Supplemental materials for “Causal inference for treatment effects in partially nested designs”

Xiao Liu[1], Fang Liu[2], Laura Miller-Graff [2], Kathryn H. Howell[3], and Lijuan Wang[2]

[1]The University of Texas at Austin

[2]University of Notre Dame

[3]The University of Memphis

Correspondence should be addressed to Xiao Liu, Department of Educational Psychology, The University of Texas at Austin, Austin, TX 78712. Email: xiao.liu@austin.utexas.edu.

Supplemental materials for "Causal inference for treatment effects in partially nested designs"

Supplemental material for the technical details

Proofs for the identification results

The proofs for the identification results under IA1 and IA2 follow from those given in VanderWeele and Hernan (2013); below we restate them. Under IA1, we have $\mathbf{E}[Y_i(0)] = \mathbf{E}[\mathbf{E}[Y_i(0)|C_i]] = \mathbf{E}[\mathbf{E}[Y_i(0)|T_i = 0, C_i]]$, which is equal to $\mathbf{E}[\mathbf{E}[Y_i|T_i = 0, C_i]]$. Under IA2, $Y_i(1, K_i(1)) \amalg T_i|\{C_i, W_i\}$, we have

$$
\begin{aligned}
\mathbf{E}[Y_i(1, K_i(1))] &= \mathbf{E}[\mathbf{E}[Y_i(1, K_i(1))|C_i, W_i]] \\
&= \mathbf{E}[\mathbf{E}[Y_i(1, K_i(1))|T_i = 1, C_i, W_i]] \\
&= \mathbf{E}[\sum_k \mathbf{E}[Y_i(1, K_i(1))|K_i(1) = k, T_i = 1, C_i, W_i]p(K_i(1) = k|T_i = 1, C_i, W_i)] \\
&= \mathbf{E}[\sum_k \mathbf{E}[Y_i(1, k)|K_i = k, T_i = 1, C_i, W_i]p(K_i = k|T_i = 1, C_i, W_i)] \\
&= \mathbf{E}[\sum_k \mathbf{E}[Y_i|K_i = k, T_i = 1, C_i, W_i]p(K_i = k|T_i = 1, C_i, W_i)] \\
&= \mathbf{E}[\mathbf{E}[Y_i|T_i = 1, C_i, W_i]].
\end{aligned}
$$

The identification of $\mathbf{E}[Y_i(1, K_i(1))]$ under IA3, namely $Y_i(1, K_i(1)) \amalg T_i|C_i, V_i^*$, mimics that under IA2. Notice that W_i in IA2 is replaced with V_i^* in IA3. Based on this observation, under IA3, we have $\mathbf{E}[Y_i(1, K_i(1))] = \mathbf{E}[\mathbf{E}[Y_i|T_i = 1, C_i, V_i^*]]$.

Application of the IAs to the outcome modeling approaches

Linear regression. For the regression in Eq. 10 adjusting for $\mathbf{C}_i$ and $\mathbf{W}_i$ both, following the identification results under IA2, we have $\mathbf{E}\left[Y_i(1, K_i(1))\right] = \mathbf{E}[\mathbf{E}(Y_i|T_i = 1, \mathbf{C}_i, \mathbf{W}_i)] = b_0 + b_1 + \mathbf{E}(\mathbf{C}_i^\top)\mathbf{b}_c + \mathbf{E}(\mathbf{W}_i^\top)\mathbf{b}_w$, and under IA1,$\mathbf{E}\left[Y_i(0)\right] = \mathbf{E}[\mathbf{E}(Y_i|T_i = 0, C_i)] = b_0 + \mathbf{E}(\mathbf{C}_i^\top)\mathbf{b}_c$. Therefore, with the grand mean centered covariates (i.e., $\mathbf{E}(\mathbf{W}_i) = 0$), the average causal effect is $\mathbf{E}[\, Y_i(1, K_i(1))\,] - \mathbf{E}[\, Y_i(0))\,] = b_1$.

For the regression in Eq. 11 without adjusting for any covariates, when IA1 and IA2 hold without the need of conditioning on $\mathbf{C}_i$ or $\mathbf{W}_i$, we have $\mathbf{E}[Y_i(1, K_i(1))] = \mathbf{E}[Y_i|T_i = 1] = b_0 + b_1$ and $\mathbf{E}[Y_i(0)] = \mathbf{E}[Y_i|T_i = 0] = b_0$. Thus, the average causal effect is $\mathbf{E}\{Y_i(1, K_i(1))\} - \mathbf{E}\{Y_i(0)\} = b_1$.

For the regression in Eq. 12 adjusting for $\mathbf{C}_i$ only, when IA2 holds without the need of conditioning on $\mathbf{W}_i$, we have $\mathbf{E}[Y_i(1, K_i(1))] = \mathbf{E}[\mathbf{E}[Y_i|T_i = 1, \mathbf{C}_i]] = b_0 + b_1 + \mathbf{E}(\mathbf{C}_i^\top)\mathbf{b}_c$; and under IA1, $\mathbf{E}[Y_i(0)] = \mathbf{E}[\mathbf{E}[Y_i|T_i = 0, \mathbf{C}_i]] = b_0 + \mathbf{E}(\mathbf{C}_i^\top)\mathbf{b}_c$. Thus, the average causal effect is identified as $\mathbf{E}\{Y_i(1, K_i(1))\} - \mathbf{E}\{Y_i(0)\} = b_1$.

Multilevel modeling for partially nested data. For the MLM in Eqs. 13-14 adjusting for $\mathbf{C}_i$ and $\mathbf{W}_i$ both, under IA2, $\mathbf{E}[Y_i(1, K_i(1))] = \mathbf{E}[\mathbf{E}(Y_i|T_i = 1, \mathbf{C}_i, \mathbf{W}_i)] = b_0^{(1)} + \mathbf{E}(\mathbf{C}_i^\top)\mathbf{b}_c^{(1)} + \mathbf{E}(\mathbf{W}_i^\top)\mathbf{b}_w^{(1)}$ Under IA1, $\mathbf{E}[Y_i(0)] = \mathbf{E}[\mathbf{E}(Y_i|T_i = 0, \mathbf{C}_i)] = b_0^{(0)} + \mathbf{E}(\mathbf{C}_i^\top)\mathbf{b}_c^{(0)}$. Therefore, with grand mean centered $\mathbf{C}_i$ and $\mathbf{W}_i$ (i.e., $\mathbf{E}(\mathbf{C}_i) = \mathbf{E}(\mathbf{W}_i) = 0$), the average causal effect is $\mathbf{E}[\, Y_i(1, K_i(1))\,] - \mathbf{E}[\, Y_i(0))\,] = b_0^{(1)} - b_0^{(0)}$.

For the MLM in Eqs. 15-16 adjusting for $\mathbf{C}_i$ only, the cluster specific intercept $v_k^{(1)}$ plays the role of the latent cluster specific factor v_k^* in IA3. Hence, applying IA3, we have $\mathbf{E}[Y_i(1, K_i(1))] = \mathbf{E}[\mathbf{E}(Y_i|T_i = 1, C_i, V_i^*)]$. Then, using the condition that $V_i^* = v_k^*$ for individuals with $K_i(1) = k$ and that $K_i(1) = K_i$ for $T_i = 1$, we further have: $\mathbf{E}[Y_i(1, K_i(1))] = \mathbf{E}[\sum_k \mathbf{E}(Y_i|T_i = 1, K_i = k, V_i^* = v_k^*, C_i)p(V_i^* = v_k^*|C_i)]$ $= \mathbf{E}[\sum_k (b_0^{(1)} + \mathbf{C}_i^\top \mathbf{b}_c^{(1)} + v_k^{(1)})p(v_k^{(1)}|\mathbf{C}_i)]$, where the expectation is taken over the values of $\mathbf{C}_i$. With the regular multilevel modeling, $v_k^{(1)}$ is independent of $\mathbf{C}_i$ and has mean 0. Thus, we further have $\mathbf{E}[Y_i(1, K_i(1))] = b_0^{(1)} + \mathbf{E}(\mathbf{C}_i^\top)\mathbf{b}_c^{(1)}$. Under IA1, $\mathbf{E}[Y_i(0)] = \mathbf{E}[\mathbf{E}(Y_i|T_i = 0, \mathbf{C}_i)] = b_0^{(0)} + \mathbf{E}(\mathbf{C}_i^\top)\mathbf{b}_c^{(0)}$. Taken together, the average causal effect is $\mathbf{E}[\, Y_i(1, K_i(1))\,] - \mathbf{E}[\, Y_i(0))\,] = b_0^{(1)} - b_0^{(0)}$, with the grand mean centered $\mathbf{C}_i$.

For the MLM in Eqs. 17-18 that does not adjust for covariates but includes the cluster specific intercept $v_k^{(1)}$, when IA1 and IA3 both hold without the need of conditioning on $\mathbf{C}_i$, we have $\mathbf{E}[Y_i(1, K_i(1))] = \mathbf{E}[\mathbf{E}(Y_i|T_i = 1, V_i^*)]$

$= \mathbf{E}[\sum_k (b_0^{(1)} + v_k^{(1)}) p(v_k^{(1)})] = b_0^{(1)}$ under IA3; and $\mathbf{E}\left[Y_i(0)\right] = \mathbf{E}[\mathbf{E}(Y_i|T_i = 0)] = b_0^{(0)}$ under IA1. Hence, the average causal effect is $\mathbf{E}[\, Y_i(1, K_i(1))\,] - \mathbf{E}[\, Y_i(0))\,] = b_0^{(1)} - b_0^{(0)}$.

For the MLM in Eqs. 19-20 with conditioning on the post-treatment observed cluster mean $\bar{\mathbf{C}}_k$, as described in the main text, the average potential outcome $\mathbf{E}\left[Y_i(1, K_i(1))\right]$ cannot be identified as the parameter $b_0^{(1)}$ in this model, because the potential cluster mean of $\mathbf{C}_i$, which can be written as $\tilde{\mathbf{C}}_k = \frac{\sum_{i=1}^n \mathbf{C}_i \mathbf{1}(T_i=1, K_i(1)=k) + \sum_{i=1}^n \mathbf{C}_i \mathbf{1}(T_i=0, K_i(1)=k)}{\sum_{i=1}^n \mathbf{1}(T_i=1, K_i(1)=k) + \sum_{i=1}^n \mathbf{1}(T_i=0, K_i(1)=k)}$ (where $\mathbf{1}(\cdot)$ is the indicator function and n is the sample size), includes the $\mathbf{C}_i$ values in both the treatment and control groups, whereas the observed cluster mean $\bar{\mathbf{C}}_k = \frac{\sum_{i=1}^n \mathbf{C}_i \mathbf{1}(T_i=1, K_i=k)}{\sum_{i=1}^n \mathbf{1}(T_i=1, K_i=k)}$ only includes the $\mathbf{C}_i$ values in the treatment group and thus differs from the potential cluster mean $\tilde{\mathbf{C}}_k$ when the treatment assignment is affected by $\mathbf{C}_i$.

For the MLM in Eqs. 21-22 with conditioning on the post-treatment observed cluster mean $\bar{\mathbf{W}}_k$, $b_0^{(1)}$ in this model does not represent the average potential outcome $\mathbf{E}\left[Y_i(1, K_i(1))\right]$ when $\mathbf{W}_i$ affects the treatment assignment. The rationales are similar as those described in the above paragraph and hence are not repeated.

Proofs for the PS models

Under IA2,

$$
\begin{aligned}
\mathbf{E}\left[Y_i(1, K_i(1))\right] &= \mathbf{E}[\mathbf{E}(Y_i|T_i = 1, C_i, W_i)] \\
&= \int_{c,w} \mathbf{E}(Y_i|T_i = 1, C_i, W_i) f(C_i, W_i)\,\mathrm{d}c\mathrm{d}w \\
&= \int_{c,w} \mathbf{E}(Y_i|T_i = 1, C_i, W_i) f(C_i, W_i|T_i = 1) \frac{\Pr(T_i = 1)}{\Pr(T_i = 1|C_i, W_i)}\,\mathrm{d}c\mathrm{d}w \\
&= \int_{c,w} \mathbf{E}(Y_i|T_i = 1, C_i, W_i) f(C_i, W_i|T_i = 1) \Pr(T_i = 1) \frac{1}{\Pr(T_i = 1|C_i, W_i)}\,\mathrm{d}c\mathrm{d}w \\
&= \mathbf{E}(\frac{1}{e_i^{(1)}} Y_i|T_i = 1) = \mathbf{E}(\frac{T_i}{e_i^{(1)}} Y_i).
\end{aligned}
$$

Under IA1,

$$\begin{aligned}
\mathbf{E}\left[Y_i(0)\right] &= \mathbf{E}[\mathbf{E}(Y_i|T_i=0,C_i)] \\
&= \int_c \mathbf{E}(Y_i|T_i=0,C_i)f(C_i)\,\mathrm{d}c \\
&= \int_c \mathbf{E}(Y_i|T_i=0,C_i)f(C_i|T_i=0)\frac{\Pr(T_i=0)}{\Pr(T_i=0|C_i)}\,\mathrm{d}c \\
&= \int_c \mathbf{E}(Y_i|T_i=0,C_i)f(C_i|T_i=0)\Pr(T_i=0)\frac{1}{1-\Pr(T_i=1|C_i)}\,\mathrm{d}c \\
&= \mathbf{E}(\frac{1}{1-e_i^{(0)}}Y_i|T_i=0) = \mathbf{E}(\frac{1-T_i}{1-e_i^{(0)}}Y_i).
\end{aligned}$$

Thus, the average causal treatment effect is

$\mathbf{E}\left\{Y_i(1,K_i(1))\right\} - \mathbf{E}\left\{Y_i(0)\right\} = \mathbf{E}(\frac{T_i}{e_i^{(1)}}Y_i) - \mathbf{E}(\frac{1-T_i}{1-e_i^{(0)}}Y_i)$.

Technical details of the IPW adjustment

Let $L_i = (C_i, W_i)^\top$ and $\mathbf{a}^{(1)} = (\mathbf{a}_c^{(1)}, \mathbf{a}_w^{(1)})$. The estimating equations for $\mathbf{a}^{(1)} = (\mathbf{a}_c^{(1)}, \mathbf{a}_w^{(1)})$ and $\mathbf{a}^{(0)} = \mathbf{a}_c^{(0)}$ are those in Eqs. 28-29 of the main text; namely:

$$\begin{aligned}
S^1(\mathbf{a}^{(1)}) &= \sum_{i=1}^n S_i^1(\mathbf{a}^{(1)}) = \sum_{i=1}^n \frac{\partial}{\partial \mathbf{a}^{(1)}} \log p(T_i|L_i;\mathbf{a}^{(1)}) \\
&= \sum_{i=1}^n (T_i - e_i^{(1)})L_i^\top = 0
\end{aligned}$$

$$\begin{aligned}
S^0(\mathbf{a}^{(0)}) &= \sum_{i=1}^n S_i^0(\mathbf{a}^{(0)}) = \sum_{i=1}^n \frac{\partial}{\partial \mathbf{a}^{(0)}} \log p(T_i|C_i;\mathbf{a}^{(0)}) \\
&= \sum_{i=1}^n (T_i - e_i^{(0)})C_i = 0.
\end{aligned}$$

The regression-based IPW approach. The proposed IPW adjustment can be combined with a marginal regression model for the outcome ignoring the treatment clustering structure. Let $\mu^{(1)} = \mathbf{E}\left\{Y_i(1,K_i(1))\right\}$ and $\mu^{(0)} = \mathbf{E}\left\{Y_i(0)\right\}$. The estimating equations for μ_1 is

$$H^1(\mu_1) = \sum_{i=1}^n H_i^1(\mu_1) = \sum_{i=1}^n \frac{T_i}{e_i^{(1)}}(Y_i - \mu_1) = 0$$

and the estimating equations for μ_0 is

$$H^0(\mu_0) = \sum_{i=1}^n H_i^0(\mu_0) = \sum_{i=1}^n \frac{1-T_i}{1-e_i^{(0)}}(Y_i - \mu_0) = 0.$$

To account for uncertainty in the PS estimation, we simultaneously solve the system of estimating equations $U(\boldsymbol{\beta}) = (S^1(\mathbf{a}_c^{(1)}, \mathbf{a}_w^{(1)}), S^0(\mathbf{a}_c^{(0)}), H^1(\mu^{(1)}), H^1(\mu^{(0)})) = 0$, where $\boldsymbol{\beta} = (\mathbf{a}_c^{(1)}, \mathbf{a}_w^{(1)}, \mathbf{a}_c^{(0)}, \mu^{(1)}, \mu^{(0)})$ contains all parameters being estimated. Based on the theory of estimating equations, as $n \to \infty$, $(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \to N(\mathbf{0}, \mathbf{A}^{-1}\mathbf{B}(\mathbf{A}^\top)^{-1})$ where $\mathbf{A} = \mathbb{E}\left[\frac{\partial}{\partial \beta^\top} U(\boldsymbol{\beta})\right]$, $\mathbf{B} = \mathbb{E}\left[U(\boldsymbol{\beta})U(\boldsymbol{\beta})^\top\right]$. Let $\hat{\mathbf{A}} = \frac{\partial}{\partial \beta^\top} U(\boldsymbol{\beta}) \mid_{\beta = \hat{\beta}}$, which has the form has the form $\hat{\mathbf{A}} = \begin{pmatrix} \hat{A}_{11} & & & \\ & \hat{A}_{22} & & \\ \hat{A}_{31} & & \hat{A}_{33} & \\ & \hat{A}_{42} & & \hat{A}_{44} \end{pmatrix}$ with

$$\hat{A}_{11} = -\sum_{i=1}^{n} \hat{e}_i^{(1)}(1 - \hat{e}_i^{(1)}) L_i L_i^\top$$

$$\hat{A}_{22} = -\sum_{i=1}^{n} \hat{e}_i^{(0)}(1 - \hat{e}_i^{(0)}) C_i C_i^\top$$

$$\hat{A}_{31} = -\sum_{i=1}^{n} \frac{T_i}{\hat{e}_i^{(1)}}(1 - \hat{e}_i^{(1)})(Y_i - \hat{\mu}_1) L_i^\top$$

$$\hat{A}_{33} = -\sum_{i=1}^{n} \frac{T_i}{\hat{e}_i^{(1)}}$$

$$\hat{A}_{42} = -\sum_{i=1}^{n} \frac{1 - T_i}{1 - \hat{e}_i^{(0)}} \hat{e}_i^{(0)}(Y_i - \hat{\mu}_0) C_i^\top$$

$$\hat{A}_{44} = -\sum_{i=1}^{n} \frac{1 - T_i}{1 - \hat{e}_i^{(0)}}.$$

Let $\hat{\mathbf{B}} = U_i(\hat{\boldsymbol{\beta}})U_i(\hat{\boldsymbol{\beta}})^\top$. The asymptotic covariance matrix can be estimated by the sandwich-type covariance matrix $ACov(\hat{\boldsymbol{\beta}}) = \hat{\mathbf{A}}^{-1}\hat{\mathbf{B}}(\hat{\mathbf{A}}^\top)^{-1}$.

The MLM-based IPW approach. The weighted estimating equation for $\boldsymbol{\theta}^{(1)}$ is

$$Q^{W1}(\boldsymbol{\theta}^{(1)}) = \frac{\partial}{\partial \boldsymbol{\theta}^{(1)}} \sum_{k \in \{1,\ldots,J\}} \log \left\{ \int_{v_k} \exp\left[\sum_{i:K_i = k} \frac{T_i}{e_i^{(1)}} \log p(Y_i | T_i = 1, v_k) \right] p(v_k)\, dv_k \right\} = 0$$

Let $l_k = \int_{v_k} \exp\left[\sum_{i:K_i=k} w_i \log p(Y_i|T_i = 1, v_k)\right] p(v_k)\, dv_k$ where $w_i = \frac{T_i}{e_i^{(1)}}$ for individuals with $T_i = 1$. Using the Laplace approximation formula which is exact when the function in the exponent is quadratic (Asparouhov, 2006), we have

$$\log l_k = -\frac{n_k^W - 1}{2}\log\left(2\pi\sigma_e^{2(1)}\right) - \frac{1}{2}\log\left(2\pi\left(\sigma_e^{2(1)} + n_k^W\sigma_v^{2(1)}\right)\right) - \frac{1}{2\sigma_e^{2(1)}}\sum_{i:K_i=k} w_i\left(Y_i - \bar{Y}_k^W\right)^2 - \frac{n_k^W\left(\bar{Y}_k^W - b_0^{(1)}\right)^2}{2\left(\sigma_e^{2(1)} + n_k^W\sigma_v^{2(1)}\right)}$$

where $n_k^W = \sum_{i:K_i=k} w_i$ and $\bar{Y}_k^W = \frac{\sum_{i:K_i=k} w_i Y_i}{\sum_{i:K_i=k} w_i} = \frac{\sum_{i:K_i=k} w_i Y_i}{n_k^W}$. Thus, the weighted estimating equation for $\boldsymbol{\theta}^{(1)}$ can be written as $Q^{W1}(\boldsymbol{\theta}^{(1)}) = \sum_{k\in\{1,\ldots,J\}} Q_k^{W1}(\boldsymbol{\theta}^{(1)})$ where

$$\begin{aligned} Q_k^{W1}(\boldsymbol{\theta}^{(1)}) &= \frac{\partial}{\partial\boldsymbol{\theta}^{(1)}}\log l_k = \begin{pmatrix} \frac{\partial}{\partial b_0^{(1)}}\log l_k \\ \frac{\partial}{\partial\sigma_v^{2(1)}}\log l_k \\ \frac{\partial}{\partial\sigma_e^{2(1)}}\log l_k \end{pmatrix} \\ &= \begin{pmatrix} \frac{n_k^W\left(\bar{Y}_k^W - b_0^{(1)}\right)}{\sigma_e^{2(1)} + n_k^W\sigma_v^{2(1)}} \\ \frac{-n_k^W}{2\left(\sigma_e^{2(1)} + n_k^W\sigma_v^{2(1)}\right)} + \frac{(Q_{k,1}^{W1})^2}{2} \\ \frac{-(n_k^W - 1)}{2\sigma_e^{2(1)}} + \frac{\sum_{i:K_i=k} w_i\left(Y_i - \bar{Y}_k^W\right)^2}{2(\sigma_e^{2(1)})^2} + \frac{Q_{k,2}^W}{n_k^W} \end{pmatrix} \end{aligned}$$

with $Q_{k,1}^{W1} = \frac{\partial}{\partial b_0^{(1)}}\log l_k$ and $Q_{k,2}^W = \frac{\partial}{\partial\sigma_v^{2(1)}}\log l_k$ denoting the first and second terms of $Q_k^{W1}(\boldsymbol{\theta}^{(1)})$.

Accordingly, we write the estimating equation for $\mathbf{a}^{(1)}$ as

$S^1(\mathbf{a}^{(1)}) = \sum_{k\in\{1,\ldots,J\}} S_k^1(\mathbf{a}^{(1)})$ where

$$\begin{aligned} S_k^1(\mathbf{a}^{(1)}) &= \sum_{i:K_i=k} S_i^1(\mathbf{a}^{(1)}) = \sum_{i:K_i=k}\frac{\partial}{\partial\mathbf{a}^{(1)}}\log p(T_i|L_i;\mathbf{a}^{(1)}) \\ &= \sum_{i:K_i=k}(T_i - e_i^{(1)})L_i^\top = 0 \end{aligned}$$

for $k = 1, ..., J$.

For estimation of $\boldsymbol{\theta}^{(0)} = (b_0^{(0)}, \sigma_e^{2(0)})^\top$, the estimating equation only involves individuals in the control group:

$$\begin{aligned} 0 &= Q^{W0}(\boldsymbol{\theta}^{(0)}) = \sum_{i:T_i=0}\frac{\partial}{\partial\boldsymbol{\theta}^{(0)}}\left\{\frac{1-T_i}{1-e_i^{(0)}}\log p(Y_i|T_i=0)\right\}. \\ &= \sum_{i:T_i=0}\begin{pmatrix} \frac{w_i(Y_i - b_0^{(0)})}{\sigma_e^{2(0)}} \\ \frac{1}{2\sigma_e^{2(0)}}\left(-w_i + \frac{w_i(Y_i - b_0^{(0)})^2}{\sigma_e^{2(0)}}\right) \end{pmatrix}. \end{aligned}$$

We obtain the parameter estimates by simultaneously solving the system of estimating equations $G(\boldsymbol{\gamma}) = (S^1(\mathbf{a}^{(1)}), S^0(\mathbf{a}^{(0)}), Q^{W1}(\boldsymbol{\theta}^{(1)}), Q^{W0}(\boldsymbol{\theta}^{(0)})) = 0$, where $\boldsymbol{\gamma} = (\mathbf{a}^{(1)}, \mathbf{a}^{(0)}, \boldsymbol{\theta}^{(1)}, \boldsymbol{\theta}^{(0)})$ contain all parameters being estimated. The treatment effect estimate is then calculated as $\hat{b}_0^{(1)} - \hat{b}_0^{(0)}$, with the sandwich-type standard error estimate given by $\mathrm{d}^\top \hat{\mathbf{A}}^{-1}\hat{\mathbf{B}}(\hat{\mathbf{A}}^\top)^{-1}\mathrm{d}$, where $\mathbf{A} = \mathbb{E}\left[\frac{\partial}{\partial \boldsymbol{\gamma}^\top} G(\boldsymbol{\gamma})\right]$, $\mathbf{B} = \mathbb{E}\left[G(\boldsymbol{\gamma})G(\boldsymbol{\gamma})^\top\right]$ and $\mathrm{d} = (\mathbf{0}, \mathbf{0}, \mathbf{0}, 1, 0, 0, -1, 0)^\top$. Matrix $\mathbf{A}$ has the form $\mathbf{A} = \begin{pmatrix} A_{11} & & & \\ & A_{22} & & \\ A_{31} & & A_{33} & \\ & A_{42} & & A_{44} \end{pmatrix}$. Let $\hat{\mathbf{A}} = \frac{\partial}{\partial \boldsymbol{\beta}^\top} U(\boldsymbol{\beta}) \mid_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}}$ where

$$\hat{A}_{11} = -\sum_{i=1}^{n} \hat{e}_i^{(1)}(1-\hat{e}_i^{(1)}) L_i L_i$$

$$\hat{A}_{22} = -\sum_{i=1}^{n} \hat{e}_i^{(0)}(1-\hat{e}_i^{(0)}) C_i C_i$$

$$\begin{aligned}
\hat{A}_{31} &= \sum_{k\in\{1,\ldots,J\}} \frac{\partial}{\partial \mathbf{a}^{(1)\top}} \begin{pmatrix} \frac{\partial}{\partial b_0^{(1)}} \log l_k \\ \frac{\partial}{\partial \sigma_v^{2(1)}} \log l_k \\ \frac{\partial}{\partial \sigma_e^{2(1)}} \log l_k \end{pmatrix} \\
&= \sum_{k\in\{1,\ldots,J\}} \begin{pmatrix} -\sum_{i:K_i=k} \frac{T_i(1-\hat{e}_i^{(1)})}{\hat{e}_i^{(1)}} \left[\frac{Y_i - \hat{b}_0^{(1)}}{\hat{\sigma}_e^{(1)2} + n_k^W \hat{\sigma}_u^{(1)2}} - \frac{\sum_{i:K_i=k} w_i (Y_i - \hat{b}_0^{(1)}) \hat{\sigma}_u^{(1)2}}{\left(\hat{\sigma}_e^{(1)2} + n_k^W \hat{\sigma}_u^{(1)2}\right)^2} \right] L_i \\ -\sum_{i:K_i=k} \frac{T_i(1-\hat{e}_i^{(1)})}{\hat{e}_i^{(1)}} \frac{\hat{\sigma}_e^{(1)2}}{2\left(\hat{\sigma}_e^{(1)2} + n_k^W \hat{\sigma}_u^{(1)2}\right)^2} L_i + \hat{Q}_{k,1}^{W1} \hat{A}_{31,1} \\ \sum_{i:K_i=k} \frac{T_i(1-\hat{e}_i^{(1)})}{\hat{e}_i^{(1)}} \left(\frac{(-\hat{\sigma}_e^{(1)2} + \left(Y_i - \bar{Y}_k^W\right)^2)}{2(\hat{\sigma}_e^{(1)2})^2} - \frac{\hat{Q}_{k,2}^{W1}}{(n_k^W)^2} \right) L_i + \frac{\hat{A}_{31,2}}{n_k^W} \end{pmatrix}
\end{aligned}$$

where $\hat{Q}_{k,2}^{W1}$ is the second row of $Q_k^{W1}(\hat{\boldsymbol{\theta}}^{(1)})$ and $\hat{A}_{31,2}$ is the second row of $\hat{A}_{31}$; and

$$\hat{A}_{33} = \sum_{k\in\{1,\ldots,J\}} \begin{pmatrix} \frac{-n_k^W}{\hat{\sigma}_e^{(1)2} + n_k^W \hat{\sigma}_u^{(1)2}} & \hat{Q}_{k,1}^{W1} \hat{A}_{33(1,1)} & \hat{A}_{33(1,2)}/n_k^W \\ \hat{Q}_{k,1}^{W1} \hat{A}_{33(1,1)} & \frac{(\hat{A}_{33(1,1)})^2}{2} + (\hat{Q}_{k,1}^{W1})^2 \hat{A}_{33(1,1)} & \hat{A}_{33(2,2)}/n_k^W \\ \hat{A}_{33(1,2)}/n_k^W & \hat{A}_{33(2,2)}/n_k^W & \frac{n_k^W - 1}{2(\hat{\sigma}_e^{(1)2})^2} - \frac{\sum_{i:K_i=k} w_i \left(Y_i - \bar{Y}_k^W\right)^2}{(\hat{\sigma}_e^{(1)2})^3} + \frac{\hat{A}_{33(2,2)}}{(n_k^W)^2} \end{pmatrix}$$

$$\hat{A}_{42} = \begin{pmatrix} -\sum_{i=1}^{n} \frac{1-T_i}{1-\hat{e}_i^{(0)}} \hat{e}_i^{(0)} (\frac{Y_i - \hat{b}_0^{(0)}}{\hat{\sigma}_e^{(0)2}}) C_i \\ -\sum_{i=1}^{n} \frac{1-T_i}{1-\hat{e}_i^{(0)}} \hat{e}_i^{(0)} \left(\frac{-1}{2\hat{\sigma}_e^{(0)2}} + \frac{(Y_i - \hat{b}_0^{(0)})^2}{2(\hat{\sigma}_e^{(0)2})^2} \right) C_i \end{pmatrix}$$

$$\hat{A}_{44} = \begin{pmatrix} -\sum_{i=1}^{n} \frac{1-T_i}{(1-\hat{e}_i^{(0)})\hat{\sigma}_e^{(0)2}} & -\sum_{i=1}^{n} \frac{1-T_i}{(1-\hat{e}_i^{(0)})\hat{\sigma}_e^{(0)2}} (\frac{Y_i - \hat{b}_0^{(0)}}{\hat{\sigma}_e^{(0)2}}) \\ -\sum_{i=1}^{n} \frac{1-T_i}{(1-\hat{e}_i^{(0)})\hat{\sigma}_e^{(0)2}} (\frac{Y_i - \hat{b}_0^{(0)}}{\hat{\sigma}_e^{(0)2}}) & -\sum_{i=1}^{n} \frac{1-T_i}{(1-\hat{e}_i^{(0)})\hat{\sigma}_e^{(0)2}} \left(\frac{-1}{2\hat{\sigma}_e^{(0)2}} + (\frac{Y_i - \hat{b}_0^{(0)}}{\hat{\sigma}_e^{(0)2}})^2 \right) \end{pmatrix}.$$

Let $\hat{\mathbf{B}} = G_i(\hat{\boldsymbol{\gamma}})U_i(\hat{\boldsymbol{\gamma}})^\top$. The asymptotic covariance matrix can be estimated by the sandwich-type covariance matrix $ACov(\hat{\boldsymbol{\beta}}) = \hat{\mathbf{A}}^{-1}\hat{\mathbf{B}}(\hat{\mathbf{A}}^\top)^{-1}$.

Supplemental material for the simulation study

The true values of the coefficients in the potential outcome generation models in Eqs. (34) and (35) and their effect sizes are listed in Tables 1 and 2, respectively.

– Tables 1 and 2 –

Additional simulation results

Randomized treatment assignment. Results from the MLM-based methods when the treatment assignment is randomized ($a_c = 0$, $a_w = 0$) with smaller numbers of clusters and cluster sizes are displayed in Table 3.

– Insert Table 3 –

Non-randomized treatment assignment with the largest number of clusters and the largest cluster size. Absolute biases and MSE values with zero (0) average causal effect when the number of clusters and cluster size are large (i.e., $J = 40$ and $m = 40$) are displayed in Tables 4 and 5, respectively.

– Insert Tables 4 and 5 –

Non-randomized treatment assignment with smaller numbers of clusters or smaller cluster sizes. Absolute biases, MSE values, and Type I error rates with zero (0) average causal effect under smaller numbers of clusters or cluster sizes are displayed in Tables 6 and 7.

– Insert Tables 6 and 7 –

Pilot simulation study: Quadratic outcome model and quadratic PS model

We conducted a pilot simulation study to examine how misspecifying the functional forms of confounders in the outcome or the PS models influences the performance of the methods.

Design. The treatment assignments were generated by $T_i \sim \text{Bernoulli}(\pi_i)$, where

$$\text{logit}(\pi_i) = \frac{a_c}{4}C_i + \frac{a_w}{4}W_i + \frac{a_c}{4}C_i^2 + \frac{a_c + a_w}{4}W_iC_i$$

with quadratic terms of the covariates being included. The potential outcome generation models also involved quadratic terms of the covariates. Specifically, the potential outcome under treatment for individual i with the potential cluster assignment $K_i(1) = k$ was $Y_i(1, K_i(1)) = Y_i(1, k)$ generated by

$$Y_i(1, k) = b_0^{(1)} + .25\tilde{W}_k + .15C_i + .75C_i^2 + .50\tilde{W}_kC_i + \tilde{v}_k^{(1)} + \tilde{\epsilon}_{ik}^{(1)}$$

where $\tilde{W}_k$ is the potential cluster mean of W. The potential outcome under control was generated by

$$Y_i(0) = b_0^{(0)} + .15C_i + .75C_i^2 + \epsilon_i^{(0)}.$$

The observed outcomes were generated via $Y_i = Y_i(1, K_i(1))T_i + Y_i(0)(1 - T_i)$.

For the manipulated factors, four types of treatment assignment mechanisms were examined, by varying a_c and a_w each at two levels as $a_c, a_w = 0, 0.50$. The ICC level of the potential outcome under treatment was varied at two levels as $ICC_{Y1} = 0.15, 0.33$. Under each ICC level, we examined two levels of the average causal treatment effect, a zero (0) level and a non-zero level, which are $b_0^{(1)} - b_0^{(0)} = 0, 0.46$ under $ICC_{Y1} = 0.15$; and $b_0^{(1)} - b_0^{(0)} = 0, 0.41$ under $ICC_{Y1} = 0.33$. The non-zero levels had Cohen's d between small (0.2) to medium (0.5). Without loss of generality, $b_0^{(0)}$ was set at 0. The number of clusters was varied as $J = 20, 40$, and the cluster size was varied as $m = 10, 20, 40$.

The data generation details were similar to the simulation study in the main text. Specifically, the random cluster-specific intercept $\tilde{v}_k^{(1)}$ and the within-cluster residual $\tilde{\epsilon}_{ik}^{(1)}$

were normally distributed with means zero and variances $\sigma_v^{2(1)}$ and $\sigma_e^{2(1)}$ respectively. The residual term $\epsilon_i^{(0)}$ was normally distributed with mean zero and variance $\sigma_e^{2(0)}$, which was set to be the same as the within-cluster residual variance of the potential outcome under treatment (i.e., $\sigma_e^{2(0)} = \sigma_e^{2(1)}$). C_i had the same ICC level as the potential outcome under treatment (i.e., $ICC_C = ICC_{Y1}$), and was simulated as the summation of its between- and within-cluster components, which were normally distributed with means 0 and variances ICC_C and $1 - ICC_C$ respectively. The ICC of W_i was set at $ICC_W = 0.30$, and W_i was also simulated as the summation of its between- and within-cluster components, which were normally distributed with means 0 and variances ICC_W and $1 - ICC_W$ respectively.

Six methods (4 outcome modeling and 2 IPW methods) were implemented in MLM and regression (REG) settings respectively. (1) qCWadj under REG ($REG.qCWadj$) or MLM ($MLM.qCWadj$): qCWadj under REG is a linear regression of Y on T, C, TW, TWC and C^2, namely

$$Y_i = b_0 + b_1 T_i + b_c C_i + b_w T_i W_i + b_{wc} T_i W_i C_i + b_{qc} C_i^2 + \epsilon_i;$$

qCWadj under MLM adjusts for C, W, WC, C^2 in the treatment group multilevel model and for C, C^2 in the control group regression, namely

$$(Y_i | T_i = 0) = b_0^{(0)} + b_c^{(0)} C_i + b_{qc}^{(0)} C_i^2 + \epsilon_i^{(0)}$$

$$(Y_i | T_i = 1, K_i = k) = b_0^{(1)} + b_c^{(1)} C_i + b_w^{(1)} W_i + b_{wc}^{(1)} W_i C_i + b_{qc}^{(1)} C_i^2 + v_k^{(1)} + \epsilon_{ik}^{(1)}.$$

The qCWadj models are specified following the identification results (IA1 and IA2) and have correct functional form specifications of the covariates.

(2) Cadj under REG ($REG.Cadj$) or CaWadj under MLM ($MLM.C\bar{W}adj$): Cadj under REG is the regression adjusting for C only (i.e., Eq. (12) in the main text), namely

$$Y_i = b_0 + b_1 T_i + b_c C_i + \epsilon_i;$$

CaWadj under MLM is the model in Eqs. (21)-(22) of the main text, which adjusts for the observed cluster mean $\bar{W}_k$, namely

$$(Y_i|T_i = 0) = b_0^{(0)} + b_c^{(0)} C_i + \epsilon_i^{(0)}$$

$$(Y_i|T_i = 1, K_i = k) = b_0^{(1)} + b_c^{(1)} C_i + b_w^{(1)} \bar{W}_k + v_k^{(1)} + \epsilon_{ik}^{(1)}.$$

(3) qCadj under REG ($REG.qCadj$) or qCaWadj under MLM ($MLM.qC\bar{W}adj$): qCadj under REG is a regression of Y on T, C and C^2, namely

$$Y_i = b_0 + b_1 T_i + b_c C_i + b_{qc} C_i^2 + \epsilon_i;$$

qCaWadj under MLM adjusts for C, $\bar{W}_k$, $\bar{W}_k C$ and C^2 in the treatment group multilevel model, and for C, C^2 in the control group regression, namely

$$(Y_i|T_i = 0) = b_0^{(0)} + b_c^{(0)} C_i + b_{qc}^{(0)} C_i^2 + \epsilon_i^{(0)}$$

$$(Y_i|T_i = 1, K_i = k) = b_0^{(1)} + b_c^{(1)} C_i + b_w^{(1)} \bar{W}_k + b_{wc}^{(1)} \bar{W}_k C_i + b_{qc}^{(1)} C_i^2 + v_k^{(1)} + \epsilon_{ik}^{(1)}.$$

(4) unadj under REG ($REG.unadj$) or MLM ($MLM.unadj$): unadj under REG is the unadjusted regression model (i.e., Eq. (11) of the main text), namely

$$Y_i = b_0 + b_1 T_i + \epsilon_i;$$

unadj under MLM is the unadjusted MLM (in Eqs. (17)-(18) of the main text), namely

$$(Y_i|T_i = 0) = b_0^{(0)} + \epsilon_i^{(0)}$$

$$(Y_i|T_i = 1, K_i = k) = b_0^{(1)} + v_k^{(1)} + \epsilon_{ik}^{(1)}.$$

(5) IPW under REG ($REG.IPW$) or MLM ($MLM.IPW$): IPW adjustment based on the PS models specified following the identification results (IA1 and IA2) and have correct functional form specifications of the covariates, which includes C, W, C^2 and WC in the treatment group PS model:

$$\text{logit}(\Pr(T_i = 1|\mathbf{C}_i, \mathbf{W}_i)) = \alpha_c C_i + \alpha_w W_i + \alpha_{qc} C_i^2 + \alpha_{wc} W_i C_i;$$

and includes C and C^2 in the control group PS model:

$$\text{logit}(\Pr(T_i = 1|\mathbf{C}_i)) = \alpha_c C_i + \alpha_{qc} C_i^2.$$

IPW under REG combines the IPW adjustment with the unadjusted regression model for the outcome, ignoring the partial nesting structure; IPW under MLM combines the IPW adjustment with the unadjusted MLM for the outcome (i.e., Eqs. (17)-(18) of the main text).

(6) IPW.mis under REG or MLM: functional-form-misspecified IPW adjustment based on the PS models in Eqs. (23)-(24), which were specified following the identification results (IA1 and IA2) but have incorrect functional form specifications of the covariates (i.e., omitting the quadratic terms); the treatment group PS models is

$$\text{logit}(\Pr(T_i = 1|\mathbf{C}_i, \mathbf{W}_i)) = \alpha_c C_i + \alpha_w W_i;$$

the control group PS model is

$$\text{logit}(\Pr(T_i = 1|\mathbf{C}_i)) = \alpha_c C_i.$$

IPW.mis under REG combines the functional-form-misspecified IPW adjustment with the unadjusted regression model for the outcome, ignoring the treatment clustering structure; IPW.mis under MLM combines the functional-form-misspecified IPW adjustment with the unadjusted MLM for the outcome.

Results. Figures 1, 2, 3, and 4 display the relative biases, MSE values, coverage rates of the 95% confidence intervals (CI), and the Type I error rates at the .05 nominal significance level of the evaluated methods, respectively.

With appropriate implementations of IA1 and IA2 and correct functional form specifications, the MLM-based IPW ($MLM.IPW$) method in general yielded estimates of the average causal effect with ignorable relative biases $< 5\%$ and had satisfactory coverage rates and Type I error rates under the studied conditions. Although specified based on the identification results, the MLM-based IPW.mis method had nonignorable relative biases, unsatisfactory coverage rates and Type I error rates when C was a true confounder ($a_c = 0.5$), due to omitting the quadratic terms of the covariates. Similarly, the regression-based IPW ($REG.IPW$) method with appropriate implementations of the IAs and correct functional form specification of the PS models led to ignorable relative biases, whereas the regression-based IPW.mis method that is specified following the identification results but omits the quadratic terms of covariates produced nonignorable relative biases. This demonstrates the lack of robustness of the IPW adjustment to functional form misspecification of the PS models.

In the cases where C was a true confounder whereas W was not (i.e., $a_c = 0.5$, $a_w = 0$), the methods that adjust for the observed cluster mean of W (i.e., qCaWadj and CaWadj under MLM) and the methods that omit W (i.e., qCadj and Cadj under REG) had appropriate implementations of the IAs. In these cases (i.e., $a_c = 0.5$, $a_w = 0$), the MLM-based qCaWadj outcome modeling ($MLM.qC\bar{W}adj$) method that includes the quadratic terms of covariates had estimates with ignorable relative biases and satisfactory coverage rates and Type I error rates overall, whereas the MLM-based CaWadj ($MLM.C\bar{W}adj$) method that omits the quadratic terms yielded estimates with nonignorable relative biases and unsatisfactory coverage rates and Type I error rates. Analogously, the regression-based qCadj ($REG.qCadj$) that incorporates the quadratic terms of covariates produced estimates with ignorable relative biases, whereas the regression-based Cadj ($REG.Cadj$) that omits such quadratic terms produced non-ignorable relative biases. This shows that the outcome modeling methods are sensitive to the functional form specification of covariates in the outcome models.

Note that for the outcome model $MLM.qC\bar{W}adj$ (i.e., qCaWadj under MLM), although it had relative biases below 5% overall due to the inclusion of quadratic terms of covariates, the inappropriate implementation of the identification results (i.e., adjusting for the observed cluster mean $\bar{W}_k$, a post-treatment variable) had consequences on its performance. Specifically, when W was not a true confounder (i.e., $a_w = 0$), the relative biases from $MLM.qC\bar{W}adj$ (i.e., qCaWadj under MLM) were comparable to those from outcome model $MLM.qCWadj$ (i.e., qCWadj under MLM) that not only incorporates the quadratic terms of covariates but also has proper implementation of the IAs. However, when W was indeed a true confounder (i.e., $a_w = 0.5$), the outcome model $MLM.qC\bar{W}adj$ that adjusts for the post-treatment $\bar{W}_k$ in general produced distinguishably larger biases than the model $MLM.qCWadj$ specified following the IAs. Similarly, for the regression model $REG.qCadj$ (i.e., qCadj under REG) that includes the quadratic term C^2 but omits W, when W was a true confounder (i.e., $a_w = 0.5$), it in general had larger relative biases than the regression model $REG.qCWadj$ that is specified following the identification results and has correct functional form specifications of the covariates.

– Insert Figure 1 –

– Insert Figure 2 –

– Insert Figure 3 –

– Insert Figure 4 –

A simulated-data example with the analysis code for implementing the proposed methods

We use a simulated-data example to illustrate the analysis code for implementing the proposed methods, including the MLM-based IPW and MLM-based outcome modeling methods. The dataset was simulated in the same way as described in the simulation study. In particular, the treatment assignment was non-randomized and affected by both C and W (same as the simulation condition with $a_c = 0.7$ and $a_w = 0.7$); the true value of the average causal treatment effect was 0.62 (same as the simulation condition with ICC being 0.1 and $b^{(1)}_{c.between} > b^{(1)}_{c.within}$). The simulated dataset contains five variables: one general confounder (“C”), one clustering confounder (“W”), treatment assignment (“Trt”, 1 for treatment and 0 for control), cluster assignment (“K”, which is 0 for individuals in the control group), and outcome (“Y”). Below we illustrate the R code for implementing the proposed methods.

First, install and load the R packages needed for conducting the analysis.

```
## Install packages

##(no need if the packages have already been installed before)

install.packages(c("devtools","MplusAutomation","lme4","tidyverse"))
```

Install and load our developed package: IPWpn.

```
devtools::install_github("xliu12/IPWpn")

library(IPWpn);
```

The simulated dataset is included in the IPWpn package.

```
## load the data, and have a look

data("dat_obs"); head(dat_obs)
> library(IPWpn);
> data("dat_obs"); tail(dat_obs,n=3)
    Trt  K          Y          C          W
798   0  0 0.57820484  0.2489332 -0.5643205
799   0  0 0.06130894 -1.0015504  0.3081488
800   1 20 0.34168196  1.3378635  0.8578789
```

Run the R function “IPW.MLMPN()” to apply the proposed MLM-based IPW methods to the data for estimating the average treatment effect of “Trt” on “Y”.

The following code runs the MLM-based IPW[C, W] method (i.e., using the PS model in Eq. 23 to adjust for both C and W in the MLM for treatment group outcomes in Eq. 18 and the PS model in Eq. 24 to adjust for C in the regression for control group outcomes in Eq. 17).

```
## MLM-based IPW[C, W] ##

res_mlmipw_cw = IPW.MLMPN( outcome = 'Y', Treatment_assignment = 'Trt',

  cluster_assignment = 'K', general_confounders = c('C'),

  clustering_confounders = c('W') ,

  data = dat_obs )

res_mlmipw_cw
> res_mlmipw_cw
$Estimate
   gamm1
0.592668

$SE
[1] 0.1755448

$z.wald
[1] 3.376165
```

In the above code, arguments "outcome", "Treatment_assignment", "cluster_assignment" read in strings indicating the names of the outcome variable, treatment assignment variable, and cluster assignment variable in the dataset, respectively. Arguments "general_confounders" and "clustering_confounders" read in character vectors containing the names of the general confounders and clustering confounders in the dataset, respectively.[1] The function returns a list containing the estimate of the average causal treatment effect, the sandwich-type standard error estimate, and the corresponding z statistic (as shown in the screenshot above).

The following code runs the MLM-based IPW[C] method (i.e., using the PS model in

[1] As elaborated in the main text, the general confounders are those baseline confounders that can directly affect both the treatment assignment and outcome within a cluster of the treatment group. The clustering confounders are those baseline confounders that can directly affect both the treatment assignment and cluster assignment in the treatment group, but do not directly influence the outcome once the treatment assignment and cluster assignment are given. When one is uncertain about whether a covariate is a general confounder or a clustering confounder, the covariate should be classified as a general confounder to be conservative, because a general confounder can directly affect the outcome, cluster assignment, and treatment assignment.

Eq. 24 to adjust for C in the MLM for treatment group outcomes in Eq. 18 and in the regression for control group outcomes in Eq. 17):

```
## MLM-based IPW[C] ##

res_mlmipw_c = IPW.MLMPN( outcome = 'Y', Treatment_assignment = 'Trt',

  cluster_assignment = 'K', general_confounders = c('C'),

  clustering_confounders = NULL ,

  data = dat_obs )

res_mlmipw_c
> res_mlmipw_c
$Estimate
   gamm1
0.593668

$SE
[1] 0.1745577

$z.wald
[1] 3.400984
```

In the above code, we input "`clustering_confounders = NULL`", because with the MLM-based IPW[C] method, the clustering confounders are not adjusted

Run the R function "Outcome.MLMPN()" to apply the proposed MLM-based outcome models to the data.

The following code runs the MLM-based Outcome[C, W] method (i.e., the outcome model in Eqs. 13-14, which adjusts for both C and W in the MLM for treatment group outcomes in Eq. 14 and adjusts for C in the regression for control group outcomes in Eq. 13):

```
## MLM-based Outcome[C, W] ##

res_mlmoutcome_cw = Outcome.MLMPN( outcome = 'Y', Treatment_assignment
= 'Trt', cluster_assignment = 'K', general_confounders = c('C'),
clustering_confounders = c('W') , data = dat_obs )

res_mlmoutcome_cw
> res_mlmoutcome_cw
$Estimate
(Intercept)
  0.5989137

$SE
[1] 0.1512346

$z.wald
(Intercept)
   3.960163
```

The following code runs the MLM-based Outcome[C] method (i.e., the outcome model in Eqs. 15-16, which adjusts for C in the MLM for treatment group outcomes in Eq. 16 and in the regression for control group outcomes Eq. 15):

```
## MLM-based Outcome[C] ##

res_mlmoutcome_c = Outcome.MLMPN( outcome = 'Y', Treatment_assignment =
'Trt',
  cluster_assignment = 'K', general_confounders = c('C'),
  clustering_confounders = NULL ,
  data = dat_obs )

res_mlmoutcome_c
> res_mlmoutcome_c
$Estimate
(Intercept)
  0.5967914

$SE
[1] 0.149958

$z.wald
(Intercept)
   3.979723
```

References

Asparouhov, T. (2006). General multi-level modeling with sampling weights. *Communications in Statistics - Theory and Methods*, *35*(3), 439–460. doi: 10.1080/03610920500476598

VanderWeele, T. J., & Hernan, M. A. (2013). Causal inference under multiple versions of treatment. *Journal of Causal Inference*, *1*(1), 1–20. doi: 10.1515/jci-2012-0002

Table 1

True values of the coefficients in the potential outcome generation models (Eqs. 34-35). The average causal effect is $b_0^{(1)} - b_0^{(0)}$, with $b_0^{(0)} = 0$. The listed $b_0^{(1)}$ values represent the true values of the non-zero levels of average causal effect; for the zero level, $b_0^{(1)}$ equals 0.

$b^{(1)}_{c.between}$ vs. $b^{(1)}_{c.within}$	Equal			Larger					Smaller				
ICC_{Y1}	$b_0^{(1)}$	$b_w^{(1)}$	$b^{(1)}_{c.between} = b^{(1)}_{c.within} = b_c^{(0)}$	$b_0^{(1)}$	$b_w^{(1)}$	$b^{(1)}_{c.between}$	$b^{(1)}_{c.within}$	$b_c^{(0)}$	$b_0^{(1)}$	$b_w^{(1)}$	$b^{(1)}_{c.between}$	$b^{(1)}_{c.within}$	$b_c^{(0)}$
0.05	0.56	0.99	0.38	0.62	0.99	2.42	0.55	0.54	0.59	0.99	0.00	0.55	0.54
0.10	0.56	0.97	0.38	0.62	0.97	1.69	0.56	0.53	0.59	0.97	0.00	0.56	0.53
0.30	0.53	0.92	0.36	0.58	0.92	0.92	0.60	0.50	0.56	0.92	0.00	0.60	0.50

Note. The coefficients of covariates (when non-zero) were set to have medium effect sizes (Cohen's $f^2 = 0.15$) under the randomized condition (i.e. $a_c = 0, a_w = 0$). $b_0^{(1)}$ was set such that the average causal effect had around medium effect size (Cohen's $d = 0.5$) under the randomized condition.

Table 2

Effect sizes of the coefficients in the potential outcome generation models (Eqs. 34-35). The effect size measure is Cohen's d for the average causal treatment effect $b_0^{(1)} - b_0^{(0)}$ ("Trt.Eff"; with $b_0^{(0)} = 0$), and Cohen's f^2 for the coefficients of covariates (i.e., $b_c^{(0)}$, $b_w^{(1)}$, $b_{c.within}^{(1)}$, and $b_{c.between}^{(1)}$).

	$b_{c.between}^{(1)}$ vs. $b_{c.within}^{(1)}$	Equal			Larger					Smaller				
$(a_c,\ a_w)$	ICC_{Y1}	Trt.Eff	$b_w^{(1)}$	$b_{c.between}^{(1)} = b_{c.within}^{(1)} = b_c^{(0)}$	Trt.Eff	$b_w^{(1)}$	$b_{c.between}^{(1)}$	$b_{c.within}^{(1)}$	$b_c^{(0)}$	Trt.Eff	$b_w^{(1)}$	$b_{c.between}^{(1)}$	$b_{c.within}^{(1)}$	$b_c^{(0)}$
(0, 0)	0.05	0.5	0.15	0.15	0.5	0.15	0.15	0.15	0.15	0.5	0.15	0.00	0.15	0.15
	0.10	0.5	0.15	0.15	0.5	0.15	0.15	0.15	0.15	0.5	0.15	0.00	0.15	0.15
	0.30	0.5	0.15	0.15	0.5	0.15	0.15	0.15	0.15	0.5	0.15	0.00	0.15	0.15
(0, 0.5)	0.05	0.6	0.17	0.15	0.6	0.18	0.15	0.15	0.15	0.6	0.18	0.00	0.15	0.15
	0.10	0.6	0.18	0.15	0.6	0.18	0.15	0.15	0.15	0.6	0.18	0.00	0.15	0.15
	0.30	0.6	0.18	0.15	0.6	0.17	0.15	0.15	0.15	0.5	0.18	0.00	0.15	0.15
(0.5, 0)	0.05	0.7	0.15	0.22	0.7	0.14	0.15	0.21	0.22	0.7	0.15	0.00	0.21	0.21
	0.10	0.7	0.15	0.22	0.7	0.14	0.16	0.20	0.21	0.7	0.15	0.00	0.21	0.21
	0.30	0.7	0.15	0.22	0.8	0.16	0.19	0.20	0.22	0.7	0.15	0.00	0.20	0.21
(0.5, 0.5)	0.05	0.7	0.17	0.21	0.8	0.17	0.16	0.20	0.22	0.8	0.18	0.00	0.20	0.21
	0.10	0.7	0.17	0.21	0.8	0.17	0.16	0.19	0.22	0.8	0.18	0.00	0.20	0.21
	0.30	0.7	0.17	0.21	0.8	0.16	0.17	0.19	0.22	0.8	0.18	0.00	0.19	0.21

Note. The coefficients of covariates (when non-zero) were set to have medium effect sizes ($f^2 = 0.15$) under the randomized condition (i.e. $a_c = 0, a_w = 0$). $b_0^{(1)}$ was set such that the average causal effect had around medium effect size ($d = 0.5$) under the randomized condition.

Table 3

Randomized treatment assignment ($a_c = 0$, $a_w = 0$): Results from the MLM-based methods for varied cluster sizes (m) and numbers of clusters (J).

	$J \times m$						$J \times m$					
	10×10	20×5	10×20	40×5	10×40	40×10	10×10	20×5	10×20	40×5	10×40	40×10
MLM	Relative bias (treatment effect: medium, $d = 0.5$)						MSE (treatment effect: medium, $d = 0.5$)					
Unadjusted	0.012	0.021	-0.013	0.009	0.006	-0.004	0.099	0.072	0.068	0.032	0.057	0.022
Outcome[$\bar{C}_k$, $C - \bar{C}_k$]	-0.001	0.009	-0.011	0.015	0.014	-0.003	0.094	0.066	0.071	0.028	0.057	0.020
Outcome[C]	0.010	0.007	-0.016	0.013	0.010	-0.002	0.091	0.064	0.066	0.028	0.055	0.019
Outcome[C, $\bar{W}_k$]	-0.000	0.007	-0.023	0.013	0.011	0.001	0.072	0.059	0.047	0.025	0.035	0.016
Outcome[C, W]	0.011	0.005	-0.015	0.013	0.008	-0.002	0.083	0.061	0.063	0.026	0.053	0.018
IPW[C]	0.009	0.008	-0.013	0.011	0.007	-0.005	0.089	0.064	0.066	0.028	0.055	0.020
IPW[C, W]	0.010	0.008	-0.013	0.012	0.006	-0.005	0.089	0.065	0.066	0.028	0.055	0.020
MLM	Coverage rates (treatment effect: medium, $d = 0.5$)						Type I error rates (treatment effect: 0)					
Unadjusted	**0.923**	0.940	0.936	0.956	0.927	0.949	**0.077**	0.060	0.064	0.044	0.073	0.051
Outcome[$\bar{C}_k$, $C - \bar{C}_k$]	**0.902**	0.940	**0.901**	0.950	**0.906**	0.952	**0.098**	0.060	**0.099**	0.050	**0.094**	0.048
Outcome[C]	**0.916**	0.944	**0.920**	0.949	**0.920**	0.955	**0.084**	0.056	**0.080**	0.051	**0.080**	0.045
Outcome[C, $\bar{W}_k$]	**0.918**	0.939	**0.916**	0.946	**0.909**	0.939	**0.082**	0.061	**0.084**	0.054	**0.091**	0.061
Outcome[C, W]	**0.921**	0.942	**0.920**	0.949	**0.921**	0.954	**0.079**	0.058	**0.080**	0.051	**0.079**	0.046
IPW[C] (sandwich SE)	0.934	0.955	0.939	0.963	0.930	0.968	0.065	0.044	0.060	0.036	0.070	0.032
(regular SE)	**0.914**	0.938	**0.925**	0.946	**0.920**	0.955	**0.087**	0.062	**0.076**	0.053	**0.080**	0.045
IPW[C, W] (sandwich SE)	0.938	0.956	0.940	0.966	0.931	0.968	0.062	0.044	0.060	0.034	0.069	0.032
(regular SE)	**0.915**	0.938	**0.922**	0.948	**0.921**	0.955	**0.085**	0.062	**0.078**	0.052	**0.079**	0.045

Note. The values were averaged across all ICC levels and combinations of the between- and within-cluster effects of C. Boldface font is used for relative biases above 0.05 or below -0.05, for coverage rates below 0.925 or above 0.975, and for Type I error rates below 0.025 or above 0.075.

Table 4

Absolute biases for varied combinations of the effects of C and W on treatment assignment (quantified by a_c and a_w, respectively) and the between- and within-cluster effects of C on the outcome under treatment (quantified by $b^{(1)}_{c.between}$ and $b^{(1)}_{c.within}$, respectively). The average causal treatment effect is zero. The number of cluster is $J = 40$ and the cluster size is $m = 40$.

	$a_c = 0.7$, $a_w = 0$			$a_c = 0$, $a_w = 0.7$			$a_c = 0.7$, $a_w = 0.7$		
$b^{(1)}_{c.between}$ vs. $b^{(1)}_{c.within}$	Equal	Larger	Smaller	Equal	Larger	Smaller	Equal	Larger	Smaller
MLM									
Unadjusted	0.222	0.328	0.323	0.011	0.008	0.012	0.216	0.313	0.312
Outcome[$\bar{C}_k$, $C - \bar{C}_k$]	0.004	-0.197	0.117	0.011	0.010	0.011	0.061	-0.124	0.164
Outcome[C]	0.000	0.001	-0.000	0.011	0.009	0.010	0.010	0.009	0.009
Outcome[C, $\bar{W}_k$]	-0.000	-0.000	-0.001	-0.206	-0.205	-0.207	-0.189	-0.183	-0.195
Outcome[C, W]	0.000	0.001	-0.000	0.001	0.001	0.000	0.000	0.001	-0.000
IPW[C]	0.007	0.012	0.011	0.012	0.009	0.013	0.021	0.024	0.027
IPW[C, W]	0.007	0.012	0.011	0.001	0.001	0.001	0.009	0.014	0.013
REG									
Unadjusted	0.235	0.379	0.320	0.088	0.087	0.088	0.298	0.429	0.376
Outcome[C]	0.001	0.001	0.001	0.089	0.088	0.089	0.089	0.089	0.090
Outcome[C, W]	0.001	0.001	0.001	-0.001	-0.001	-0.001	-0.001	0.001	-0.002
IPW[C]	0.001	0.002	0.001	0.089	0.088	0.089	0.090	0.091	0.091
IPW[C, W]	0.001	0.002	0.001	-0.000	0.000	-0.000	0.001	0.002	0.001

Note. The values were averaged across all ICC levels.

Table 5

Mean squared errors (MSE) values for varied combinations of the effects of C and W on treatment assignment (quantified by a_c and a_w, respectively) and the between- and within-cluster effects of C on the outcome under treatment (quantified by $b^{(1)}_{c.between}$ and $b^{(1)}_{c.within}$, respectively). The average causal treatment effect is 0. The number of cluster is $J = 40$ and the cluster size is $m = 40$.

	$a_c = 0.7$, $a_w = 0$			$a_c = 0$, $a_w = 0.7$			$a_c = 0.7$, $a_w = 0.7$		
$b^{(1)}_{c.between}$ vs. $b^{(1)}_{c.within}$	Equal	Larger	Smaller	Equal	Larger	Smaller	Equal	Larger	Smaller
MLM									
Unadjusted	0.063	0.125	0.119	0.013	0.017	0.014	0.060	0.115	0.111
Outcome[$\bar{C}_k$, $C - \bar{C}_k$]	0.021	0.071	0.036	0.013	0.014	0.013	0.024	0.041	0.047
Outcome[C]	0.013	0.016	0.014	0.013	0.016	0.013	0.013	0.016	0.014
Outcome[C, $\bar{W}_k$]	0.007	0.010	0.008	0.050	0.053	0.052	0.044	0.045	0.047
Outcome[C, W]	0.012	0.015	0.013	0.012	0.015	0.013	0.012	0.015	0.013
IPW[C]	0.013	0.016	0.014	0.012	0.016	0.013	0.013	0.016	0.014
IPW[C, W]	0.013	0.016	0.014	0.012	0.016	0.013	0.013	0.017	0.014
REG									
Unadjusted	0.069	0.162	0.117	0.021	0.025	0.022	0.102	0.202	0.156
Outcome[C]	0.013	0.017	0.014	0.021	0.024	0.021	0.021	0.024	0.022
Outcome[C, W]	0.010	0.013	0.010	0.010	0.013	0.010	0.010	0.013	0.010
IPW[C]	0.013	0.017	0.014	0.021	0.024	0.021	0.022	0.025	0.023
IPW[C, W]	0.013	0.017	0.014	0.012	0.016	0.013	0.014	0.017	0.015

Note. The values were averaged across all ICC levels.

Table 6

Absolute biases and MSE values for varied cluster sizes (m) and numbers of clusters (J) from the MLM-based methods adjusting for C only (i.e., Outcome[C] and IPW[C] under MLM) and the MLM-based methods adjusting for C and W both (i.e., Outcome[C, W] and IPW[C, W] under MLM). The average causal treatment effect is zero.

	Absolute bias						MSE					
	$J \times m = 100$		$J \times m = 200$		$J \times m = 400$		$J \times m = 100$		$J \times m = 200$		$J \times m = 400$	
	10×10	20×5	10×20	40×5	10×40	40×10	10×10	20×5	10×20	40×5	10×40	40×10
MLM						$a_c = 0.7$, $a_w = 0$						
Outcome[C]	0.002	0.007	-0.012	0.001	0.003	-0.004	0.097	0.070	0.068	0.031	0.056	0.021
Outcome[C, W]	0.001	0.004	-0.013	0.001	0.002	-0.004	0.089	0.068	0.064	0.029	0.054	0.019
IPW[C]	0.031	0.053	0.006	0.047	0.011	0.022	0.099	0.076	0.069	0.033	0.056	0.022
IPW[C, W]	0.031	0.053	0.006	0.048	0.011	0.022	0.099	0.076	0.069	0.033	0.057	0.022
MLM						$a_c = 0$, $a_w = 0.7$						
Outcome[C]	0.044	0.057	0.014	0.054	0.015	0.031	0.094	0.067	0.065	0.031	0.054	0.022
Outcome[C, W]	0.012	0.008	-0.001	0.006	0.006	0.002	0.089	0.062	0.063	0.028	0.052	0.020
IPW[C]	0.043	0.058	0.018	0.055	0.015	0.033	0.094	0.068	0.066	0.031	0.054	0.022
IPW[C, W]	0.013	0.020	0.002	0.018	0.006	0.003	0.095	0.065	0.066	0.030	0.054	0.022
MLM						$a_c = 0.7$, $a_w = 0.7$						
Outcome[C]	0.037	0.062	0.013	0.059	0.012	0.031	0.093	0.073	0.068	0.032	0.056	0.022
Outcome[C, W]	0.004	0.012	-0.004	0.011	0.003	0.002	0.088	0.068	0.065	0.028	0.053	0.020
IPW[C]	0.068	0.105	0.036	0.101	0.025	0.059	0.099	0.083	0.070	0.039	0.056	0.024
IPW[C, W]	0.042	0.081	0.020	0.077	0.015	0.036	0.101	0.080	0.072	0.037	0.058	0.023

Note. The values were averaged across all ICC levels and combinations of the between- and within-cluster effects of C on the outcome under treatment.

Table 7

Type I error rates for varied cluster sizes (m) and numbers of clusters (J) from the MLM-based methods adjusting for C only (i.e., Outcome[C] and IPW[C] under MLM) and adjusting for both C and W (i.e., Outcome[C, W] and IPW[C, W] under MLM). The average causal treatment effect is zero.

	$J \times m = 100$		$J \times m = 200$		$J \times m = 400$	
	10×10	20×5	10×20	40×5	10×40	40×10
MLM						
Outcome[C]	**0.088**	0.063	**0.083**	0.053	**0.081**	0.050
Outcome[C, W]	**0.083**	0.067	**0.082**	0.051	**0.080**	0.053
IPW[C] (sandwich SE)	0.073	0.063	0.067	0.045	0.070	0.040
(regular SE)	**0.102**	**0.091**	**0.084**	0.065	**0.084**	0.057
IPW[C, W] (sandwich SE)	0.072	0.061	0.069	0.044	0.070	0.039
(regular SE)	**0.100**	**0.092**	**0.085**	0.065	**0.085**	0.058
MLM						
Outcome[C]	**0.091**	0.073	**0.080**	0.065	**0.078**	0.063
Outcome[C, W]	**0.089**	0.063	**0.082**	0.056	0.075	0.053
IPW[C] (sandwich SE)	0.069	0.055	0.065	0.048	0.069	0.049
(regular SE)	**0.088**	**0.079**	**0.079**	0.067	**0.078**	0.066
IPW[C, W] (sandwich SE)	0.068	0.043	0.066	0.047	0.071	0.039
(regular SE)	**0.095**	0.066	**0.081**	0.068	**0.080**	0.054
MLM						
Outcome[C]	**0.085**	**0.077**	**0.085**	0.062	**0.080**	0.059
Outcome[C, W]	**0.084**	0.069	**0.083**	0.052	0.073	0.051
IPW[C] (sandwich SE)	**0.081**	**0.086**	0.070	0.066	0.075	0.057
(regular SE)	**0.107**	**0.116**	**0.093**	**0.088**	**0.089**	**0.083**
IPW[C, W] (sandwich SE)	**0.078**	0.072	0.068	0.059	0.069	0.044
(regular SE)	**0.106**	**0.103**	**0.088**	**0.080**	**0.084**	0.061

Note. The values were averaged across all ICC levels and combinations of the between- and within-cluster effects of C on the outcome under treatment. Boldface font is used for Type I error rates below 0.025 or above 0.075.

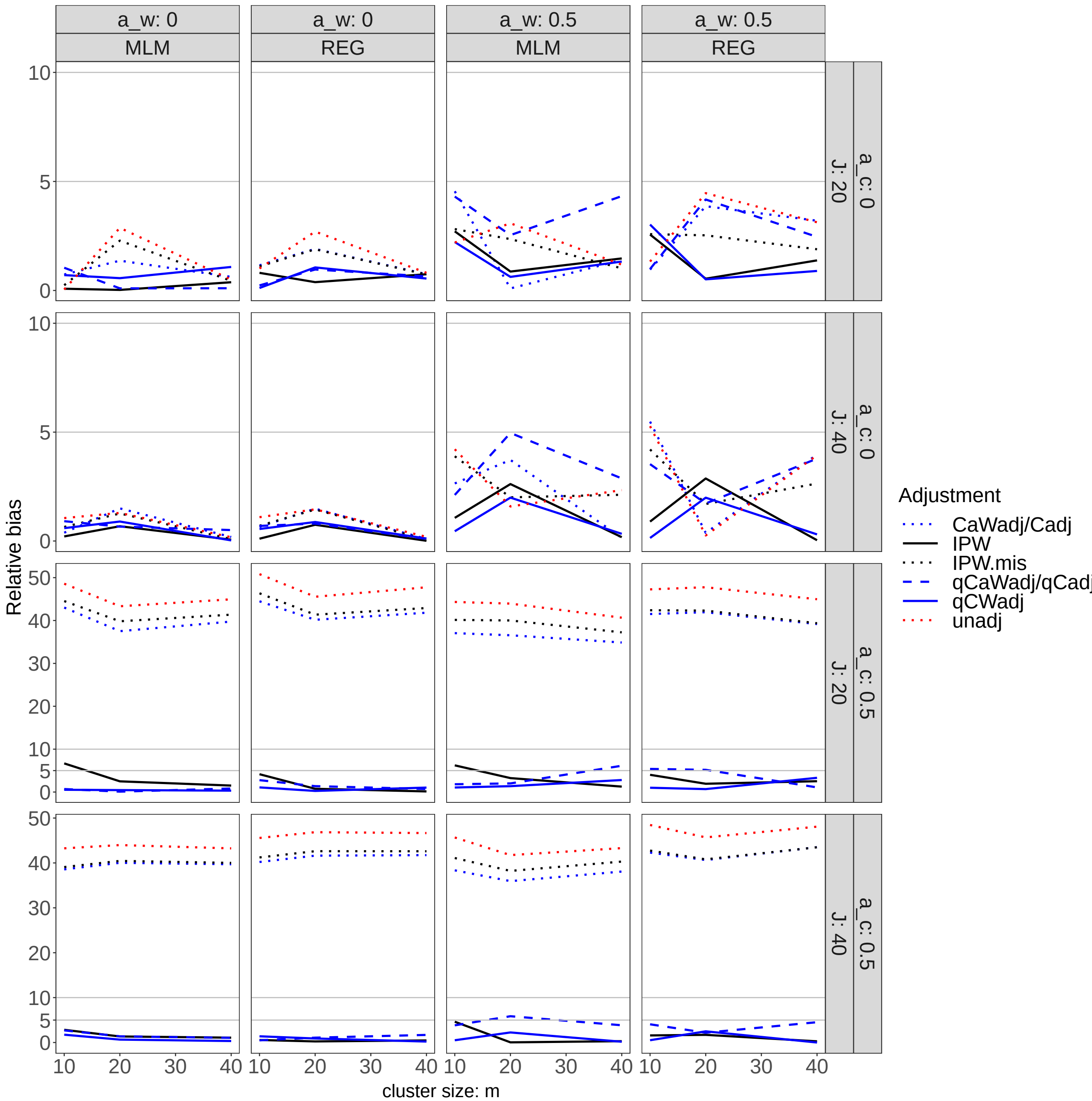

a_w: 0
MLM
a_w: 0
REG
a_w: 0.5
MLM
a_w: 0.5
REG
J: 20
a_c: 0
J: 40
a_c: 0
J: 20
a_c: 0.5
J: 40
a_c: 0.5
Relative bias
cluster size: m
Adjustment
CaWadj/Cadj
IPW
IPW.mis
qCaWadj/qCadj
qCWadj
unadj

Figure 1. Results from the pilot simulation study: Relative bias (%) when average causal effect is 0.41 and ICC is 0.33. CaWadj/Cadj under MLM: adjustment for C, $\bar{W}_k$ in the treatment group multilevel model and for C in the control group regression; CaWadj/Cadj under REG: regression model adjusted for C; IPW: inverse propensity weighting adjusted for C, W, WC, C^2 in the treatment group PS model and for C, C^2 in the control group PS model; IPW.mis: inverse propensity weighting adjusted for C and W in the treatment group PS model and for C in the control group PS model; qCaWadj/qCadj under MLM: adjustment for C, $\bar{W}_k$, $\bar{W}_k C$, C^2 in the treatment group multilevel model and for C, C^2 in the control group regression; qCaWadj/qCadj under REG: regression model adjusted for C, C^2; qCWadj under MLM: adjustment for C, W, WC, C^2 in the treatment group multilevel model and for C, C^2 in the control group regression; qCWadj under REG: regression model adjusted for C, W, WC, C^2 in the treatment group and for C, C^2 in the control group; unadj: outcome modeling without adjusting for covariates.

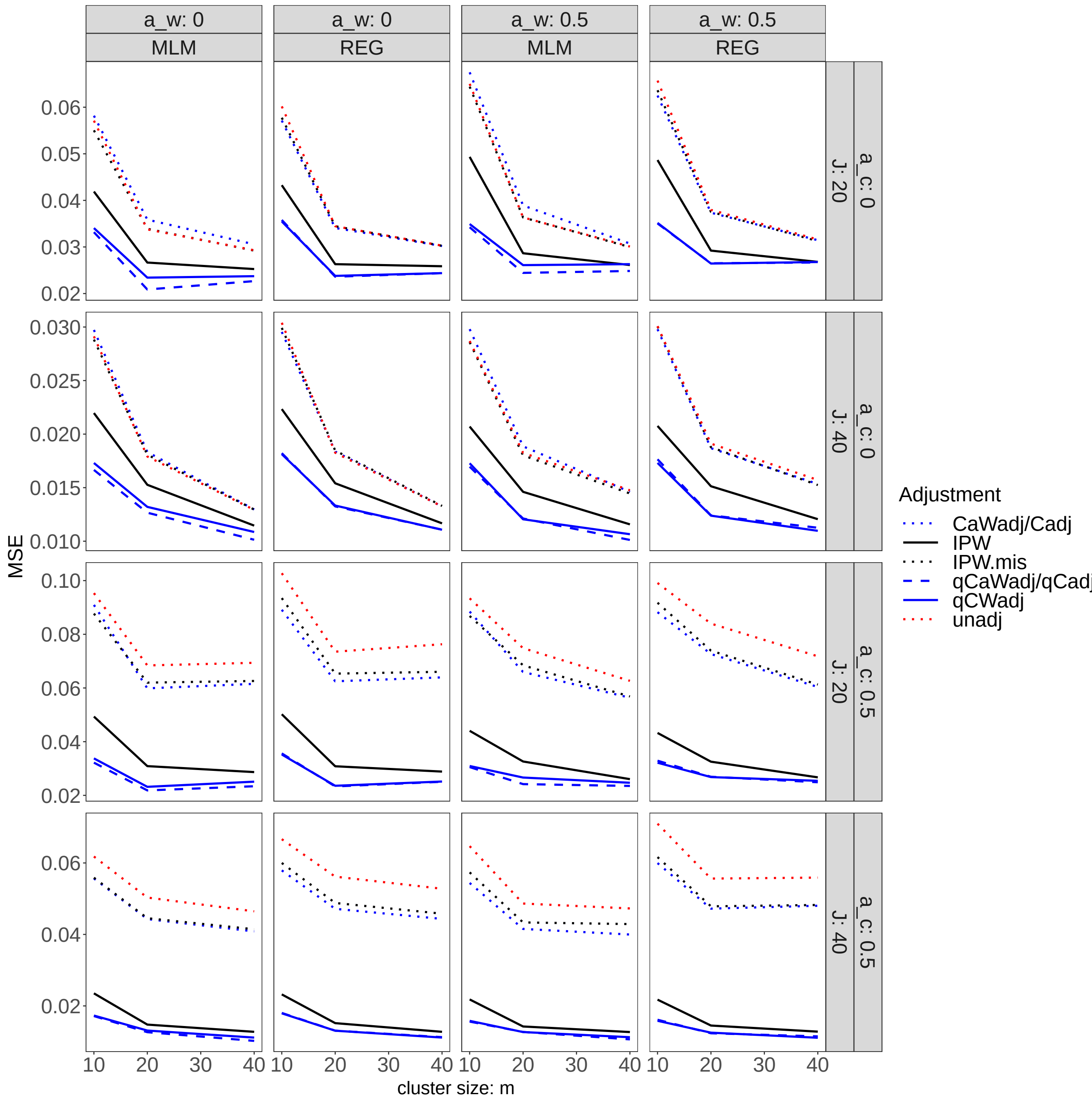
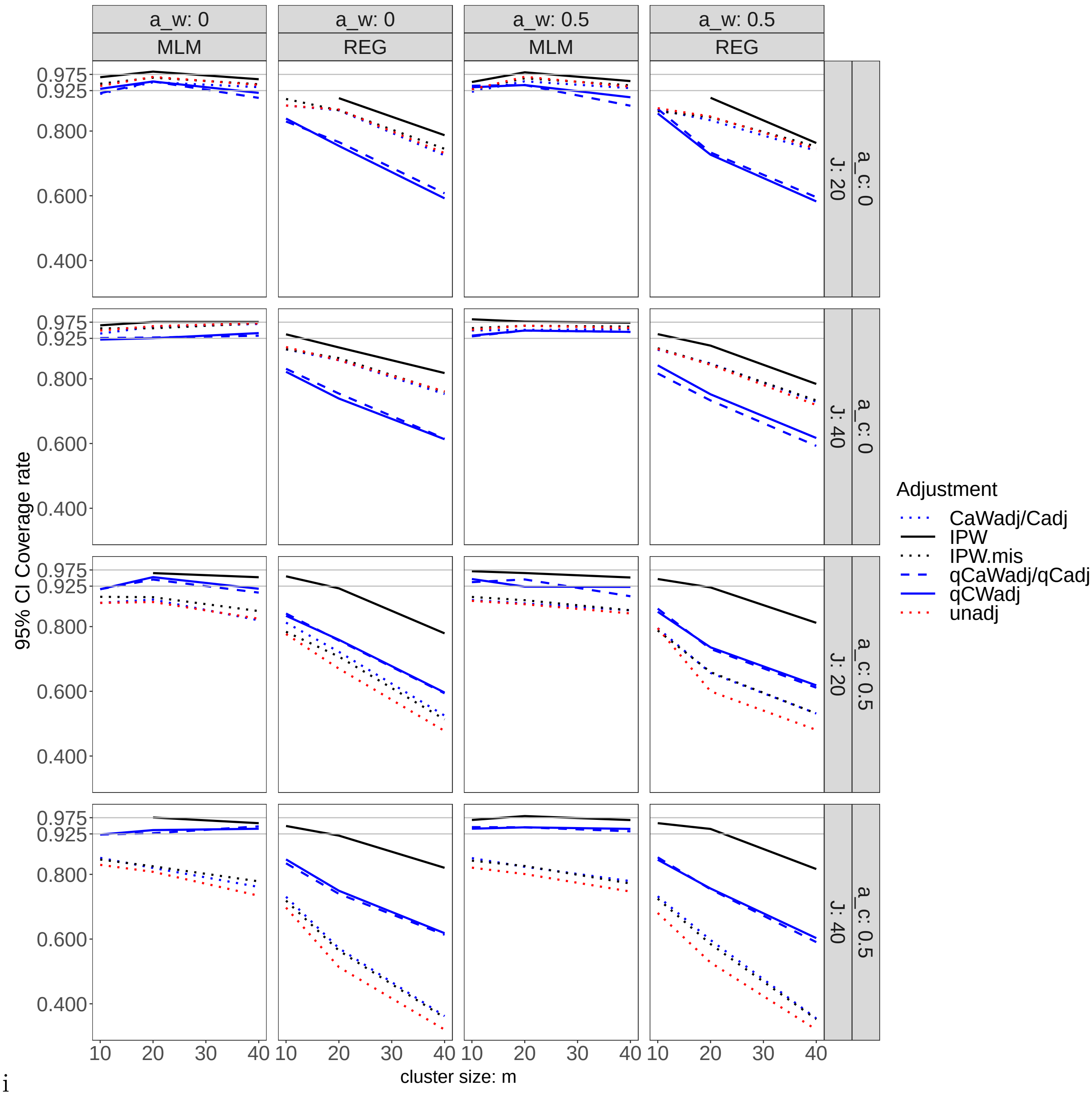
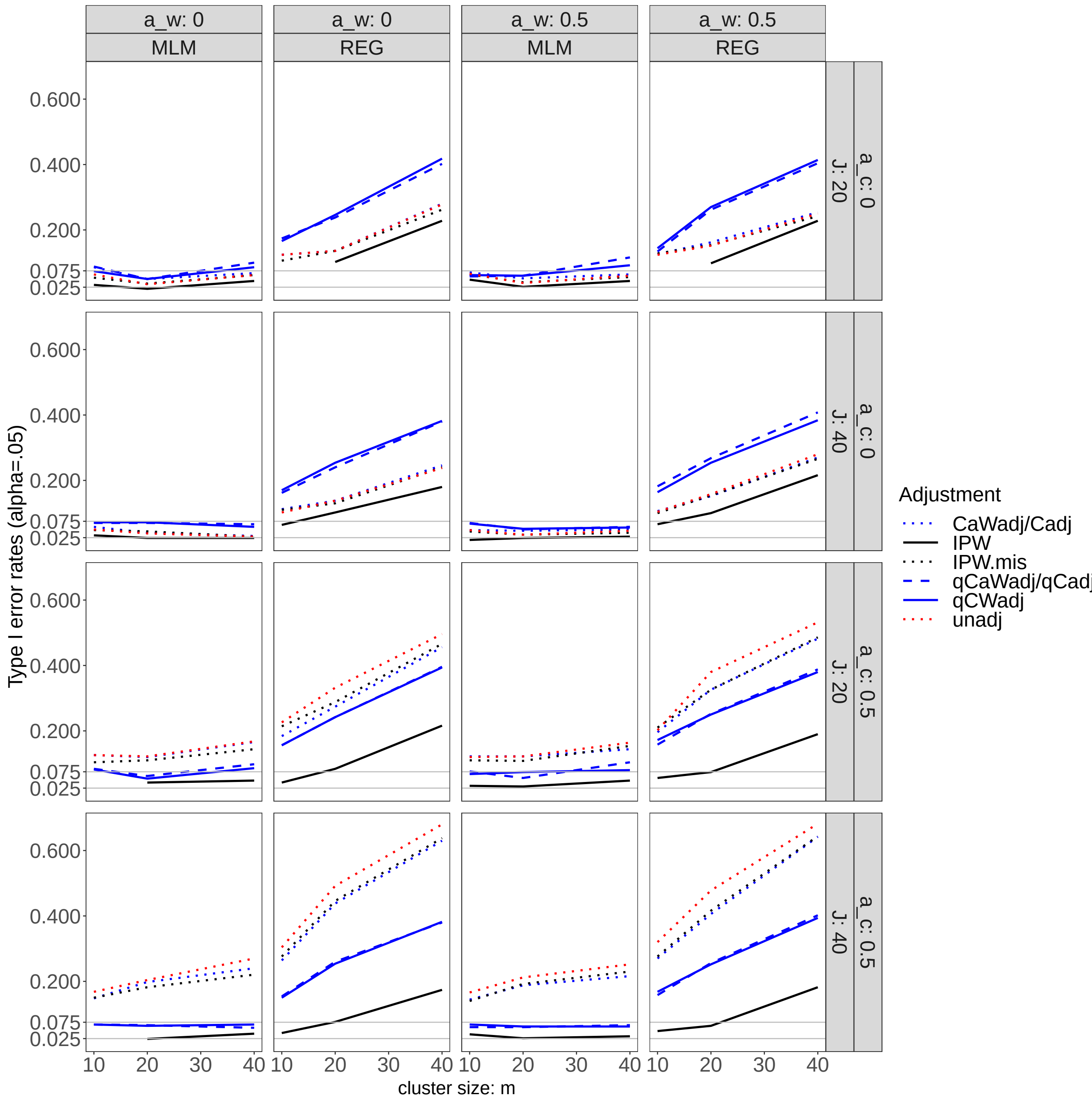

a_w: 0
MLM
a_w: 0
REG
a_w: 0.5
MLM
a_w: 0.5
REG
J: 20
a_c: 0
J: 40
a_c: 0
J: 20
a_c: 0.5
J: 40
a_c: 0.5
MSE
cluster size: m
Adjustment
CaWadj/Cadj
IPW
IPW.mis
qCaWadj/qCadj
qCWadj
unadj

Figure 2. Results from the pilot simulation study: MSE in when average causal effect is 0.41 and ICC is 0.33. CaWadj/Cadj under MLM: adjustment for C, $\bar{W}_k$ in the treatment group multilevel model and for C in the control group regression; CaWadj/Cadj under REG: regression model adjusted for C; IPW: inverse propensity weighting adjusted for C, W, WC, C^2 in the treatment group PS model and for C, C^2 in the control group PS model; IPW.mis: inverse propensity weighting adjusted for C and W in the treatment group PS model and for C in the control group PS model; qCaWadj/qCadj under MLM: adjustment for C, $\bar{W}_k$, $\bar{W}_k C$, C^2 in the treatment group multilevel model and for C, C^2 in the control group regression; qCaWadj/qCadj under REG: regression model adjusted for C, C^2; qCWadj under MLM: adjustment for C, W, WC, C^2 in the treatment group multilevel model and for C, C^2 in the control group regression; qCWadj under REG: regression model adjusted for C, W, WC, C^2 in the treatment group and for C, C^2 in the control group; unadj: outcome modeling without adjusting for covariates.


a_w: 0
MLM
a_w: 0
REG
a_w: 0.5
MLM
a_w: 0.5
REG
J: 20
a_c: 0
J: 40
a_c: 0
J: 20
a_c: 0.5
J: 40
a_c: 0.5
95% CI Coverage rate
0.975
0.925
0.800
0.600
0.400
10
20
30
40
cluster size: m
Adjustment
CaWadj/Cadj
IPW
IPW.mis
qCaWadj/qCadj
qCWadj
unadj

i

Figure 3. Results from the pilot simulation study: Coverage rates of 95% confidence interval when average causal effect is 0.41 and ICC is 0.33. CaWadj/Cadj under MLM: adjustment for C, $\bar{W}_k$ in the treatment group multilevel model and for C in the control group regression; CaWadj/Cadj under REG: regression model adjusted for C; IPW: inverse propensity weighting adjusted for C, W, WC, C^2 in the treatment group PS model and for C, C^2 in the control group PS model; IPW.mis: inverse propensity weighting adjusted for C and W in the treatment group PS model and for C in the control group PS model; qCaWadj/qCadj under MLM: adjustment for C, $\bar{W}_k$, $\bar{W}_k C$, C^2 in the treatment group multilevel model and for C, C^2 in the control group regression; qCaWadj/qCadj under REG: regression model adjusted for C, C^2; qCWadj under MLM: adjustment for C, W, WC, C^2 in the treatment group multilevel model and for C, C^2 in the control group regression; qCWadj under REG: regression model adjusted for C, W, WC, C^2 in the treatment group and for C, C^2 in the control group; unadj: outcome modeling without adjusting for covariates.



a_w: 0
MLM
a_w: 0
REG
a_w: 0.5
MLM
a_w: 0.5
REG
J: 20
a_c: 0
J: 40
a_c: 0
J: 20
a_c: 0.5
J: 40
a_c: 0.5
0.600
0.400
0.200
0.075
0.025
Type I error rates (alpha=.05)
10
20
30
40
cluster size: m
Adjustment
CaWadj/Cadj
IPW
IPW.mis
qCaWadj/qCadj
qCWadj
unadj

Figure 4. Results from the pilot simulation study: Type I error rates at the 0.05 significance level in the null hypothesis testing of the average causal effect when average causal effect is 0 and ICC is 0.33. CaWadj/Cadj under MLM: adjustment for C, $\bar{W}_k$ in the treatment group multilevel model and for C in the control group regression; CaWadj/Cadj under REG: regression model adjusted for C; IPW: inverse propensity weighting adjusted for C, W, WC, C^2 in the treatment group PS model and for C, C^2 in the control group PS model; IPW.mis: inverse propensity weighting adjusted for C and W in the treatment group PS model and for C in the control group PS model; qCaWadj/qCadj under MLM: adjustment for C, $\bar{W}_k$, $\bar{W}_k C$, C^2 in the treatment group multilevel model and for C, C^2 in the control group regression; qCaWadj/qCadj under REG: regression model adjusted for C, C^2; qCWadj under MLM: adjustment for C, W, WC, C^2 in the treatment group multilevel model and for C, C^2 in the control group regression; qCWadj under REG: regression model adjusted for C, W, WC, C^2 in the treatment group and for C, C^2 in the control group; unadj: outcome modeling without adjusting for covariates.