A Bias Correction Method in Meta-analysis of Randomized Clinical Trials with no Adjustments for Zero-inflated Outcomes: Supplementary Materials

Zhengyang Zhou, Ph.D.

Department of Biostatistics and Epidemiology

University of North Texas Health Science Center, Fort Worth, TX

Minge Xie, Ph.D.

Department of Statistics

Rutgers University, Piscataway, NJ

David Huh, Ph.D.

School of Social Work

University of Washington, Seattle, WA

Eun-Young Mun, Ph.D.

Department of Health Behavior and Health Systems

University of North Texas Health Science Center, Fort Worth, TX

Correspondence should be sent to:

Zhengyang Zhou, Ph.D.

Email: zhengyang.zhou@unthsc.edu

Eun-Young Mun, Ph.D.

Email: eun-young.mun@unthsc.edu

Appendix A. Analogue between the ZIBC method and EM algorithm

We can obtain Equation (5) following the idea of the EM algorithm. Denote $\log L_{CV}(\boldsymbol{\beta}) \triangleq \sum_{i=1}^{n}\{-\exp(\mathbf{x}_i^t\boldsymbol{\beta}) + y_i\mathbf{x}_i^t\boldsymbol{\beta}\}$ as the log-likelihood function of $\boldsymbol{\beta}$ under conventional model minus a constant, the analogous expectation and maximization steps are shown as follows.

- (Analogous) expectation-step:

  Taking the expectation of $\log L_{CV}(\boldsymbol{\beta})$, we have

$$\begin{aligned}
\mathbb{E}[\log L_{CV}(\boldsymbol{\beta})] &= \sum_{i=1}^{n}\{-\exp(\mathbf{x}_i^t\boldsymbol{\beta}) + \mathbb{E}[y_i]\cdot\mathbf{x}_i^t\boldsymbol{\beta}\} \\
&= \sum_{i=1}^{n}\{-\exp(\mathbf{x}_i^t\boldsymbol{\beta}) + (1-\pi_i)\exp(\mathbf{x}_i^t\boldsymbol{\beta}^0)\cdot\mathbf{x}_i^t\boldsymbol{\beta}\} \\
&\approx \sum_{i=1}^{n}\{-\exp(\mathbf{x}_i^t\boldsymbol{\beta}) + (1-\bar{\pi})\exp(\bar{\mathbf{x}}^t\boldsymbol{\beta}^0)\cdot\mathbf{x}_i^t\boldsymbol{\beta}\} \\
&\triangleq Q(\boldsymbol{\beta}).
\end{aligned}$$

- (Analogous) maximization step:

  To maximize $Q(\boldsymbol{\beta})$, we derive its first derivative with respect to $\boldsymbol{\beta}$, which is given by

$$\begin{aligned}
\frac{\partial Q(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}} &= \sum_{i=1}^{n}\{-\exp(\mathbf{x}_i^t\boldsymbol{\beta}) + (1-\bar{\pi})\exp(\bar{\mathbf{x}}^t\boldsymbol{\beta}^0)\}\mathbf{x}_i \\
&= \left[(1-\bar{\pi})\exp(\bar{\mathbf{x}}^t\{\boldsymbol{\beta}^0-\boldsymbol{\beta}\}) - 1\right]\sum_{i=1}^{n}\exp(\mathbf{x}_i^t\boldsymbol{\beta})\mathbf{x}_i.
\end{aligned}$$

Note that $\boldsymbol{\beta}^*$ is the solution of $S_{\mathrm{CV}}(\boldsymbol{\beta}) = 0$. By plugging $\boldsymbol{\beta}^*$ in $\frac{\partial Q(\boldsymbol{\beta})}{\partial\boldsymbol{\beta}}$, we obtain Equation (5).

Appendix B. Proof of Lemma 1

Proof. Consider three "average" subjects in the control group, intervention group, and the overall sample (denoted as $\{\text{average}, C\}$, $\{\text{average}, T\}$, and $\{\text{average}\}$), with $\mathbf{x}_{\text{average},C,p-2} = \bar{\mathbf{x}}_{C,p-2}$, $\mathbf{x}_{\text{average},T,p-2} = \bar{\mathbf{x}}_{T,p-2}$, and $\mathbf{x}_{\text{average},p-2} = \bar{\mathbf{x}}_{p-2}$, respectively. Without loss of generality, assuming that observed covariates excluding the intervention assignment are grand mean centered before data analysis, we have $\mathbf{x}_{\text{average},p-2} = \bar{\mathbf{x}}_{p-2} = \mathbf{0}$. Since $\bar{\mathbf{x}}_{C,p-2} = \bar{\mathbf{x}}_{T,p-2}$, we also have $\mathbf{x}_{\text{average},C,p-2} = \mathbf{x}_{\text{average},T,p-2} = \mathbf{x}_{\text{average},p-2} = \mathbf{0}$. Therefore, we have

$$\log(\mu_{\text{average},C}) = \beta^0_{0,C}$$
$$\log(\mu_{\text{average},T}) = \beta^0_{0,T} + \beta^0_{1,T}$$
$$\log(\mu_{\text{average}}) = \beta^0_0 + \beta^0_1 \mathbb{1}_{\{A_{\text{average}}=T\}}$$

under the true method. If an average subject in the overall sample belongs to the control group, then

$$\begin{aligned} &\log(\mu_{\text{average}}) = \log(\mu_{\text{average},C}) \\ \Rightarrow &\beta^0_{0,C} = \beta^0_0 \\ \Rightarrow &\hat{\beta}_{0,C,\text{MLE}} \approx \hat{\beta}_{0,MLE}. \end{aligned} \tag{A.1}$$

Similarly, if an average subject belongs to the intervention group, then

$$\begin{aligned} &\log(\mu_{\text{average}}) = \log(\mu_{\text{average},T}) \\ \Rightarrow &\beta^0_{0,T} + \beta^0_{1,T} = \beta^0_0 + \beta^0_1 \\ \Rightarrow &(\widehat{\beta_0 + \beta_1})_{T,\text{MLE}} \approx \hat{\beta}_{0,MLE} + \hat{\beta}_{1,MLE}. \end{aligned} \tag{A.2}$$

Under similar arguments, for the conventional method, we have

$$\hat{\beta}_{0,C,\text{CV}} \approx \hat{\beta}_{0,\text{CV}} \tag{A.3}$$

and

$$(\widehat{\beta_0 + \beta_1})_{T,\text{CV}} \approx \hat{\beta}_{0,CV} + \hat{\beta}_{1,CV}. \tag{A.4}$$

Plug Equations (A.1) and (A.3) into Equation (8), and plug Equations (A.2) and (A.4) into Equation (9), we have

$$\begin{aligned} \hat{\beta}_{0,\text{MLE}} &\approx \hat{\beta}_{0,\text{CV}} - \log(1 - \bar{\pi}_C) \\ \hat{\beta}_{0,\text{MLE}} + \hat{\beta}_{1,\text{MLE}} &\approx \hat{\beta}_{0,\text{CV}} + \hat{\beta}_{1,\text{CV}} - \log(1 - \bar{\pi}_T), \end{aligned} \tag{A.5}$$

which directly gives Equation (10). □

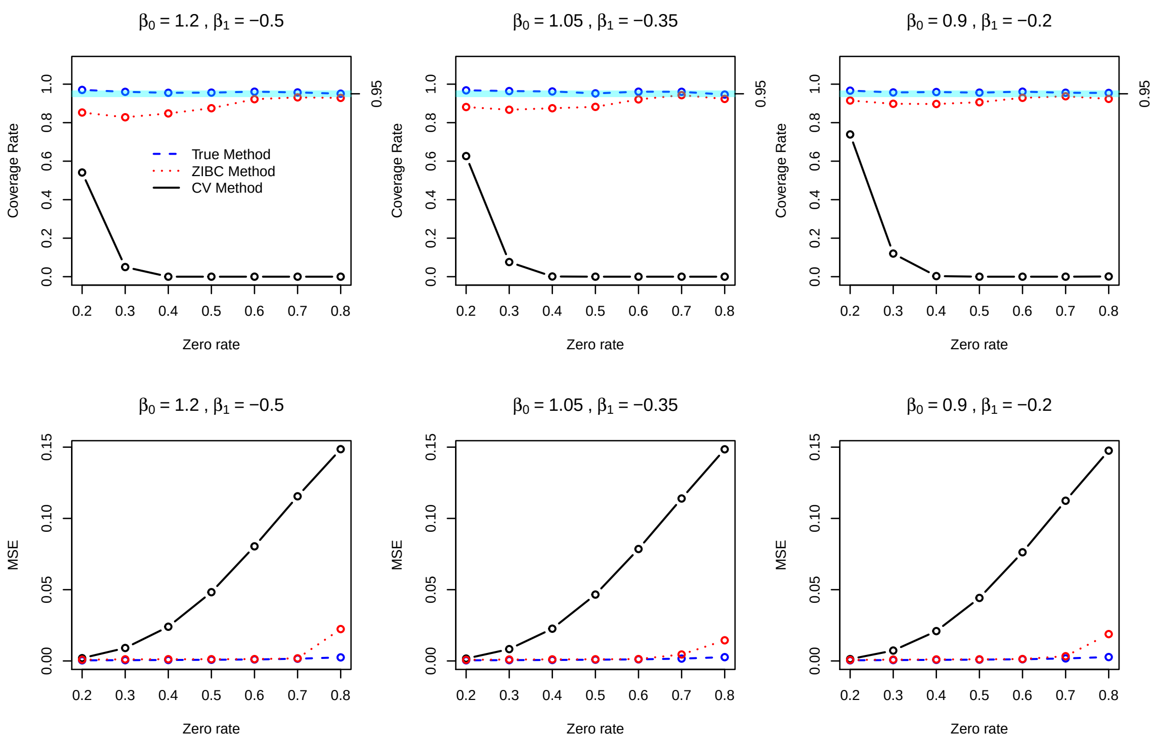

Figure S1: Coverage rates and MSE values of the true (blue dashed line), ZIBC (red dotted line) and conventional (black solid line) methods from 1000 replications (K=16)

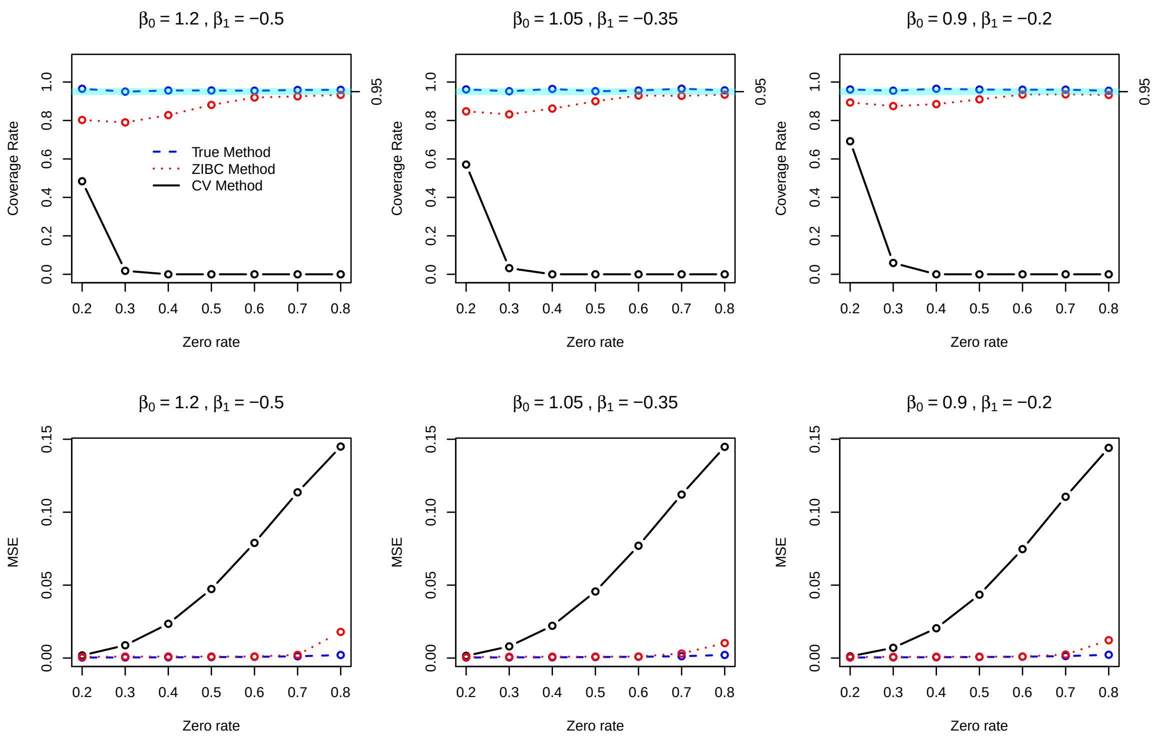

Figure S2: Coverage rates and MSE values of the true (blue dashed line), ZIBC (red dotted line) and conventional (black solid line) methods from 1000 replications (K=20)