\alpha}_1 - \hat{\alpha}_0 + (\hat{\beta}_1 - \hat{\beta}_0)^\top \sum_{i=1}^{n^t} \frac{W_i^t}{n^t}.$$

3.2 Relative Efficiency Estimation

Now we want to quantify the relative efficiency of the adjusted (fully adjusted and working-model-based) and unadjusted estimators under a **sharp null** distribution, i.e. the treatment has no effect. Note that our quantities of the relative efficiencies under the sharp null also correspond to the relative efficiencies under local alternatives, where the treatment effect is not zero but not too far away. (Li et al. 2023)

In the future trial data, under the sharp null, the joint distribution of (W^t, Y^t), denoted as P, is independent of treatment A. Consider an external data $X = (W, Y)$ which can accurately reflects the distribution of W^t and Y^t in future trials where the treatment has no effect, i.e. the external data $X = (W, Y)$ also has distribution P. These external data may come from the control arm in a previous trial or from an observational study.

Then we could estimate the relative efficiency for the future trial using the external data, by estimating the asymptotic variances of the adjusted and unadjusted estimator in section 3.1 and then take the ratio. According to the Lemma 3.1 in the paper, when $(W, Y) \sim P$, the asymptotic variances of $\hat{\psi}_u, \hat{\psi}_f$, and $\hat{\psi}_m$ are $\sigma_u^2 = Var_P(Y)$, $\sigma_f^2 = E_P(Var_P(Y|W))$, and $\sigma_u^2 = E_P(Y - \alpha^* - W^T\beta^*)$, where $(\alpha^*, \beta^*) = argmin_{\alpha,\beta \in R \times R^d} E_P[(Y - \alpha - W^\top \beta)^2]$.

Using external data $\{(W_i, Y_i), i = 1, \ldots, n\}$, σ_u^2, σ_f^2, and σ_m^2 are estimated by: (Here we ignore the proof. Readers can refer to the paper and its supplement.)

- $\hat{\sigma}_u^2 = \sum_{i=1}^n (Y_i - \bar{Y})^2/n$, where $\bar{Y} = \sum_{i=1}^n Y_i/n$. (Sample variance)
- $\hat{\sigma}_f^2 = \sum_{i=1}^n \{Y_i - \hat{r}(W_i)\}^2/n$, where $\hat{r}(w)$ be an estimator of $r(w) := E_P[Y \mid W = w]$
- $\hat{\sigma}_m^2 = \sum_{i=1}^n (Y_i - \hat{\alpha} - W_i^\top \hat{\beta})^2/n$, where $(\hat{\alpha}, \hat{\beta})$ are the coefficients in the linear regression of Y on W

Finally, the relative efficiencies are estimated by

$$\hat{\phi}_f = \hat{\sigma}_f^2/\hat{\sigma}_u^2 \text{ and } \hat{\phi}_m = \hat{\sigma}_m^2/\hat{\sigma}_u^2.$$

4 Other Settings

According to Li et al., the above relative efficiency estimation framework also works in similar ways for other settings: (Readers can refer to the paper for detailed information since we won't go to details in the report.)

Ordinal outcomes: Li et al. considered three treatment effect estimands for ordinal outcomes: difference in means (DIM), Mann–Whitney estimand (MW) and average log odds ratio (LOR). The procedure of estimating relative efficiencies for DIM is same with that in continuous outcomes in Section 3, except using a treatment-specific proportional odds model as the working parametric model, instead of linear models. Readers can refer to the paper for detailed information about MW and LOR.

Bootstrap procedure for working-model-based estimators: Li et al. also proposed an automated double bootstrap procedure for statistical inference about relative efficiency of a new working-model-based adjusted estimator ϕ_m that avoids the need to perform the potentially-tedious analytic calculations.

Partially observed outcomes: Li et al. solve settings where the outcome in the trial is partially observed, by applying coarsening-at-random (Gill et al., 1997). A typical example is the time-to-event data.

5 Simulation and Result for Ordinal Outcomes

They conducted simulation studies on both ordinal and time-to-event outcomes. To illustrate their simulation mechanism, let's take an example of their simulation "Setting II" with ordinal outcomes. What they generated is the external data only for estimation as following steps:

1. Generate three continuous covariates W_1, W_2, W_3:

$$(W_1, W_2) \sim N_2(\mu = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Sigma = \begin{pmatrix} 1 & 0.3 \\ 0.3 & 1 \end{pmatrix}),$$

$$W_3 \sim \text{Uniform}(-1, 1), \text{ independent of } (W_1, W_2)$$

2. Generate ordinal outcome $Y \in \{1, 2, 3\}$:

$$P(Y \leq 1 \mid W_1 = w_1, W_2 = w_2, W_3 = w_3) = \text{expit}(1.5w_1 - 2w_3^2),$$

$$P(Y \leq 2 \mid W_1 = w_1, W_2 = w_2, W_3 = w_3) = \text{expit}(1.5w_1 + w_3^2).$$

They estimated the relative efficiency of both fully adjusted and working-model-based models. They considered three estimands of the treatment effect: DIM, MW and average LOR. For DIM method, the estimation of the asymptotic variance are almost same to the continuous setting in section 3:

- For **unadjusted analysis**, they use sample variance $\hat{\sigma}_u^2$
- For **fully adjusted** analysis: $\hat{\sigma}_f^2 = \sum_{i=1}^n \{Y_i - \hat{r}(W_i)\}^2/n$. To get non-parametric estimator of $r(w) := E_P[Y|W = w]$, they use Super Learner (van der Laan et al., 2007; Polley & Van Der Laan, 2010) implemented as in the R package *SuperLearner*
- For **working-model-based** analysis they use the proportional odds model instead of the linear model in Section 3, due to the ordinal outcome.

For MW and average LOR, readers can refer to their paper for detail.

They built Wald-type confidence intervals (built on the logit scale first and transformed to the original scale), in the form of $\hat{\phi} \pm \hat{var}(\hat{\phi})$, where $\hat{var}(\hat{\phi}) = \hat{var}(\frac{\hat{\sigma}^2}{\hat{\sigma}_u^2})$ was calculated using Delta method (readers can refer to the supplement of the paper for detail). They do 1,000 replications of the simulations, with 1,000 sample size. The results are shown in Table 1. Note that in Table 1, the "Truth" means the true relative efficiency under the sharp null, which was calculated by the analytical value or the estimator under large sample size.

Table 1: Simulation results for ordinal outcome: the relative efficiency of fully adjusted and working-model-based estimators for DIM, MW, and LOR. 'F' and 'W' stands for the fully adjusted and the working-model-based estimators, respectively.

	Truth	Bias	MSE	%RMSE	Coverage	CI width
DIM (F)	0.677	0.011	0.001	0.042	0.923	0.095
DIM (W)	0.681	-0.003	0.001	0.038	0.935	0.098
MW (F)	0.677	0.011	0.001	0.042	0.922	0.095
MW (W)	0.681	-0.002	0.001	0.038	0.934	0.098
LOR (F)	0.677	0.007	0.001	0.040	0.939	0.101
LOR (W)	0.682	-0.002	0.001	0.038	0.942	0.098

Table 1 showed that under the null, their estimations have biases around 0.00., and 95% confidence interval with over 92% coverage of the truth. In general, their estimations are relatively close to the true value under the sharp null.

6 Our Investigation

However, a basic assumption of Li et al.'s framework is the sharp null distribution, when treatment doesn't actually involve even in the true value. Would their estimation framework of the relative efficiencies using external "controlled data" still work well under the alternative, i.e. when there exists treatment effect in the trial?

To investigate this question, we conducted a new simulation, aiming to explore the performance of the above estimation framework under the alternative. Focusing on the simple continuous outcome, we generated the true values of ϕ_f and ϕ_m under the alternative by simulating a clinical trial data with treatment effect exists.

6.1 Simulation Mechanism

Below is the future clinical trial data generation mechanism:

- Generate three continuous covariates $(W_1, W_2, W_3) \sim N_3(\mu = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \Sigma = \begin{pmatrix} 1 & 0.3 & 0 \\ 0.3 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix})$
- Generate treatment $A \sim$ Bernoulli(0.5)
- Generate outcome $Y = (W_1 + 2W_2 + 3W_3^2) + A * m|W_1| + \epsilon$, where $\epsilon \sim N(0, 1)$ and m is a scalar that controls the amount of treatment effect.

Our treatment effect estimand is **ATE**, so the unadjusted, fully adjusted and working-model-based estimators are all same with that in Section 3. Since $|W_1| \geq 0$, the treatment effect increases as m increases, with ATE $= E(m|W_1|)$. This means we move further from the sharp null as m increases.

To calculate the true value of the the asymptotic variances of ATE estimators, considering the asymptotic property of the Monte Carlo estimator, we use sample variances of their influence functions, under a large enough sample size $n = 10^5$. Applying the delta method:

- The influence function of the unadjusted estimator $\hat{\psi}_u$ is

$$D_u(w^t, a^t, y^t) = \left(\frac{a^t}{\pi_1} - \frac{1 - a^t}{\pi_0} \right) \left\{ u(y^t) - E[u(Y^t) \mid A = a^t] \right\},$$

- The influence function of fully-adjusted AIPW estimator $\hat{\psi}_f$ is

$$D_f(w^t, a^t, y^t) = \frac{\{u(y^t) - r_1(w^t)\}a^t}{\pi_1} + r_1(w^t) - \frac{\{u(y^t) - r_0(w^t)\}(1 - a^t)}{\pi_0} - r_0(w^t) - \psi,$$

  where $\hat{r}_1(w)$ and $\hat{r}_0(w)$ are estimators of $r_a(w) = E[u(Y^t) \mid A = a, W^t = w]$.

- The working-model-based influence function is

$$D_m(w^t, a^t, y^t) = D_m(w^t, 1, y^t) - D_m(w^t, 0, y^t),$$

  where $D_m(w^t, a, y^t) = (\frac{a^t}{\pi_1} + \frac{1-a^t}{\pi_0}) \left\{ y^t - \alpha_a^* - (\beta_a^*)^\top w^t \right\} + \alpha_a^* + (\beta_a^*)^\top w^t - \mu_a$, $\pi_a = \Pi(A = a)$, and $\mu_a = E(Y|A = a)$.

Given simulated data $(W_i^t, A_i^t, Y_i^t), i = 1, ..., n$, we calculated the value of influence function D_i for each sample, and calculate their sample variance as asymptotic values of $\sigma_u^2, \sigma_f^2, \sigma_m^2$. Then the asymptotic true relative efficiencies are ratios $\phi_f = \sigma_f^2/\sigma_u^2$ and $\phi_m = \sigma_m^2/\sigma_u^2$.

We calculate the estimators of relative efficiencies under the sharp null using this paper's method, i.e. using the controlled data where $Y = (W_1 + 2W_2 + 3W_3^2) + \epsilon$. The calculation of 95% CI is similar with their simulations in Section 5.

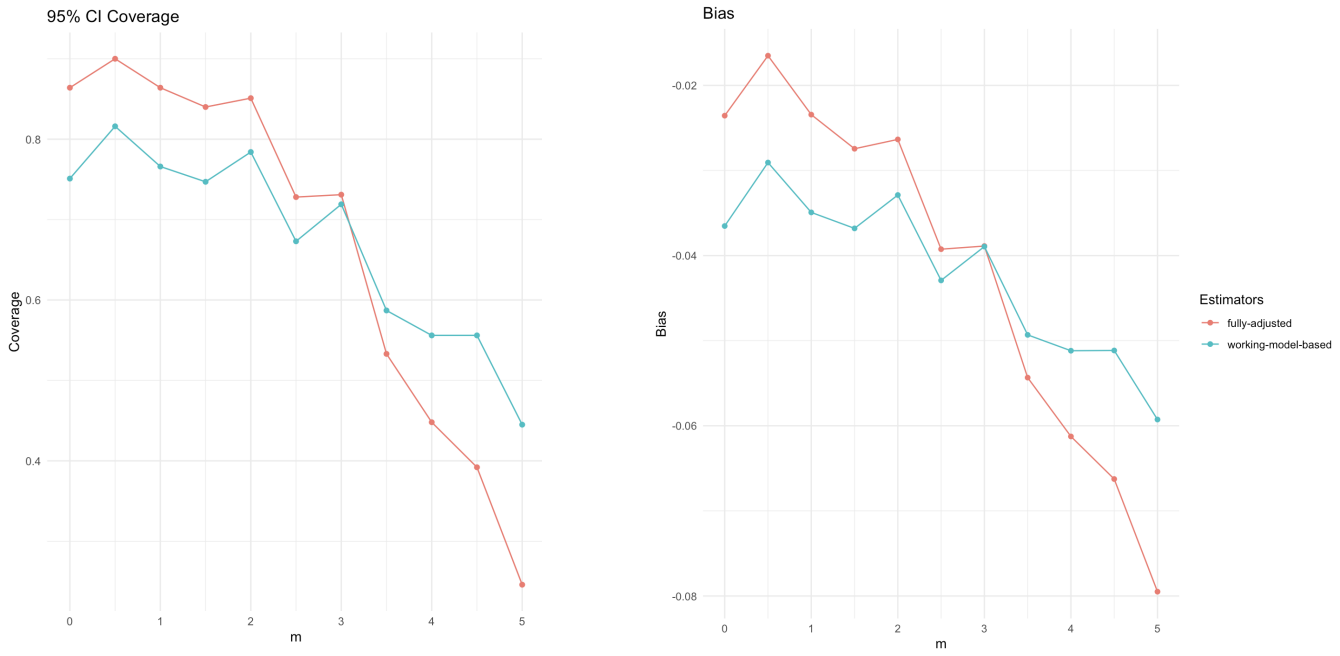
6.2 Result

We gradually increase the value of m from 0 to 5, and draw plots of the 95% CI coverage and bias of the estimator of relative efficiency in Figure 1. We noticed that:

- In general, as m increasing, the 95% CI coverage decreases from about 0.8 to nearly 0, and the bias becomes larger (negatively).
- At the local alternative, when $m \leq 2$, the accuracy of the estimator doesn't have significant change. In contrast, when $m > 2$, the accuracy decreases much faster.

This implies that as treatment effect goes larger, i.e. we move further from the null, their estimator of the relative efficiency based on the null distribution work worse. In addition, their estimation works well under local alternative when there are small perturbations to the underlying distribution, but doesn't work when the treatment effect becomes more obvious.

Figure 1: Simulation result for continuous data when the true distribution is under the alternative.



7 Discussions

When a future clinical trial involves a large collection of baseline covariates, investigators may wish to select a smaller subset for adjustment in the analysis. Li et al.'s method can compare the relative efficiency of different adjusted estimators that use various sets of covariates. In scenarios where certain covariates are costly to measure, the relative efficiency estimate can help balance the cost of measuring these covariates against recruiting more participants to achieve the desired power.

In Li et al., the relative efficiency they considered is based on a sharp null setting where the treatment has no effect. Their estimation works also well under local alternative when there are small treatment effect. However, according to our further investigation, when there exists large treatment effect, their estimation doesn't work. An alternative approach would involve specifying a particular alternative distribution.

Another limitation of their method is that it relies solely on the external dataset. This reliance may subject the estimation to sample variability, especially if the external dataset has a small sample size.

References

Steingrimsson JA, Hanley DF, & Rosenblum M (2017), *Improving precision by adjusting for prognostic baseline variables in randomized trials with binary outcomes, without regression model assumptions*, Contemporary Clinical Trials.

Wang B, Ogburn EL, & Rosenblum M (2019), *Analysis of covariance in randomized trials: More precision and valid confidence intervals, without model assumptions*, Biometrics.

Colantuoni E, & Rosenblum M (2015), *Leveraging prognostic baseline variables to gain precision in randomized trials*, Statistics in Medicine.

Kahan BC, Jairath V, Doré CJ, & Morris TP (2014), *The risks and rewards of covariate adjustment in randomized trials: An assessment of 12 outcomes from 8 studies*, Trials.

Benkeser D, Díaz I, Luedtke A, Segal J, Scharfstein D, & Rosenblum M (2021), *Improving precision and power in randomized trials for Covid-19 treatments using covariate adjustment, for binary, ordinal, and time-to-event outcomes*, Biometrics.

Xiudi L, Sijia L, & Alex L (2023), *Estimating the efficiency gain of covariate-adjusted analyses in future clinical trials using external data*, Journal of the Royal Statistical Society Series B: Statistical Methodology.

Moore KL, Neugebauer R, Valappil T, & van der Laan MJ (2011), *Robust extraction of covariate information to improve estimation efficiency in randomized trials*, Statistics in Medicine.

Gill RD, van der Laan MJ, & Robins JM (1997), *Coarsening at random: Characterizations, conjectures, counter-examples*, In Proceedings of the First Seattle Symposium in Biostatistics.

Van der Laan MJ, Polley EC, & Hubbard AE (2007), *Super learner*, Statistical Applications in Genetics and Molecular Biology.

Polley EC, & Van Der Laan MJ (2010), *Super learner in prediction*, U.C. Berkeley Division of Biostatistics Working Paper Series.