Web Appendix for "Sample size requirements for detecting treatment effect heterogeneity in cluster randomized trials" by Yang et al.

A. DERIVING THE VARIANCE EXPRESSION WITH MULTIPLE COVARIATES

Similar to Section 3.1 in the main text, with multiple covariates ($p \geq 2$), we can write $U_n = cS_n + dT_n$, where

$$S_n = \sum_{i=1}^{n}\sum_{j=1}^{m} Z_{ij}Z_{ij}^T = \sum_{i=1}^{n}\sum_{j=1}^{m}\begin{bmatrix} 1 & W_i - \bar{W} & X_{ij}^T & (W_i-\bar{W})X_{ij}^T \\ W_i - \bar{W} & (W_i-\bar{W})^2 & (W_i-\bar{W})X_{ij}^T & (W_i-\bar{W})^2X_{ij}^T \\ X_{ij} & (W_i-\bar{W})X_{ij} & X_{ij}X_{ij}^T & (W_i-\bar{W})X_{ij}X_{ij}^T \\ (W_i-\bar{W})X_{ij} & (W_i-\bar{W})^2X_{ij} & (W_i-\bar{W})X_{ij}X_{ij}^T & (W_i-\bar{W})^2X_{ij}X_{ij}^T \end{bmatrix}$$

and

$$T_n = \sum_{i=1}^{n}\left(\sum_{j=1}^{m} Z_{ij}\right)\left(\sum_{j=1}^{m} Z_{ij}\right)^T = \sum_{i=1}^{n} m^2 \begin{bmatrix} 1 & W_i - \bar{W} & \bar{X}_i^T & (W_i-\bar{W})\bar{X}_i^T \\ W_i - \bar{W} & (W_i-\bar{W})^2 & (W_i-\bar{W})\bar{X}_i^T & (W_i-\bar{W})^2\bar{X}_i^T \\ \bar{X}_i & (W_i-\bar{W})\bar{X}_i & \bar{X}_i\bar{X}_i^T & (W_i-\bar{W})\bar{X}_i\bar{X}_i^T \\ (W_i-\bar{W})\bar{X}_i & (W_i-\bar{W})^2\bar{X}_i & (W_i-\bar{W})\bar{X}_i\bar{X}_i^T & (W_i-\bar{W})^2\bar{X}_i\bar{X}_i^T \end{bmatrix},$$

where $\bar{X}_i = (m^{-1}\sum_{j=1}^m X_{ij})$. We define $h_1 = \lim_{n\to\infty}(nm)^{-1}\sum_{i=1}^n\sum_{j=1}^m X_{ij}$, $H_2 = \lim_{n\to\infty}(nm)^{-1}\sum_{i=1}^n\sum_{j=1}^m X_{ij}X_{ij}^T$, $L_2 = n^{-1}\sum_{i=1}^n \bar{X}_i\bar{X}_i^T$ as the moment vector and matrices. Then the limits of $n^{-1}S_n$ and $n^{-1}T_n$ are obtained as

$$S = \lim_{n\to\infty} n^{-1}S_n = m\begin{bmatrix} 1 & 0 & h_1^T & 0_{p\times1}^T \\ 0 & \sigma_w^2 & 0_{p\times1}^T & \sigma_w^2 h_1^T \\ h_1 & 0_{p\times1} & H_2 & 0_{p\times p} \\ 0_{p\times1} & \sigma_w^2 h_1 & 0_{p\times p} & \sigma_w^2 H_2 \end{bmatrix}$$

and

$$T = \lim_{n\to\infty} n^{-1}T_n = m^2\begin{bmatrix} 1 & 0 & h_1^T & 0_{p\times1}^T \\ 0 & \sigma_w^2 & 0_{p\times1}^T & \sigma_w^2 h_1^T \\ h_1 & 0_{p\times1} & L_2 & 0_{p\times p} \\ 0_{p\times1} & \sigma_w^2 h_1 & 0_{p\times p} & \sigma_w^2 L_2 \end{bmatrix}$$

The expressions of S and T allow us to write

$$U = cS + dT = \begin{bmatrix} A & B \\ C & D \end{bmatrix} = \left[\begin{array}{cc|cc} m(c+dm) & 0 & m(c+dm)h_1^T & 0_{p\times1}^T \\ 0 & m(c+dm)\sigma_w^2 & 0_{p\times1}^T & m(c+dm)\sigma_w^2 h_1^T \\ \hline m(c+dm)h_1 & 0_{p\times1} & m(cH_2+dmL_2) & 0_{p\times p} \\ 0_{p\times1} & m(c+dm)\sigma_w^2 h_1 & 0_{p\times p} & m\sigma_w^2(cH_2+dmL_2) \end{array}\right].$$

Let $\otimes$ denote the Kronecker product. Notice that the off-diagonal block $C = B^T = A \otimes h_1$, therefore by block matrix inversion, we obtain the lower-right $p\times p$ block of U^{-1} as

$$(D - CA^{-1}B)^{-1} = m^{-1}\begin{bmatrix} c(H_2 - h_1h_1^T) + dm(L_2 - h_1h_1^T) & 0_{p\times p} \\ 0_{p\times p} & \sigma_w^2[c(H_2 - h_1h_1^T) + dm(L_2 - h_1h_1^T)] \end{bmatrix}^{-1},$$

and therefore

$$\Omega_4 = \frac{\sigma_{y|x}^2(1-\rho_{y|x})}{m\sigma_w^2}\left\{\frac{1-\rho_{y|x}}{1+(m-1)\rho_{y|x}}(H_2 - h_1h_1^T) + \frac{m\rho_{y|x}}{1+(m-1)\rho_{y|x}}(H_2 - L_2)\right\}^{-1},$$

where the inverse operator is applied toward the weighted combination between the marginal covariance matrix $H_2 - h_1h_1^T$ and the difference between two covariances $H_2 - L_2$.

Similar to Section 3.1 of the main text, we rewrite the above variance expression by introducing the two correlation matrices. The first correlation matrix summarizes the marginal correlation between p covariates of interest, and is defined as

$$\Gamma_x^1 = \Omega_x^{-1/2}\left\{E(X_{ij}X_{ij}^T) - h_1h_1^T\right\}\Omega_x^{-1/2} = \Omega_x^{-1/2}\left\{H_2 - h_1h_1^T\right\}\Omega_x^{-1/2},$$

where $\Omega_x = \text{diag}(H_2 - h_1 h_1^T)$ is the diagonal matrix containing the marginal variances of all covariates. In other words, the diagonal element of Γ_x^1 is one and the off-diagonal elements represent the marginal correlation between each pair of covariates. The second correlation differs from the first one in a subtle way, and is defined as

$$\Gamma_x^0 = \Omega_x^{-1/2} \left\{ E(X_{ij} X_{ik}^T) - h_1 h_1^T \right\} \Omega_x^{-1/2},$$

which could be regarded as a multivariate extension of ICC ρ_x. Specifically, the diagonal element of Γ_x^0 is the ICC of each covariate, and the off-diagonal elements are the intraclass cross-correlations between different covariates. Now observe that

$$L_2 = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \left(\frac{1}{m} \sum_{j=1}^{m} X_{ij} \right) \left(\frac{1}{m} \sum_{j=1}^{m} X_{ij}^T \right) = \frac{1}{m} H_2 + \frac{1}{m} (m-1) \left(h_1 h_1^T + \Omega_x^{1/2} \Gamma_x \Omega_x^{1/2} \right).$$

Some algebra then gives the variance expression

$$\Omega_4 = \frac{\sigma_{y|x}^2 (1-\rho_{y|x}) \left\{ 1 + (m-1)\rho_{y|x} \right\}}{m\sigma_w^2} \Omega_x^{-1/2} \left\{ \Gamma_x^1 + (m-2)\rho_{y|x}\Gamma_x^1 - (m-1)\rho_{y|x}\Gamma_x^0 \right\}^{-1} \Omega_x^{-1/2}. \tag{1}$$

A.1 An illustrative example of definitions of Γ_x^1 and Γ_x^0

We use a simple multilevel exchangeable model to illustrate the definition of the two matrices Γ_x^1 and Γ_x^0 introduced above. Assume the covariates are generated from the following random effects model

$$X_{ijp} = \tau_{ip} + c_i + s_j + (cs)_{ij} + \zeta_{ijp}, \tag{2}$$

where τ_{ip} is the average of pth covariate in cluster i, $c_i \sim \mathcal{N}(0, \sigma_c^2)$, $s_j \sim \mathcal{N}(0, \sigma_s^2)$, $(cs)_{ij} \sim \mathcal{N}(0, \sigma_{cs}^2)$, and $\zeta_{ijp} \sim \mathcal{N}(0, \sigma_\zeta^2)$ are independent random effects that represent the cluster effect, individual effect, cluster-by-individual interaction and measurement error. Across all clusters, the marginal (exchangeable) correlation between two covariates X_{ijp} and $X_{ijp'}$ $(p \neq p')$ is

$$\rho_x^1 = \frac{\sigma_c^2 + \sigma_s^2 + \sigma_{cs}^2}{\sigma_c^2 + \sigma_s^2 + \sigma_{cs}^2 + \sigma_\zeta^2},$$

and the matrix Γ_x^1 is then given by

$$\Gamma_x^1 = \begin{pmatrix} 1 & \rho_x^1 & \rho_x^1 & \dots & \rho_x^1 \\ \rho_x^1 & 1 & \rho_x^1 & \dots & \rho_x^1 \\ \rho_x^1 & \rho_x^1 & 1 & \dots & \rho_x^1 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_x^1 & \rho_x^1 & \rho_x^1 & \dots & \rho_x^1 \end{pmatrix} = (1-\rho_x^1) I_p + \rho_x^1 J_p.$$

Further, recall that the diagonal element of Γ_x^0 describes the ICC between X_{ijp} and X_{ikp} $(j \neq k)$, and is

$$\rho_x^0 = \frac{\sigma_c^2}{\sigma_c^2 + \sigma_s^2 + \sigma_{cs}^2 + \sigma_\zeta^2},$$

while in this multilevel exchangeable model, the intraclass cross-correlation between X_{ijp} and $X_{ikp'}$ $(j \neq k, p \neq p')$ is still ρ_x^0. This leads to an expression of

$$\Gamma_x^0 = \begin{pmatrix} \rho_x^0 & \rho_x^0 & \rho_x^0 & \dots & \rho_x^0 \\ \rho_x^0 & \rho_x^0 & \rho_x^0 & \dots & \rho_x^0 \\ \rho_x^0 & \rho_x^0 & \rho_x^0 & \dots & \rho_x^0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_x^0 & \rho_x^0 & \rho_x^0 & \dots & \rho_x^0 \end{pmatrix} = \rho_x^0 J_p.$$

In this simple and special example, we can rewrite variance expression (1) as

$$\begin{aligned} \Omega_4 &= \frac{\sigma_{y|x}^2 (1-\rho_{y|x}) \left\{ 1 + (m-1)\rho_{y|x} \right\}}{m\sigma_w^2} \Omega_x^{-1/2} \left\{ \Gamma_x^1 + (m-2)\rho_{y|x}\Gamma_x^1 - (m-1)\rho_{y|x}\Gamma_x^0 \right\}^{-1} \Omega_x^{-1/2} \\ &= \frac{\sigma_{y|x}^2 (1-\rho_{y|x}) \left\{ 1 + (m-1)\rho_{y|x} \right\}}{m\sigma_w^2} \Omega_x^{-1/2} \left\{ (1-\rho_x^1)\{1+(m-2)\rho_{y|x}\} I_m + \{\rho_x^1 + (m-2)\rho_{y|x}\rho_x^1 - (m-1)\rho_{y|x}\rho_x^0\} J_m \right\}^{-1} \Omega_x^{-1/2} \end{aligned}$$

A.2 Explicit variance expression when $p = 2$

When $p = 2$ covariates are considered, it is possible to provide an explicit inverse of the covariance in equation (1). We provide such an expression here to obtain some understanding of the impact of covariate ICC in this slightly more general setting. In this case, we generically denote the marginal correlation between X_{ij1} and X_{ij2} as ρ_x^1, in which case the matrix Γ_x^1 is

$$\Gamma_x^1 = \begin{pmatrix} 1 & \rho_x^1 \\ \rho_x^1 & 1 \end{pmatrix} = (1 - \rho_x^1)I_2 + \rho_x^1 J_2.$$

Further, we generically write pth diagonal element of Γ_x^0 as $\rho_{x_p}^0$, which defines the ICC between X_{ijp} and X_{ikp} $(j \neq k)$. Similarly, we write intraclass cross-correlation between X_{ij1} and X_{ik2} $(j \neq k)$ as $\rho_{x_{12}}^0$. This leads to the expression

$$\Gamma_x^0 = \begin{pmatrix} \rho_{x_1}^0 & \rho_{x_{12}}^0 \\ \rho_{x_{12}}^0 & \rho_{x_2}^0 \end{pmatrix}.$$

Denote

$$\begin{aligned} M &= \Gamma_x^1 + (m-2)\rho_{y|x}\Gamma_x^1 - (m-1)\rho_{y|x}\Gamma_x^0 \\ &= \begin{bmatrix} 1 + (m-2)\rho_{y|x} - (m-1)\rho_{y|x}\rho_{x_1}^0 & \left\{1 + (m-2)\rho_{y|x}\right\}\rho_x^1 - (m-1)\rho_{y|x}\rho_{x_{12}}^0 \\ \left\{1 + (m-2)\rho_{y|x}\right\}\rho_x^1 - (m-1)\rho_{y|x}\rho_{x_{12}}^0 & 1 + (m-2)\rho_{y|x} - (m-1)\rho_{y|x}\rho_{x_2}^0 \end{bmatrix}, \end{aligned}$$

the determinant is

$$\det(M) = \left[1 + (m-2)\rho_{y|x} - (m-1)\rho_{y|x}\rho_{x_1}^0\right]\left[1 + (m-2)\rho_{y|x} - (m-1)\rho_{y|x}\rho_{x_2}^0\right] - \left[\left\{1 + (m-2)\rho_{y|x}\right\}\rho_x^1 - (m-1)\rho_{y|x}\rho_{x_{12}}^0\right]^2.$$

Therefore, we have

$$\begin{aligned} \Omega_4 &= \frac{\sigma_{y|x}^2(1-\rho_{y|x})\left\{1 + (m-1)\rho_{y|x}\right\}}{m\sigma_w^2}\Omega_x^{-1/2}M^{-1}\Omega_x^{-1/2} \\ &= \frac{\sigma_{y|x}^2(1-\rho_{y|x})\left\{1 + (m-1)\rho_{y|x}\right\}}{m\sigma_w^2\det(M)} \times \\ &\begin{bmatrix} \sigma_1^{-2}\left\{1 + (m-2)\rho_{y|x} - (m-1)\rho_{y|x}\rho_{x_2}^0\right\} & (\sigma_1\sigma_2)^{-1}\left\{(m-1)\rho_{y|x}\rho_{x_{12}}^0 - \left\{1 + (m-2)\rho_{y|x}\right\}\rho_x^1\right\} \\ (\sigma_1\sigma_2)^{-1}\left\{(m-1)\rho_{y|x}\rho_{x_{12}}^0 - \left\{1 + (m-2)\rho_{y|x}\right\}\rho_x^1\right\} & \sigma_2^{-2}\left\{1 + (m-2)\rho_{y|x} - (m-1)\rho_{y|x}\rho_{x_1}^0\right\} \end{bmatrix}, \end{aligned}$$

where σ_p^2 is the marginal variance for covariate X_p, $p = 1, 2$. From this expression, we can see that the $(1, 1)$th element of Ω_4 is proportional to

$$\left(\left[1 + (m-2)\rho_{y|x} - (m-1)\rho_{y|x}\rho_{x_1}^0\right] - \frac{\left[\left\{1 + (m-2)\rho_{y|x}\right\}\rho_x^1 - (m-1)\rho_{y|x}\rho_{x_{12}}^0\right]^2}{1 + (m-2)\rho_{y|x} - (m-1)\rho_{y|x}\rho_{x_2}^0}\right)^{-1},$$

which is an increasing function of $\rho_{x_1}^0$ and ρ_x^1. This confirms that larger covariate ICC and larger marginal correlation between the two covariates both increase the variance of the estimator of the interaction coefficient and so will inflate the required sample size. The same reasonsing can be applied to the $(2, 2)$ element of Ω_4. However, the role of the intraclass cross-correlation is not as clear in the expression, and remains to be explored in future work.

B. MARGINALIZE THE COVARIATE-ADJUSTED LINEAR MIXED MODEL IN SECTION 4.2

Recall that the conditional outcome model with a single covariate X_{ij} is defined as

$$Y_{ij} = \beta_1 + \beta_2 W_i + \beta_3 X_{ij} + \beta_4 X_{ij} W_i + \gamma_i + \epsilon_{ij} \tag{3}$$

where we assume in model (3) $\gamma_i \sim \mathcal{N}(0, \sigma_\gamma^2)$ and $\epsilon_{ij} \sim \mathcal{N}(0, \sigma_\epsilon^2)$, and independence between γ_i and ϵ_{ij}. In Section 4.2 (and also in our simulation study with a continuous covariate), we assume $X_{ij} = \mu + \mu_i + \tau_{ij}$, where $\mu_i \sim \mathcal{N}(0, \sigma_\mu^2)$, and $\tau_{ij} \sim \mathcal{N}(0, \sigma_\tau^2)$.

The marginal variance of X_{ij} is therefore $\sigma_x^2 = \sigma_\mu^2 + \sigma_\tau^2$, and the marginal covariate ICC is $\rho_x = \sigma_\mu^2/(\sigma_\mu^2 + \sigma_\tau^2)$. In this case, we can expand model (3) as

$$\begin{aligned} Y_{ij} &= \beta_1 + \beta_2 W_i + \beta_3(\mu + \mu_i + \tau_{ij}) + \beta_4(\mu + \mu_i + \tau_{ij})W_i + \gamma_i + \epsilon_{ij} \\ &= (\beta_1 + \beta_3\mu) + (\beta_2 + \beta_4\mu)W_i + (\beta_3\mu_i + \beta_4 W_i\mu_i + \gamma_i) + (\beta_3\tau_{ij} + \beta_4 W_i\tau_{ij} + \epsilon_{ij}) \\ &= \alpha_1 + \alpha_2 W_i + \lambda_i + \xi_{ij}, \end{aligned}$$

where we treat $\alpha_1 = \beta_1 + \beta_3\mu$, $\alpha_2 = \beta_2 + \beta_4\mu$, $\lambda_i = \beta_3\mu_i + \beta_4 W_i\mu_i + \gamma_i$ and $\xi_{ij} = \beta_3\tau_{ij} + \beta_4 W_i\tau_{ij} + \epsilon_{ij}$. The following observations can be made on this expansion:

- In the special case where the covariate is mean centered so that $\mu = 0$, it is evident that $\alpha_1 = \beta_1$ and $\alpha_2 = \beta_2$. In other words, we can interpret the main effect of W_i as the OTE. Similar results have been extensively discussed in the literature of individually randomized trials; see, for example, the "ANCOVA II" model in Yang and Tsiatis[1]. This observation can be treated as a re-interpretation of Yang and Tsiatis[1] in the context of random-effects ANCOVA with a cluster-level treatment. In what follows, we will consistently assume $\mu = 0$. Notice we have made the assumption $\mu = 0$ in Section 4.2 of the manuscript.

- If the interaction effect $\beta_4 = 0$, then $\lambda_i = \beta_3\mu_i + \gamma_i \sim \mathcal{N}(0, \beta_3^2\sigma_\mu^2 + \sigma_\gamma^2)$ and $\xi_{ij} = \beta_3\tau_{ij} + \epsilon_{ij} \sim \mathcal{N}(0, \beta_3^2\sigma_\tau^2 + \sigma_\epsilon^2)$. In this special case, we obtain the main-effects-only model as in Raudenbush[2], Yang et al.[3] and Li et al.[4], among others. Because the induced unadjusted random effects λ_i and ξ_{ij} are exactly normally distributed. The unadjusted model (2) in the main text and the adjusted model (3) hold simultaneously.

- In the more general case (which is the case we are interested in) where $\beta_4 \neq 0$, then we have $W_i \sim \text{Bernoulli}(\bar{W})$, $\lambda_i|W_i = \beta_3\mu_i + \beta_4 W_i\mu_i + \gamma_i \sim \mathcal{N}(0, \beta_3^2\sigma_\mu^2 + \beta_4^2 W_i\sigma_\mu^2 + 2\beta_3\beta_4 W_i\sigma_\mu^2 + \sigma_\gamma^2)$ and $\xi_{ij}|W_i = \beta_3\tau_{ij} + \beta_4 W_i\tau_{ij} + \epsilon_{ij} \sim \mathcal{N}(0, \beta_3^2\sigma_\tau^2 + \beta_4^2 W_i\sigma_\tau^2 + 2\beta_3\beta_4 W_i\sigma_\tau^2 + \sigma_\epsilon^2)$. Marginalizing over W_i, both λ_i and ξ_{ij} follow mixtures of two normal distributions, and so are no longer univariate normal. In this case, one could identify the best approximating unadjusted model to model (3) by obtaining the unadjusted variance components. However, because λ_i and ξ_{ij} follow mixture normal distributions, the unadjusted model (2) in the main text (which assumes normal random effects) does not hold exactly. This explains why in general unadjusted model (2) in the main text and the adjusted model (3) do not hold simultaneously.

Following the last bullet point, the best approximating unadjusted model can be identified once we compute the unadjusted variance components for λ_i and ξ_{ij}. By the law of total variance,

$$\begin{aligned} \sigma_\lambda^2 &= \text{var}(\lambda_i) = E[\text{var}(\lambda_i|W_i)] + \text{var}[E(\lambda_i|W_i)] \\ &= E[\sigma_\gamma^2 + (\beta_3^2 + \beta_4^2 W_i + 2\beta_3\beta_4 W_i)\sigma_\mu^2] = \sigma_\gamma^2 + (\beta_3^2 + \beta_4^2\bar{W} + 2\beta_3\beta_4\bar{W})\sigma_\mu^2, \\ \sigma_\xi^2 &= \text{var}(\xi_{ij}) = E[\text{var}(\xi_{ij}|W_i)] + \text{var}[E(\xi_{ij}|W_i)] \\ &= E[\sigma_\epsilon^2 + (\beta_3^2 + \beta_4^2 W_i + 2\beta_3\beta_4 W_i)\sigma_\tau^2] = \sigma_\epsilon^2 + (\beta_3^2 + \beta_4^2\bar{W} + 2\beta_3\beta_4\bar{W})\sigma_\tau^2, \\ \sigma_{\lambda,\xi} &= \text{cov}(\lambda_i, \xi_{ij}) = E[\text{cov}(\lambda_i, \xi_{ij}|W_i)] + \text{cov}[E(\lambda_i|W_i), E(\xi_{ij}|W_i)] = 0 \end{aligned}$$

Define the unadjusted variance components $\sigma_y^2 = \sigma_\lambda^2 + \sigma_\xi^2$, and write $B = \beta_3^2 + \beta_4^2\bar{W} + 2\beta_3\beta_4\bar{W}$, the unadjusted outcome ICC can be reasonably approximated by

$$\rho_y = \frac{\sigma_\lambda^2}{\sigma_y^2} = \frac{\sigma_\gamma^2 + B\sigma_\mu^2}{\sigma_\gamma^2 + \sigma_\epsilon^2 + B(\sigma_\mu^2 + \sigma_\tau^2)} = \frac{\sigma_{y|x}^2}{\sigma_{y|x}^2 + B\sigma_x^2}\rho_{y|x} + \frac{B\sigma_x^2}{\sigma_{y|x}^2 + B\sigma_x^2}\rho_x = \omega\rho_{y|x} + (1-\omega)\rho_x,$$

which clearly is a weighted combination between the adjusted outcome ICC and the covariate ICC.

Now in order to gain some insights from equation (24) in Section 4.2, we plug equation (23) into equation (24) in the main text. This gives us

$$\Theta(m) = \frac{\sigma_{y|x}^2(1-\rho_{y|x})}{\sigma_y^2\sigma_x^2\left\{1 + (m-2)\rho_{y|x} - (m-1)\rho_x\rho_{y|x}\right\}} \times \frac{1 + (m-1)\rho_{y|x}}{1 + (m-1)\omega\rho_{y|x} + (m-1)(1-\omega)\rho_x} \times \frac{1}{\text{RDES}^2}. \tag{4}$$

Next we define

$$f(\rho_x) = \left\{1 + (m-2)\rho_{y|x} - (m-1)\rho_x\rho_{y|x}\right\}\left\{1 + (m-1)\omega\rho_{y|x} + (m-1)(1-\omega)\rho_x\right\}$$
$$= -(1-\omega)(m-1)^2\rho_x^2 - \left[(m-1)\rho_{y|x} + (m-1)^2\omega\rho_{y|x}^2\right]\rho_x + \left[(1-\omega)(m-1) + (m-2)(1-\omega)(m-1)\rho_{y|x}\right]\rho_x +$$
$$\left[1 + (m-2)\rho_{y|x}\right]\left[1 + (m-1)\omega\rho_{y|x}\right],$$

which is a quadratic function of ρ_x. We can immediately recognize two roots of $f(\rho_x)$=0:

$$\rho_{x,1} = \frac{1 + (m-2)\rho_{y|x}}{(m-1)\rho_{y|x}} > 1 \quad \text{and} \quad \rho_{x,2} = \frac{-1 - (m-1)\omega\rho_{y|x}}{(1-\omega)(m-1)} < 0.$$

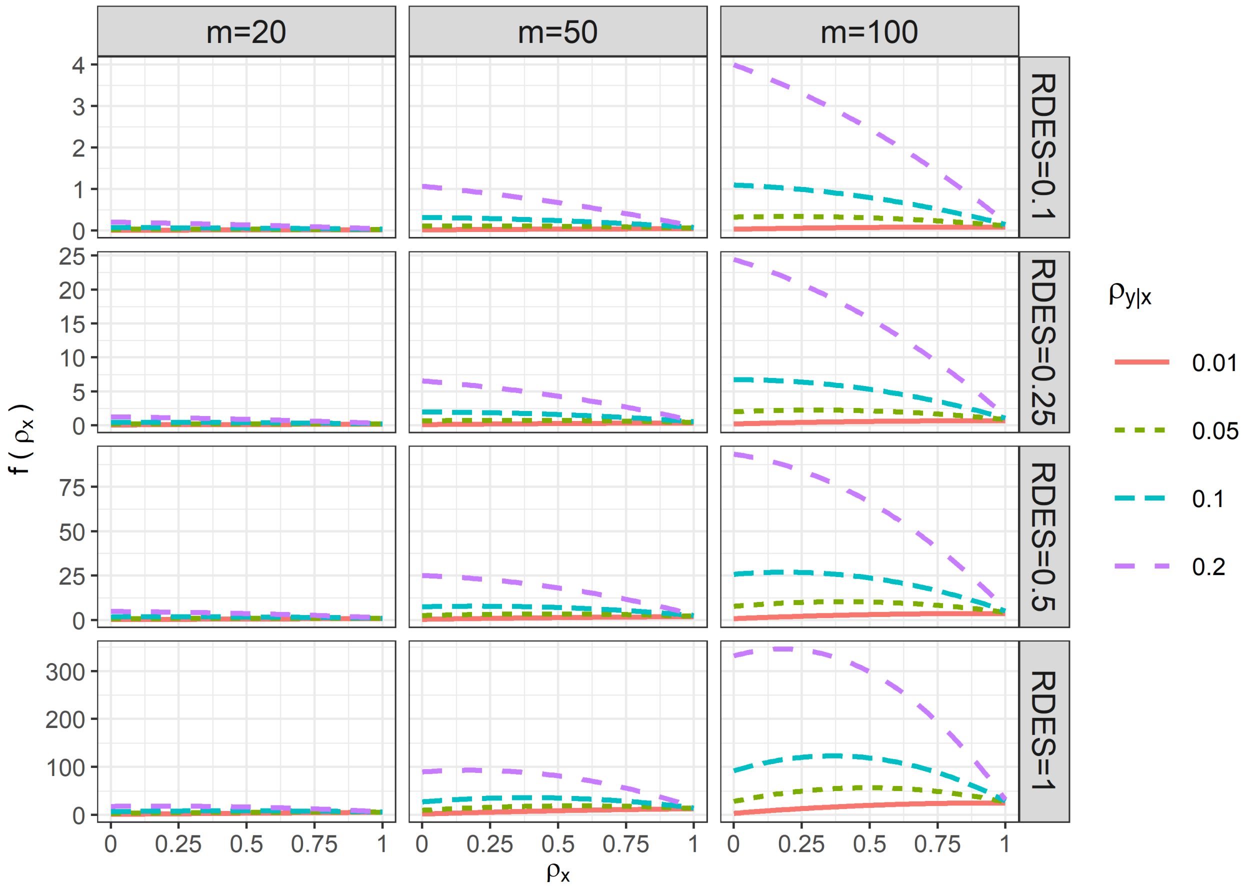
The line of symmetry is

$$\rho_x^* = \frac{\omega(m-1)\rho_{y|x}^2 + [1 - (1-\omega)(m-2)]\,\rho_{y|x} - (1-\omega)}{-2(1-\omega)(m-1)\rho_{y|x}}.$$

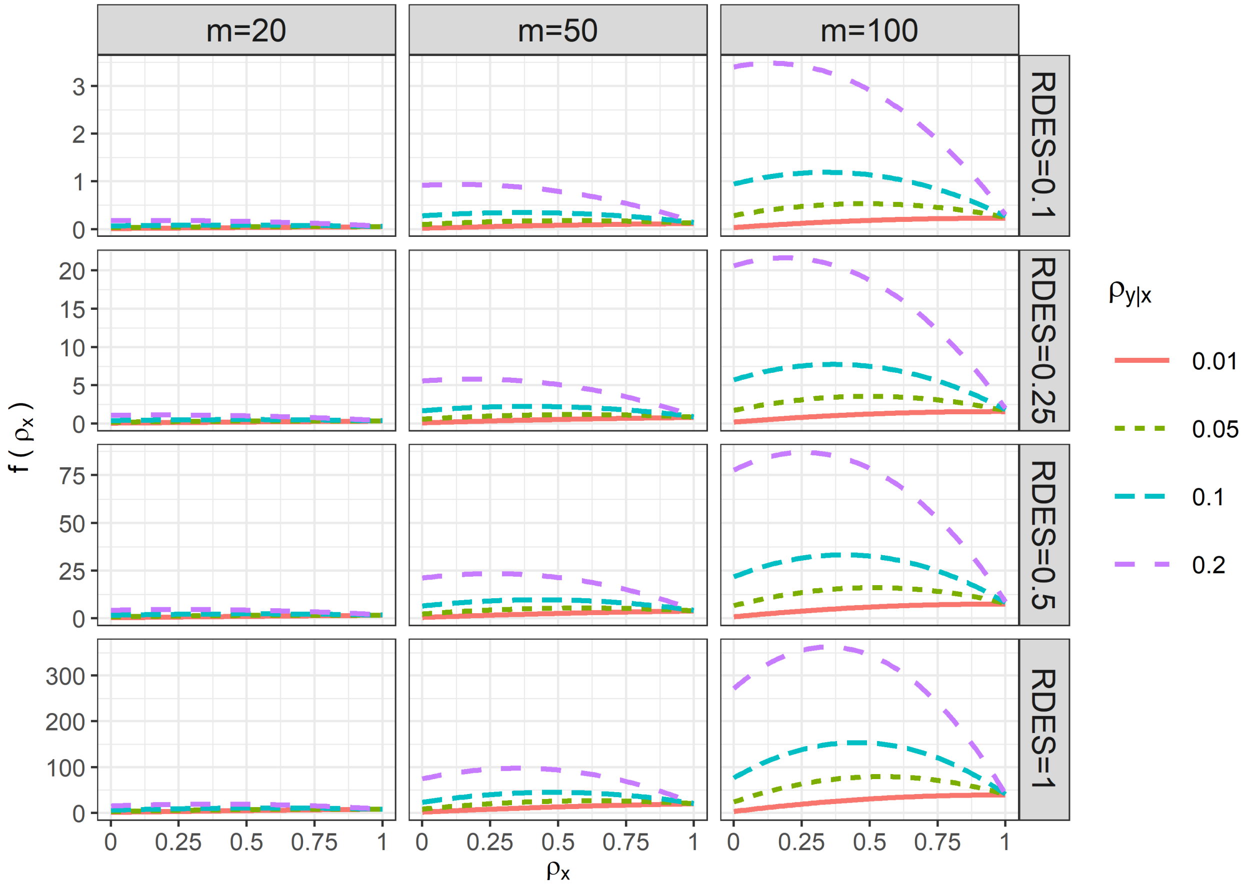
This means that the relationship between ρ_x and $f(\rho_x)$ over the support [0, 1], and hence the relationship between ρ_x and $\Theta(m)$, depends on the location of ρ_x^*. Specifically

- when $\rho_x^* < 0$, then $f(\rho_x)$ is monotonically decreasing in ρ_x; equivalently, $\Theta(m)$ will be monotonically increasing in ρ_x;
- when $\rho_x^* \in [0, 1)$, then $f(\rho_x)$ first increases and then decreases as ρ_x increases; equivalently, $\Theta(m)$ first decreases and then increases as ρ_x increases;
- when $\rho_x^* \geq 1$, then $f(\rho_x)$ is monotonically increasing in ρ_x; equivalently, $\Theta(m)$ will be monotonically decreasing in ρ_x.

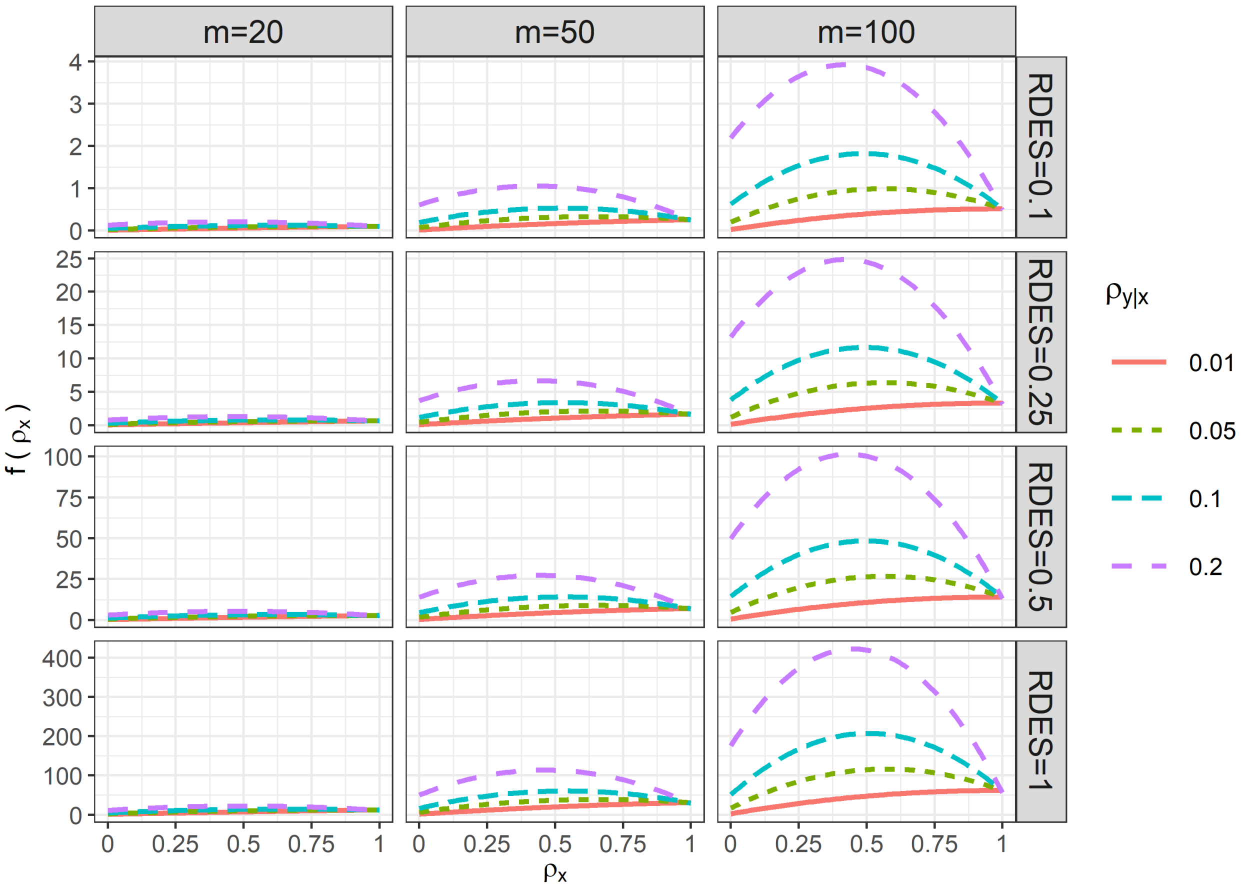
We visualize the function of $f(\rho_x)$ under various settings in Web Figure 1-3 to facilitate the understanding of relationship between $\Theta(m)$ and ρ_x. The parameter values are chosen to represent scenarios we discussed in the main text in Section 4.2. When the cluster size is small ($m = 20$), $f(\rho_x)$ is relatively flat as ρ_x increases, explaining the insensitivity of $\Theta(m)$ to changes in ρ_x in this scenario. When the cluster size is moderate and large ($m = 50$, 100), $f(\rho_x)$ increases monotonically when $\rho_{y|x} = 0.01$ because the line of symmetry is much greater than 1. In contrast, when $\rho_{y|x} > 0.01$, the line of symmetry is likely between 0 and 1, so that $f(\rho_x)$ first increases and then decreases. Taken together, these patterns could explain the pattern we observe in Figure 2 in the main text.



Web Figure 1 The graph of of $f(\rho_x)$ under various cluster size m, adjusted outcome ICC $\rho_{y|x}$, and ratio of detectable effect sizes (RDES), assuming $\sigma_x^2 = \sigma_{y|x}^2 = 1$, $\beta_2 = 0.5$, $\beta_3 = 0.25$.



Web Figure 2 The graph of of $f(\rho_x)$ under various cluster size m, adjusted outcome ICC $\rho_{y|x}$, and ratio of detectable effect sizes (RDES), assuming $\sigma_x^2 = \sigma_{y|x}^2 = 1$, $\beta_2 = 0.5$, $\beta_3 = 0.5$.



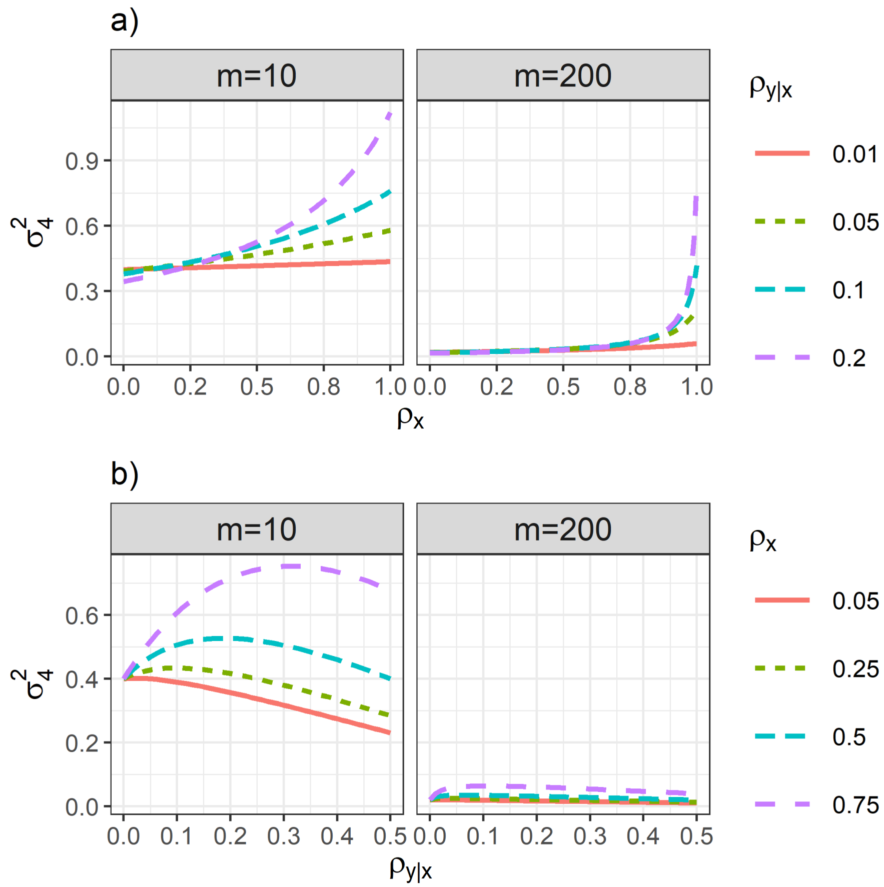
Web Figure 3 The graph of of $f(\rho_x)$ under various cluster size m, adjusted outcome ICC $\rho_{y|x}$, and ratio of detectable effect sizes (RDES), assuming $\sigma_x^2 = \sigma_{y|x}^2 = 1$, $\beta_2 = 0.5$, $\beta_3 = 1$.

C. R FUNCTION FOR POWER CALCULATION

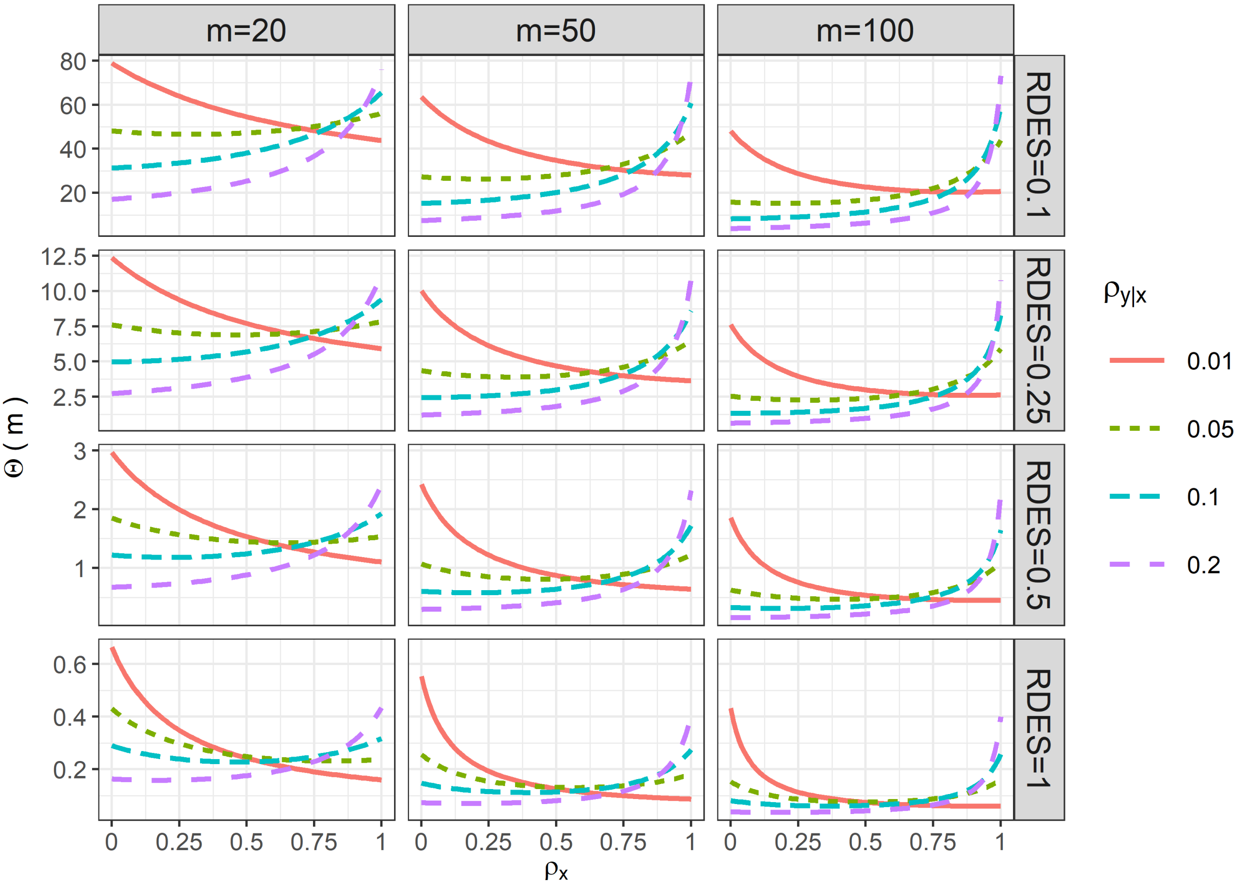
Below we provide simple R functions for implementing the sample size calculation procedure proposed in the main text, with a single covariate.

```
sample_size_cal<- function(eff, rhox, rhoy, varx, vary, beta=0.2, alpha=0.05, m, p=0.5){
  # Input:
  # eff - effect size of the interaction
  # rhox - covariate ICC
  # rhoy - adjusted outcome ICC
  # varx - marginal variance of covariate
  # vary - adjusted outcome variance components
  # beta -type II error (default 0.2)
  # alpha - type I error (default 0.05)
  # m - cluster size
  # p - proportion of treated (default 0.5)

  kappa = 1/(1-rhoy)*(1 + (m-2)*rhoy - (m-1)*rhox*rhoy)/(1+(m-1)*rhoy)
  n = (qnorm(1-alpha/2) + qnorm(1-beta))^2*vary/(eff^2*m*p*(1-p)*kappa*varx)
  return (n)
}


power_cal<- function(eff, rhox, rhoy, varx, vary, k, alpha=0.05, m, p=0.5){
  # Input:
  # eff - effect size of the interaction
  # rhox - covariate ICC
  # rhoy - adjusted outcome ICC
  # varx - marginal variance of covariate
  # vary - adjusted outcome variance components
  # k -number of clusters
  # alpha - type I error (default 0.05)
  # m - cluster size
  # p - proportion of treated (default 0.5)

  kappa = 1/(1-rhoy)*(1 + (m-2)*rhoy - (m-1)*rhox*rhoy)/(1+(m-1)*rhoy)
  power =pnorm( sqrt(k*eff^2*m*p*(1-p)*kappa*varx/vary)-qnorm(1-alpha/2))
  return (power)
}
```

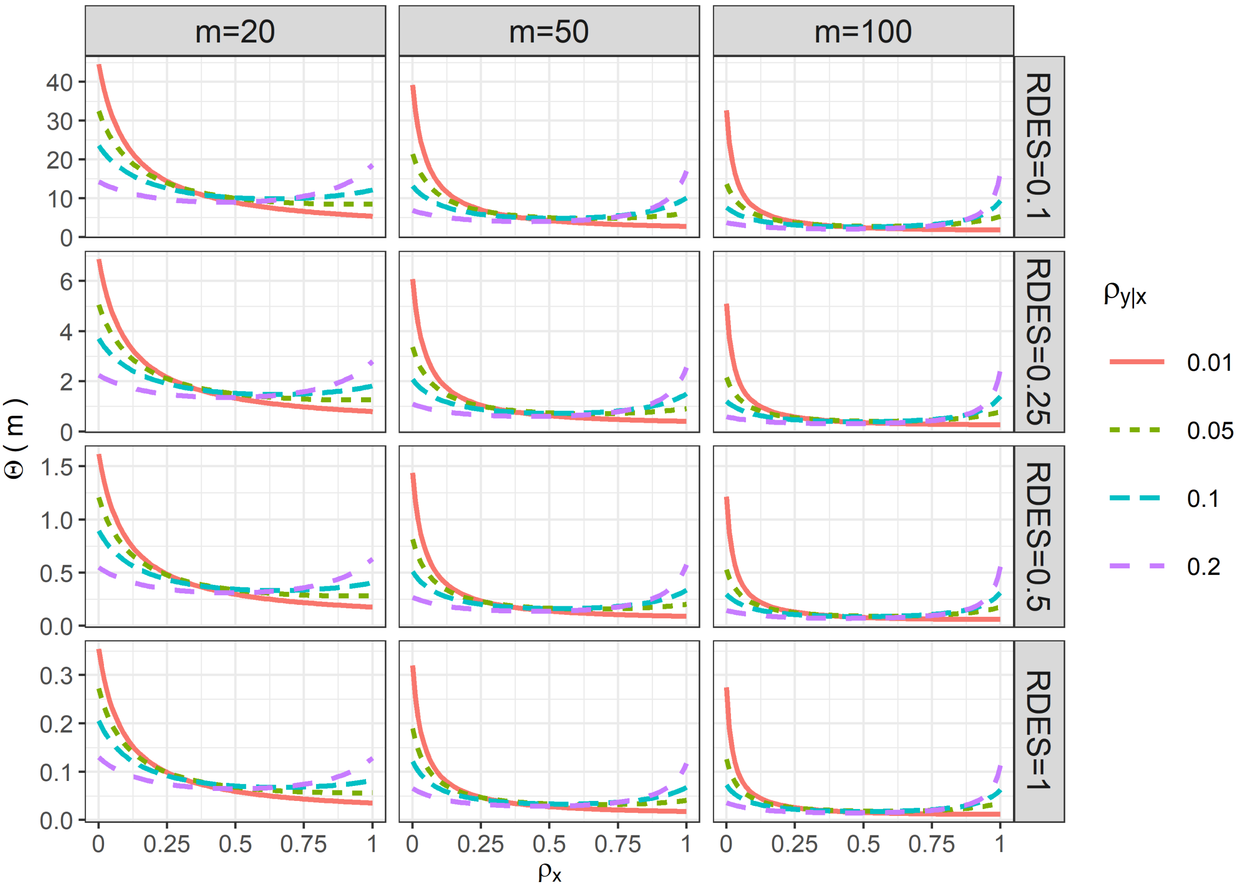
D. ADDITIONAL WEB FIGURES



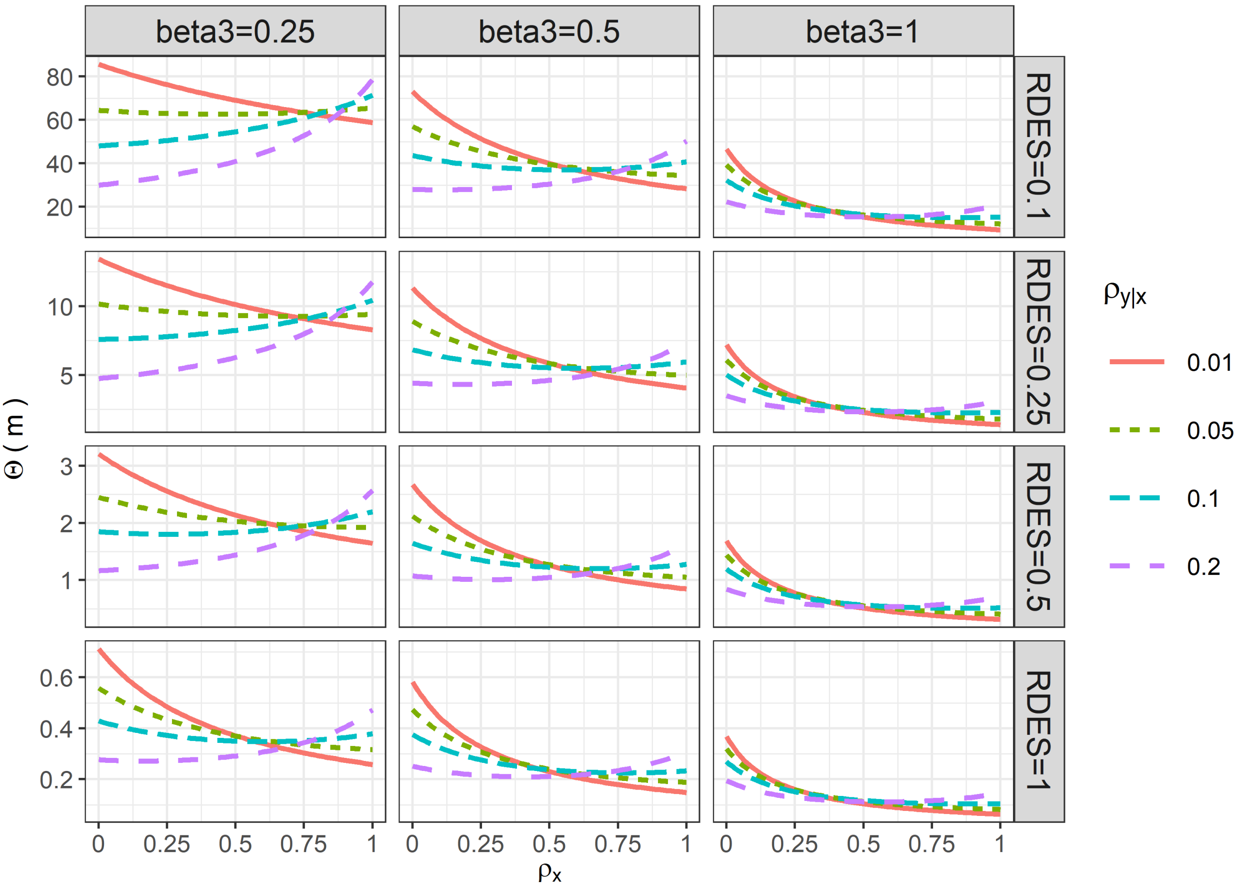
Web Figure 4 Variance of the GLS estimator for the treatment-by-covariate interaction, σ_4^2, as a function of the a) covariate ICC ρ_x, and b) adjusted outcome ICC $\rho_{y|x}$ with cluster sizes $m \in \{10, 200\}$, assuming $\sigma_{y|x}^2 = \sigma_x^2 = 1$, and $\sigma_w^2 = 1/4$.



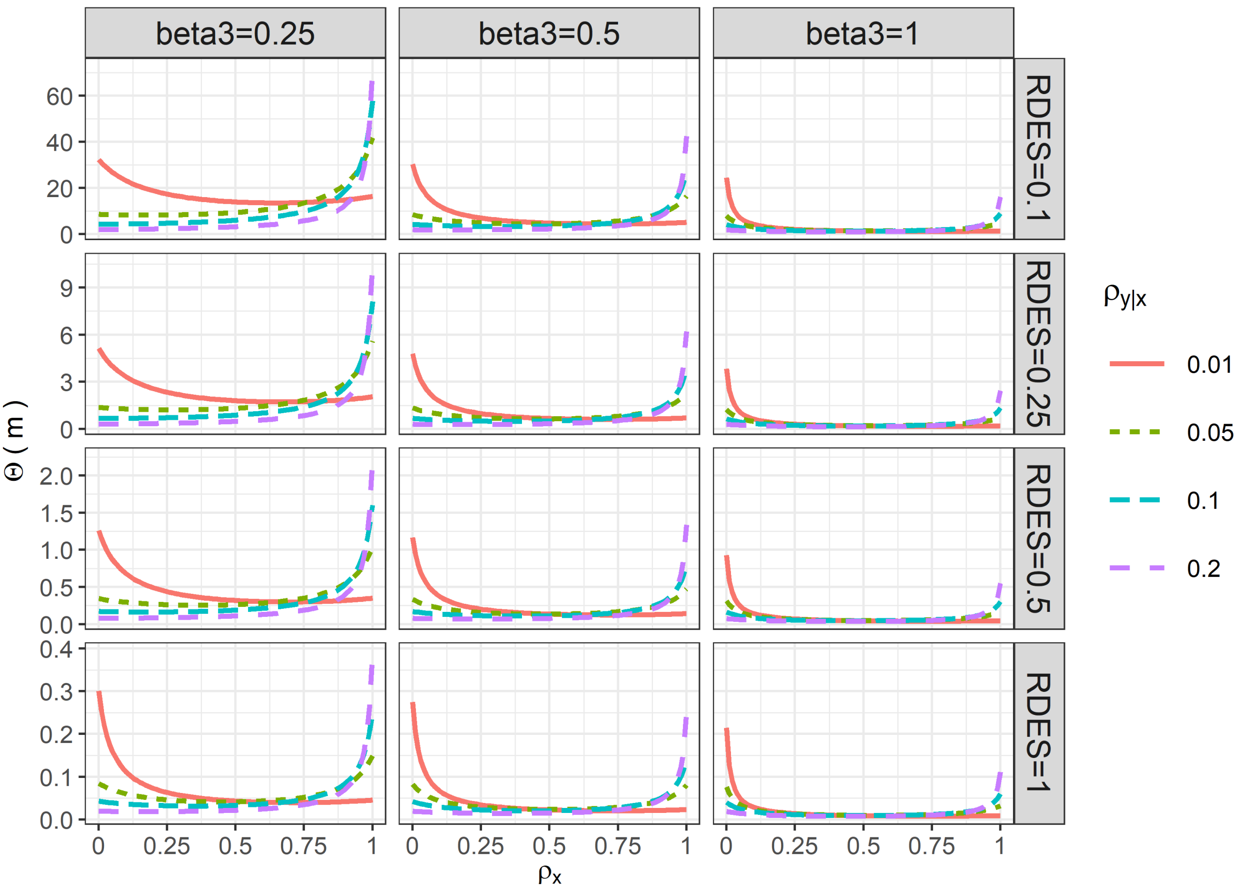
Web Figure 5 Ratio of total sample size required for testing HTE versus OTE as a function of the cluster size m, covariate ICC ρ_x, adjusted outcome ICC $\rho_{y|x}$, and ratio of detectable effect sizes (RDES), assuming $\sigma_x^2 = \sigma_{y|x}^2 = 1$, $\beta_2 = 0.5$, $\beta_3 = 0.25$, and $\bar{W} = 1/2$.



Web Figure 6 Ratio of total sample size required for testing HTE versus OTE as a function of the cluster size m, covariate ICC ρ_x, adjusted outcome ICC $\rho_{y|x}$, and ratio of detectable effect sizes (RDES), assuming $\sigma_x^2 = \sigma_{y|x}^2 = 1$, $\beta_2 = 0.5$, $\beta_3 = 1$, and $\bar{W} = 1/2$.



Web Figure 7 Ratio of total sample size required for testing HTE versus OTE as a function of covariate main effect β_3, covariate ICC ρ_x, adjusted outcome ICC $\rho_{y|x}$, and ratio of detectable effect sizes (RDES), assuming an extremely small cluster size $m = 10$, $\sigma_x^2 = \sigma_{y|x}^2 = 1$, $\beta_2 = 0.5$, and $\bar{W} = 1/2$.



Web Figure 8 Ratio of total sample size required for testing HTE versus OTE as a function of covariate main effect β_3, covariate ICC ρ_x, adjusted outcome ICC $\rho_{y|x}$, and ratio of detectable effect sizes (RDES), assuming a large cluster size $m = 200$, $\sigma_x^2 = \sigma_{y|x}^2 = 1$, $\beta_2 = 0.5$, and $\bar{W} = 1/2$.

References

1. Yang L, Tsiatis AA. Efficiency study of estimators for a treatment effect in a pretest–posttest trial. *The American Statistician* 2001; 55(4): 314–321.

2. Raudenbush SW. Statistical analysis and optimal design for cluster randomized trials.. *Psychological Methods* 1997; 2(2): 173–185.

3. Yang S, Starks MA, Hernandez AF, et al. Impact of baseline covariate imbalance on bias in treatment effect estimation in cluster randomized trials: Race as an example. *Contemporary Clinical Trials* 2019.

4. Li F, Lokhnygina Y, Murray DM, Heagerty PJ, DeLong ER. An evaluation of constrained randomization for the design and analysis of group-randomized trials. *Statistics in Medicine* 2016; 35(10): 1565–1579.

□