Supporting Information for
"Two Weights Make a Wrong: Cluster Randomized Trials with Variable Cluster Sizes and Heterogeneous Treatment Effects"

Appendix A

Convergence of different estimators obtained from the four analytic approaches

Suppose we have n clusters and m_i is the cluster size. Let Y_{ij} denote the outcome for participant $j = 1, \ldots, m_i$ in cluster $i = 1, \ldots, n$, and Z_i denote the cluster-level treatment indicator. Assume the identity link, and covariates only include an intercept and the cluster-level treatment indicator Z_i. Let $\mu_i = \mathrm{E}(Y_{ij}|Z_i)$ be the marginal mean outcome given Z_i, which is specified via the following generalized linear model

$$\mu_i = \beta_0 + \beta_1 Z_i.$$

The marginal variance function is specified as $\mathrm{var}(Y_{ij}|Z_i) = \phi\nu_i$, where ϕ is the dispersion parameter and ν_i is a variance function that depends on the marginal mean.

For each cluster, let $\boldsymbol{Y}_i = (Y_{i1}, \ldots, Y_{im_i})'$ and $\boldsymbol{\mu}_i = (\mu_i, \ldots, \mu_i)'$ be the $m_i \times 1$ vector of outcomes and $m_i \times 1$ vector of marginal means, respectively; let $\boldsymbol{X}_i$ be the $m_i \times 2$ covariate matrix with all ones in the first column and all Z_i's in the second column. We use the generalized estimating equations (GEE) approach (Liang and Zeger, 1986) to estimate the treatment effect in the above mean model. Define $\boldsymbol{D}_i = \partial\boldsymbol{\mu}_i/\partial\boldsymbol{\beta}'$, where $\boldsymbol{\beta} = (\beta_0, \beta_1)'$, and let $\boldsymbol{V}_i = \boldsymbol{A}_i^{1/2}\boldsymbol{R}_i\boldsymbol{A}_i^{1/2}$ be a working covariance matrix for $\boldsymbol{Y}_i$, where $\boldsymbol{A}_i$ is a m_i-dimensional diagonal matrix with elements of $\phi\nu_i$, and $\boldsymbol{R}_i$ is a working correlation matrix. In our case, $\boldsymbol{R}_i$ can be an independence working correlation, i.e. $\boldsymbol{R}_i = \boldsymbol{I}_{m_i}$ where $\boldsymbol{I}_u$ is a $u \times u$ identity matrix, or an exchangeable working correlation, i.e. $\boldsymbol{R}_i = (1-\rho)\boldsymbol{I}_{m_i} + \rho\boldsymbol{J}_{m_i}$ where ρ is the intraclass correlation coefficient (ICC) and $\boldsymbol{J}_u = \boldsymbol{1}_u\boldsymbol{1}_u'$ is a matrix of ones.

The GEE estimator $\hat{\boldsymbol{\beta}}$ is obtained by solving

$$\sum_{i=1}^{n} \boldsymbol{D}_i' \boldsymbol{V}_i^{-1} \left(\boldsymbol{Y}_i - \boldsymbol{\mu}_i \right) / w_i = \boldsymbol{0}, \tag{1}$$

where w_i is a cluster-specific weight, equal to 1 for no weighting or m_i for inverse cluster size weight. Since we focus on the marginal mean model parameters, the details of variance estimators are omitted here. For the rest of the derivations, we assume Y_{ij} is continuous.

1. IEE (independence working correlation matrix with no cluster size weighting)

Under IEE, we have $\boldsymbol{D}_i = \boldsymbol{X}_i$, $\boldsymbol{V}_i = \phi \boldsymbol{I}_{m_i}$, and $w_i = 1$ in Equation (1), which becomes

$$\begin{aligned}
&\sum_i \begin{pmatrix} 1 & \dots & 1 \\ Z_i & \dots & Z_i \end{pmatrix} \begin{pmatrix} Y_{i1} - \mu_i \\ \dots \\ Y_{im_i} - \mu_i \end{pmatrix} = \boldsymbol{0}, \\
\Rightarrow &\sum_i \begin{pmatrix} 1 \\ Z_i \end{pmatrix} \sum_j (Y_{ij} - \mu_i) = \boldsymbol{0}, \\
\Rightarrow &\sum_i \begin{pmatrix} 1 \\ Z_i \end{pmatrix} \sum_j (Y_{ij} - \beta_0 - \beta_1 Z_i) = \boldsymbol{0}.
\end{aligned}$$

By solving the above equation, we obtain

$$\begin{aligned}
\hat{\beta}_0 &= \frac{\sum_i (1 - Z_i) \sum_j Y_{ij}}{\sum_i (1 - Z_i) m_i}, \\
\hat{\beta}_1 &= \frac{\sum_i Z_i \sum_j Y_{ij}}{\sum_i Z_i m_i} - \frac{\sum_i (1 - Z_i) \sum_j Y_{ij}}{\sum_i (1 - Z_i) m_i}.
\end{aligned}$$

Therefore, we have

$$\lim_{n \to \infty} \hat{\beta}_1 = E\left\{ \frac{1}{E(m_i)} \sum_j \left[Y_{ij}(1) - Y_{ij}(0) \right] \right\}.$$

2. IEEW (independence working correlation matrix with inverse cluster size weighting)

Similar to IEE, under IEEW, We have $\boldsymbol{D}_i = \boldsymbol{X}_i$, $\boldsymbol{V}_i = \phi \boldsymbol{I}_{m_i}$, and $w_i = m_i$ in Equation (1), which becomes

$$\sum_i \begin{pmatrix} 1 \\ Z_i \end{pmatrix} \sum_j \frac{Y_{ij} - \beta_0 - \beta_1 Z_i}{m_i} = \mathbf{0}.$$

By solving the above equation, we obtain

$$\hat{\beta}_0 = \frac{\sum_i \frac{1-Z_i}{m_i} \sum_j Y_{ij}}{\sum_i (1 - Z_i)},$$
$$\hat{\beta}_1 = \frac{\sum_i \frac{Z_i}{m_i} \sum_j Y_{ij}}{\sum_i Z_i} - \frac{\sum_i \frac{1-Z_i}{m_i} \sum_j Y_{ij}}{\sum_i (1 - Z_i)}.$$

Therefore, we have

$$\lim_{n\to\infty} \hat{\beta}_1 = E\left\{ \frac{1}{m_i} \sum_j \left[Y_{ij}(1) - Y_{ij}(0) \right] \right\}.$$

3. EEE (exchangeable working correlation matrix with no cluster size weighting)

Under EEE, We have $\boldsymbol{D}_i = \boldsymbol{X}_i$, $\boldsymbol{V}_i = \phi \left[(1-\rho) \boldsymbol{I}_{m_i} + \rho \boldsymbol{J}_{m_i} \right]$, and $w_i = 1$ in Equation (1), which becomes

$$\sum_i \begin{pmatrix} 1 & \dots & 1 \\ Z_i & \dots & Z_i \end{pmatrix} \begin{pmatrix} 1 & \rho & \dots & \rho \\ \rho & 1 & \dots & \rho \\ \rho & \rho & \dots & 1 \end{pmatrix}^{-1} \begin{pmatrix} Y_{i1} - \mu_i \\ \dots \\ Y_{im_i} - \mu_i \end{pmatrix} = \mathbf{0},$$
$$\Rightarrow \sum_i \begin{pmatrix} 1 \\ Z_i \end{pmatrix} \sum_j \frac{Y_{ij} - \mu_i}{1 + (m_i - 1)\rho} = \mathbf{0},$$

$$\Rightarrow \sum_i \begin{pmatrix} 1 \\ Z_i \end{pmatrix} \sum_j \frac{Y_{ij} - \beta_0 - \beta_1 Z_i}{1 + (m_i - 1)\,\rho} = \mathbf{0}.$$

By solving the above equation, we obtain

$$\hat{\beta}_0 = \frac{\sum_i \frac{1-Z_i}{1+(m_i-1)\rho} \sum_j Y_{ij}}{\sum_i \frac{m_i(1-Z_i)}{1+(m_i-1)\rho}},$$
$$\hat{\beta}_1 = \frac{\sum_i \frac{Z_i}{1+(m_i-1)\rho} \sum_j Y_{ij}}{\sum_i \frac{m_i Z_i}{1+(m_i-1)\rho}} - \frac{\sum_i \frac{1-Z_i}{1+(m_i-1)\rho} \sum_j Y_{ij}}{\sum_i \frac{m_i(1-Z_i)}{1+(m_i-1)\rho}}.$$

Therefore, we have

$$\lim_{n\to\infty} \hat{\beta}_1 = \frac{E\left\{\frac{1}{1+(m_i-1)\rho} \sum_j \left[Y_{ij}(1) - Y_{ij}(0)\right]\right\}}{E\left\{\frac{m_i}{1+(m_i-1)\rho}\right\}}.$$

4. EEEW (exchangeable working correlation matrix with inverse cluster size weighting)

Similar to EEE, under EEEW, We have $\boldsymbol{D}_i = \boldsymbol{X}_i$, $\boldsymbol{V}_i = \phi\left[(1-\rho)\boldsymbol{I}_{m_i} + \rho \boldsymbol{J}_{m_i}\right]$, and $w_i = m_i$ in Equation (1), which becomes

$$\sum_i \begin{pmatrix} 1 \\ Z_i \end{pmatrix} \sum_j \frac{Y_{ij} - \beta_0 - \beta_1 Z_i}{\left[1 + (m_i - 1)\,\rho\right] m_i} = \mathbf{0}.$$

By solving the above equation, we obtain

$$\hat{\beta}_0 = \frac{\sum_i \frac{1-Z_i}{[1+(m_i-1)\rho]m_i} \sum_j Y_{ij}}{\sum_i \frac{1-Z_i}{1+(m_i-1)\rho}},$$
$$\hat{\beta}_1 = \frac{\sum_i \frac{Z_i}{[1+(m_i-1)\rho]m_i} \sum_j Y_{ij}}{\sum_i \frac{Z_i}{1+(m_i-1)\rho}} - \frac{\sum_i \frac{1-Z_i}{[1+(m_i-1)\rho]m_i} \sum_j Y_{ij}}{\sum_i \frac{1-Z_i}{1+(m_i-1)\rho}}.$$

Therefore, we have

$$\lim_{n\to\infty} \hat{\beta}_1 = \frac{E\left\{\frac{1}{[1+(m_i-1)\rho]m_i}\sum_j [Y_{ij}(1) - Y_{ij}(0)]\right\}}{E\left\{\frac{1}{1+(m_i-1)\rho}\right\}}.$$

Appendix B

Web Table 1: Additional simulation results for continuous outcomes.

Scenario	Parameters	Results[a]	Methods			
			IEE[b]	IEEW[c]	EEE[d]	EEEW[e]
Homogeneous treatment effect						
1	$\beta_l = 2$	Mean Estimand	2.0000	2.0000	2.0000	2.0000
	$\beta_s = 2$	Mean Estimate	2.0025	2.0009	2.0014	2.0003
	$\sigma_b = 0.2$	Mean Bias	0.0025	0.0009	0.0014	0.0003
	$\sigma_e = 1$	Mean Relative Bias (%)	0.1274	0.0439	0.0725	0.0159
	$\rho = 0.038$	Empirical Coverage	0.9200	0.9430	0.9340	0.9210
2	$\beta_l = 2$	Mean Estimand	2.0000	2.0000	2.0000	2.0000
	$\beta_s = 2$	Mean Estimate	2.0047	2.0032	2.0040	2.0030
	$\sigma_b = 0.2$	Mean Bias	0.0047	0.0032	0.0040	0.0030
	$\sigma_e = 2$	Mean Relative Bias (%)	0.2341	0.1621	0.2006	0.1517
	$\rho = 0.010$	Empirical Coverage	0.9210	0.9460	0.9230	0.9320
Heterogeneous treatment effects						
3	$\beta_l = 4$	Mean Estimand	3.6659	3.0000	3.0431	2.3537
	$\beta_s = 2$	Mean Estimate	3.6505	2.9963	3.0363	2.3495
	$\sigma_b = 0.2$	Mean Bias	-0.0154	-0.0037	-0.0068	-0.0042
	$\sigma_e = 1$	Mean Relative Bias (%)	-0.4195	-0.1246	-0.2466	-0.1976
	$\rho = 0.038$	Empirical Coverage	0.9850	0.9860	0.9860	0.9740
4	$\beta_l = 4$	Mean Estimand	3.6659	3.0000	3.1757	2.4431
	$\beta_s = 2$	Mean Estimate	3.6526	2.9986	3.1603	2.3786
	$\sigma_b = 0.2$	Mean Bias	-0.0132	-0.0014	-0.0154	-0.0645
	$\sigma_e = 2$	Mean Relative Bias (%)	-0.3612	-0.0458	-0.5674	-2.7275
	$\rho = 0.010$	Empirical Coverage	0.9710	0.9750	0.9790	0.9620

[a] For each method, results of the mean bias, mean relative bias (%), and empirical coverage are relative to the corresponding theoretical estimand calculated using the expression provided in Table 1 of the main paper.

[b] IEE: Independence working correlation matrix with no cluster size weighting.

[c] IEEW: Independence working correlation matrix with inverse cluster size weighting.

[d] EEE: Exchangeable working correlation matrix with no cluster size weighting.

[e] EEEW: Exchangeable working correlation matrix with inverse cluster size weighting.

References

Liang, K. Y., & Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika, 73*(1), 13-22.