Data-Generating Model for Simulations in Partial EIC Simulation Experiment

December 8, 2015

In these simulations, the pretest is included as a covariate. We assume the pretest P_{ij} is centered and scaled so that $P_{ij} \sim N(0,1)$ and that the posttest Y_{ij} is centered and scaled so that, after adjusting for treatment and cluster effects, it also has a mean of zero and variance of 1. We assume a partial nesting design where individuals are clustered ($C = 1$) if and only if $X_1 = +1$, and unclustered ($C = 0$) if and only if $X_i = -1$. Cluster members have a random cluster effect. In the notation, we treat everyone as clustered (even if the individual is the only member of his or her own cluster), but people who are clustered in this trivial sense do not get a random cluster effect. Thus, for individual i in "cluster" j,

$$Y_{ij} = u_j C_j + \frac{d}{2} x_{1j} + \frac{d}{2} x_{3j} + \frac{d}{2} x_{1j} x_{3j} + \gamma_p P_{ij} + e_{ij}$$

where $u_j \sim N(0, \tau_u^2)$ and $e_{ij} \sim N(0, \sigma_e^2)$.

Equal Variance Scenario

In scenarios where clustered and nonclustered individuals have **equal error variances** σ_e^2 after adjusting for cluster, we assume this error variance is $\sigma_e^2 = 1 - \gamma_p^2$. Thus, the overall posttest variance for unclustered members is $Var(Y_{ij}|\mathbf{x}_j, u_j, C_j = 0) = \gamma_p^2 Var(P_{ij}) + Var(e_{ij}) = \gamma_p^2 + 1 - \gamma_p^2 = 1$. The overall posttest variance for clustered members within a cluster is also $Var(Y_{ij}|\mathbf{x}_j, u_j, C_j = 1) = \gamma_p^2 Var(P_{ij}) + Var(e_{ij}) = \gamma_p^2 + 1 - \gamma_p^2 = 1$. The overall posttest variance for clustered members if cluster label is ignored is slightly higher, namely $Var(Y_{ij}|\mathbf{x}_j, C_j = 1) = 1 + \tau_u^2$. The pretest-posttest correlation for unclustered members, or for clustered members within a cluster, is

$$\frac{Cov(P_{ij}, Y_{ij}|u_j)}{\sqrt{Var(P_{ij}|u_j) Var(Y_{ij}|u_j)}} = \frac{\gamma_p}{1} = \gamma_p,$$

and we set $\gamma_p = .65$ for a pretest-posttest correlation of .65 because this value is considered reasonable in the literature. Thus, we set $\sigma_e^2 = 1 - .65^2 = 0.5775$.

Because the posttest variance inside a cluster is $1+\tau_u^2$, the within-cluster correlation for clustered members inside a cluster is

$$\frac{Cov(Y_{ij}, Y_{i'j}|u_j)}{Var(Y_{ij}|u_j)} = \frac{\tau_u^2}{1+\tau_u^2}.$$

We assume a posttest ICC of .15, so we must set $\tau_u^2 = \frac{.15}{1-.15} \approx 0.17647$.

Unequal Variance Scenario

In scenarios where clustered and nonclustered individuals have **unequal error variances** σ_e^2, i.e., cluster members have smaller posttest error variances, we assume that the error variance for clustered individuals is $\sigma_{e1}^2 = \frac{2}{3}\left(1-\gamma_p^2\right)$ and the error variance for unclustered individuals is $\sigma_{e0}^2 = \frac{4}{3}\left(1-\gamma_p^2\right)$. This means that even without renormalizing the outcomes to get rid of the contribution of τ_u^2, members of a cluster are more similar to each other than unclustered individuals are to each other. Specifically, continuing to use $\gamma_p = .65$, we get $\sigma_{e1}^2 = 0.385$ and $\sigma_{e0}^2 = 0.770$, so that $\sigma_{e1}^2 = \frac{1}{2}\sigma_{e0}^2$.

The person-level pretest-posttest correlation is now

$$\frac{Cov(P_{ij}, Y_{ij}|u_j)}{\sqrt{Var(P_{ij}|u_j)Var(Y_{ij}|u_j)}} = \frac{\gamma_p}{1\sqrt{\gamma_p^2 + \frac{4}{3}\left(1-\gamma_p^2\right)}} = \frac{\gamma_p}{\sqrt{\frac{4}{3} - \frac{1}{3}\gamma_p^2}}$$

for unclustered individuals, i.e., about .595 if $\gamma_p = .65$. It is

$$\frac{Cov(P_{ij}, Y_{ij}|u_j)}{\sqrt{Var(P_{ij}|u_j)Var(Y_{ij}|u_j)}} = \frac{\gamma_p}{1\sqrt{\gamma_p^2 + \frac{2}{3}\left(1-\gamma_p^2\right)}} = \frac{\gamma_p}{\sqrt{\frac{2}{3} + \frac{1}{3}\gamma_p^2}}$$

for clustered individuals, i.e., about .723 if $\gamma_p = .65$.

For clustered individuals, the intraclass correlation is now

$$\frac{Cov(Y_{ij}, Y_{i'j}|u_j)}{Var(Y_{ij}|u_j)} = \frac{\tau_u^2}{\sqrt{\gamma_p^2 + \frac{2}{3}\left(1-\gamma_p^2\right) + \tau_u^2}}$$

If we keep the same $\gamma_p = .65$ and $\tau_u^2 = \frac{.15}{1-.15} \approx 0.17647$ then we have a new ICC of about 0.1779.

Variance of Regression Coefficients for Treatment Components

For purposes of predicting power, we need to calculate $Var(\gamma_1)$ which is the sampling variance of the regression coefficient for x_1. (Somewhat surprisingly, the regression coefficients for the other components and interactions have the same variance as that of x_1.)

The suggested formula is

$$Var(\gamma_1) = \frac{\tau_u^2}{4J_1} + \frac{\sigma_{e1}^2}{4J_1 n} + \frac{\sigma_{e0}^2}{4J_0}$$

In the simulation we assume 20% dropout so we replace n with $0.8n$.

As mentioned above, in the equal-error-variance scenario, with unclustered posttest variance set to 1, we have $\tau_u^2 = \rho/(1-\rho)$ and $\sigma_{e0}^2 = \sigma_{e1}^2 = 1 - \gamma_p^2$ where ρ is the posttest ICC and γ_p is the pretest-posttest correlation. In the unequal-variance scenario, these simple relationships no longer hold.