## Abstract Title Page

*Not included in page count.*

**Title:**

The Statistical Power of the Cluster Randomized Block Design with Matched Pairs – a Simulation Study

**Author(s):**

Nianbo Dong, Ph.D. and Mark Lipsey, Ph.D.
*Vanderbilt University*

## Abstract Body

*Limit 5 pages single spaced.*

**Background/context:**

Group-randomized experiments, in which groups (clusters) are randomly assigned to the treatment and control conditions, are widely used in educational research (e.g., Borman, Slavin, Cheung, Chamberlain, and Madden, et al., 2005; Cook, Hunt, & Murphy, 2000). In designing such studies, it is important to have sufficient power to detect the effect of the intervention. Raudenbush, Martinez, & Spybrook (2007) summarized five factors associated with the power in group-randomized experiments: (1) the size of the true average effect size (ES) of the intervention, (2) the intra-class correlation (ICC) indicating the fraction of the variance that lies between clusters, (3) the number of clusters studied (*J*), (4) the number of individuals per cluster (*n*), and (5) the adopted level of statistical significance ( $\alpha$ ) for either a one-tailed or two-tailed test.

In addition, using covariance adjustment can improve precision (Bloom, 2006). The bigger the $R^2$ for the relationship between a covariate and the dependent variable, the smaller the minimum detectable effect size (MDES) is. This is a commonly used approach in group-randomized studies (e.g., Borman et al, 2005).

Furthermore, pre-randomization blocking or matching is another important approach that can control background variability and increase power for detecting the effects of treatment (Hedges, 2007; Raudenbush, Martinez, & Spybrook, 2007). One application of this approach is the randomized matched-pairs design, one variant of the randomized block design, i.e., first matching clusters (e.g., schools) into pairs on a covariate, then randomly assigning the clusters in each pair to two conditions. Matching can improve face validity, and improve precision in some circumstances (Bloom, 2007; Raudenbush, Martinez, & Spybrook, 2007). In addition to the factors listed above, two additional factors are also associated with the power in the matched-pairs (MP) randomized-block design, i.e., the within-pair correlation and the true effect size variability (ESV) across pairs (Raudenbush, Martinez, & Spybrook, 2007).

Pairing will improve statistical power when "the gain in predictive power outweighs the loss of degrees of freedom" associated with the blocked pairs (Bloom, 2007). Bloom (2007) derived a formula to calculate the minimum incremental $R^2$ that will improve power, and illustrated the required $R^2$ for varying numbers of clusters. Furthermore, Raudenbush, Martinez, & Spybrook (2007) compared power for the completely randomized cluster design *with* and *without* covariance adjustment, and the matched-pair (MP) randomized-block design for different parameter combinations. Their conclusions are consistent with Bloom (2007) that matching will not improve power unless the percentage of the variance explained by matching reaches a threshold.

These papers provided a good theoretical framework and mathematical formulation for determining the threshold above which matching will improve power. However, as Raudenbush, Martinez, & Spybrook (2007) noted "there is no closed-form mathematical expression for determining the variance explained by matching on *W* when *W* randomly varies within pairs" (p.21). In practice it is unclear how large the incremental $R^2$ of matching can be under various situations. Furthermore, power was not directly compared between the models with random and fixed pair effects *with* covariance adjustment and the model for the group randomized design *with* covariance adjustment.

**Purpose / objective / research question / focus of study:**

This study uses simulation techniques to examine the statistical power of the group-randomized design and the matched-pair (MP) randomized block design under various parameter combinations. Both nearest neighbor matching and random matching are used for the MP design. The power of each design for any parameter combination was calculated from 1,000 simulated datasets. The random pair effect HLM model *with* and *without* covariance adjustment, and the fixed pair effect HLM model *with* and *without* covariance adjustment were used to analyze the data from the matched-pair design, and the HLM with covariance adjustment was used to analyze the data from the completely randomized cluster design.

**Setting:**

Simulated setting, in which the individuals are nested within clusters, e.g., students are nested within schools.

**Population / Participants / Subjects:**

Samples were generated for the simulation using statistical models with various combinations of parameters.

**Intervention / Program / Practice:**

Simulated interventions with various effect sizes and effect size variability across pairs/.

**Research Design:**

Data were generated using various parameters: (1) the number of clusters ($J$ = 20, 40, and 60), (2) the number of individuals per cluster ($n$ = 20), (3) the intra-class correlation (ICC = .1 and .2), (4) $R_1^2$ (= .2 and .5) for individual level covariate, (5) $R_2^2$ (= $R_1^2$ = .2 and .5) for cluster level covariate, (6) the effect size (ES = .2 and .5), and (7) the effect size variability (ESV = 0, and .01) across pairs. The number of total parameter combinations is 48.

Both nearest neighbor matching and random matching were used for the matched-pair (MP) group-randomized design. The nearest neighbor matching procedure rank orders clusters by a cluster-level covariate ($W$) and pairs adjoining clusters. In random matching, clusters are paired randomly. The results from random matching provide information about how much the quality of matching matters.

Treatment effects were estimated for the group-randomized design using a two-level HLM model with covariance adjustment. They were estimated for the nearest neighbor and random matching using four models: *random* pair effects *with* and *without* covariance adjustment, and *fixed* pair effects *with* and *without* covariance adjustment. The $p$-value, the effect size, the $R_1^2$ for the individual level covariate, $R_2^2$ for the cluster level covariate, the unconditional and conditional ICC etc. were calculated for each model with each simulated data set.

For each of 48 parameter combinations, 1,000 datasets were generated and analyzed using SAS. The percentage of the studies whose *p*-value were less than .05, two-tailed, estimated the power. The average ES, $R_1^2$, $R_2^2$, and ICC etc. are calculated.

**Data Collection and Analysis:**

*Data Generation*

Using the parameters specified above, data were generated following two stages below.

1. Generating a master dataset satisfying the models in Appendix C.

The outcome variable, $Y_{ij}$, was created satisfying; (1) the specified unconditional ICC, which was satisfied by varying level-1 variance and level-2 variance, (2) an individual level covariate, *X*, and a cluster-level covariate, *W*, which were normally distributed, and their variance explained at level 1 and level 2 were $R_1^2$ and $R_2^2$, respectively. The overall pooled standard deviation ($\sqrt{\tau_0 + \sigma_0^2}$) of $Y_{ij}$ was calculated.

2. Adding treatment effect on $Y_{ij}$.

Using the same master dataset generated above, clusters were randomly assigned to two conditions, clusters in the nearest neighbor matched pairs were randomly assigned to two conditions, and clusters in the randomly matched pairs were randomly assigned to two conditions to produce three study designs. The same treatment effect, $(ES + v) * \sqrt{\tau_0 + \sigma_0^2}$, where $v \sim N(0, \sqrt{ESV})$ was a random term indicating that the treatment effect varied across pairs, was added to the $Y_{ij}$ for the treatment group to produce final dependent variable, $Y_{ij}^{'}$.

*Data Analysis*

A two-level HLM model with covariance adjustment was used to estimate the treatment effect for the data from completely group-randomized design.

Level 1 (individual)

(1) $$Y_{ij}^{'} = \beta_{0j} + \beta_{1j} C_{ij} + r_{ij} \qquad r_{ij} \sim N(0, \sigma^2)$$

Level 2 (cluster)

(2) $$\beta_{0j} = \gamma_{00} + \gamma_{01} W_j + \gamma_{02} TREATMENT_j + \mu_0 \qquad \mu_0 \sim N(0, \tau)$$

(3) $$\beta_{1j} = \gamma_{10}$$

The reduced model is:

(4) $$Y_{ij}^{'} = \gamma_{00} + \gamma_{01} W_j + \gamma_{02} TREATMENT_j + \gamma_{10} X_{ij} + \mu_0 + r_{ij}$$

As mentioned in Research Design, four models were used to estimate the treatment effect for the matched-pair group-randomized design. The *random* pair effect HLM model *with* covariance adjustment is presented below:

(5)
$$Y'_{ijk} = \gamma_{000} + \gamma_{010}W_{jk} + \gamma_{020}TREATMENT_{jk} + \gamma_{100}\ X_{ijk}$$
$$+\mu_{00k} + \mu_{0jk} + \mu_{02k}TREATMENT_{jk} + r_{ijk}$$
$$r_{ijk} \sim N(0,\sigma^2)\,,\ \mu_{0jk} \sim N(0,\tau_{00}^{(2)})\,,\ \begin{pmatrix}\mu_{00k}\\ \mu_{02k}\end{pmatrix} \sim N\begin{pmatrix}\tau_{00}^{(3)} & \\ \tau_{10}^{(3)} & \tau_{11}^{(3)}\end{pmatrix}$$

Where $k$ is the third level, pair.

The *fixed* pair effect HLM model *with* covariance adjustment is presented below:

(6)
$$Y'_{ijk} = \gamma_{000} + \gamma_{010}W_{jk} + \gamma_{020}TREATMENT_{jk} + \gamma_{100}\ X_{ijk} + \sum_m \gamma_{00\ m}P_n$$
$$+\mu_{0jk} + r_{ijk}$$
$$r_{ijk} \sim N(0,\sigma^2)\,, \mu_{0jk} \sim N(0,\tau_{00}^{(2)})$$

Where $k$ is the third level, pair. $P_m$ is a dummy variable indicating pair $m$.

The *random* and *fixed* pair effect HLM models *without* covariance adjustment are similar with Models (5) and (6), respectively except that no covariate variables (*X* and *W*) are included.

**Findings / Results:**

Partial results are reported below although the overall simulation and analyses have not been completed. Table 1 presents the simulation results (with 1,000 replications) for the parameter combination: $J$ = 40, $n$ = 20, ICC = .1, ES = .3, $R_1^2$ = $R_2^2$ = .5, and ESV = 0 and .01, respectively. First, notice that the power of the group-randomized design under this parameter combination for ESV = 0 is .93. The Optimal Design software (Liu, Spybrook, Congdon, Martinez, & Raudenbush, 2006) produced a power of .84 for the same parameter combination but did not take account of $R_1^2$ (Figure 1). Based on Bloom's (2006, p. 17) Minimum Detectable Effect Size (MDES) formula, the power associated with this parameter combination is .93, and the power ignoring $R_1^2$ is .85. Hence this simulation result is consistent with the results derived from Bloom's mathematical formulation. Second, the group-randomized design had the largest power, but the MP randomized block design, when analyzed *with* covariance-adjustment, had almost identical power. In addition, there was no difference between the random and fixed pair effects MP models or between the two matching procedures for models *with* covariance-adjustment. However, for the models *without* covariance-adjustment, nearest neighbor matching had more power than random matching (.81 vs. .72). Third, for the MP randomized block design, the analytic models *with* covariance-adjustment had more power than those *without* covariance-adjustment. Fourth, the power with ESV (effect size variability) = 0 was slightly higher than the power with ESV = .01 when the models *with* covariance-adjustment were used, but did not vary much for the models *without* covariance adjustment.

Table 2 presents the simulation results with the same parameter combination in Table 1 except with $R_1^2$ = $R_2^2$ = .2. The results have the same pattern as in Table 1 but power is lower than in Table 1.

Table 3 presents the simulation results for $J$ = 20, $n$ = 20, ICC = .1, ES = .3, $R_1^2 = R_2^2 =$ .2 and 5, and ESV = 0 and .01, respectively for the models *with* covariance-adjustment. The group-randomized design still had the largest power. The random pair effects models produced similar power, but the fixed pair effects models produce slightly smaller power.

Table 4 presents the average estimated $R^2$ of various study designs. First, notice that the simulated data were as designed in terms of $R^2$. Next, *with* covariance-adjustment for ESV = .1, $R_2^2$ from the random pair effects model was a little bigger than the other models. However, the overall $R^2$ did not vary across study designs and analytic models when covariance was adjusted.

**Conclusions:**

Under the parameter combination illustrated, the simulation showed that the group-randomized design had the greatest power. By including the covariate in the model, the power was improved, and the quality of the matching had only a small effect on power. However, when $R^2$ was big, but the covariate was omitted, the nearest neighbor matching had more power than random matching. In addition, when the sample size was small, the random pair effects model produced bigger power than the fixed pair model, which seems to be against the current theory and needs further exploration. Finally, when covariance was adjusted, the incremental $R^2$ of matching was limited.

## Appendices

*Not included in page count.*

### Appendix A. References

*References are to be in APA version 6 format.*


Bloom, H. S. (2006). The Core Analytics of Randomized Experiments for Social Research. MDRC Working Papers on Research Methodology. Available online at: http://www.mdrc.org/ publications/437/full.pdf

Bloom, H. S. (2007). *Sample Design for Group-Randomized Trials*. Lecture in the IES Summer Research Training Institute, Vanderbilt University.

Cook, T. D., Hunt, H. D., & Murphy, R. F. (2000). Comer's School Development Program in Chicago: A theory-based evaluation. *American Educational Research Journal*, *37*, 535–597.

Hedges, L. (2007). *Basic Experimental Design*. Lecture in the IES Summer Research Training Institute, Vanderbilt University.

Liu, X, Spybrook, J., Congdon, R., Martinez, A., & Raudenbush, S. (2006). Optimal Design for Multi-level and Longitudinal Research. (Version 1.77): HLM Software.

Raudenbush, S. W., Martinez, A., & Spybrook, J. (2007). Strategies for Improving Precision in Group-Randomized Experiments. *Educational Evaluation and Policy Analysis. 29 (1*): 5-29.

## Appendix B. Tables and Figures

*Not included in page count.*

**Table 1**

*Powers of various study designs from simulation*[a] *(J=40, n=20, ICC=.1, ES=.3,* $R_1^2 = R_2^2 = .5$*,* $\alpha = .05$ *for two-tailed test)*

| Study Design | Analytic Model | *With* covariance-adjustment | | *Without* covariance-adjustment | |
|---|---|---|---|---|---|
| | | ESV[b] | | ESV[b] | |
| | | 0 | .01 | 0 | .01 |
| Completely group-randomized design | 2-level HLM | .93[c] | .90 | | |
| Matched-pair Randomized Block Design (Nearest neighbor matching) | Random Pair Effects Model | .91 | .89 | .81 | .81 |
| | Fixed Pair Effects Model | .90 | .88 | .81 | .81 |
| Matched-pair Randomized Block Design (Random matching) | Random Pair Effects Model | .91 | .88 | .72 | .72 |
| | Fixed Pair Effects Model | .90 | .87 | .72 | .71 |

[a]1,000 replications.

[b]ESV: Effect Size Variability across pair or cluster, represented by the variance of the Effect Size adopted.

[c]Under this parameter combination, the Optimal Design software (v.1.77) produced a power of .84, which is lower than .93 because Optimal Design did not take account of $R_1^2$. Based on the formula on page 17 in Bloom (2006), the power associated with this parameter combination is .93, and the power ignoring $R_1^2$ is .85.

**Table 2**

*Powers of various study designs from simulation*[a] *(J=40, n=20, ICC=.1, ES=.3,*

$R_1^2 = R_2^2 = .2,\ \alpha\ = .05$ *for two-tailed test)*

| Study Design | Analytic Model | *With* covariance-adjustment ESV[b] 0 | *With* covariance-adjustment ESV[b] .01 | *Without* covariance-adjustment ESV[b] 0 | *Without* covariance-adjustment ESV[b] .01 |
|---|---|---|---|---|---|
| Completely group-randomized design | 2-level HLM | .76 | .73 | | |
| Matched-pair Randomized Block Design (Nearest neighbor matching) | Random Pair Effects Model | .77 | .73 | .70 | .72 |
| | Fixed Pair Effects Model | .74 | .71 | .70 | .72 |
| Matched-pair Randomized Block Design (Random matching) | Random Pair Effects Model | .75 | .71 | .70 | .68 |
| | Fixed Pair Effects Model | .73 | .70 | .70 | .68 |

[a]1,000 replications.

[b]ESV: Effect Size Variability across pair or cluster, represented by the variance of the Effect Size adopted.

**Table 3**

*Powers of various study designs from simulation*[a] *(J=20, n=20, ICC=.1, ES=.3,* $\alpha$ *= .05 for two-tailed test)*

| Study Design | Analytic Model | $R_1^2 = R_2^2 = .2$ | | $R_1^2 = R_2^2 = .5$ | |
|---|---|---|---|---|---|
| | | ESV[b] | | ESV[b] | |
| | | 0 | .01 | 0 | .01 |
| Completely group-randomized design | 2-level HLM | .44 | .42 | .64 | .61 |
| Matched-pair Randomized Block Design (Nearest neighbor matching) | Random Pair Effects Model | .43 | .38 | .61 | .57 |
| | Fixed Pair Effects Model | .38 | .33 | .56 | .53 |
| Matched-pair Randomized Block Design (Random matching) | Random Pair Effects Model | .40 | .37 | .56 | .52 |
| | Fixed Pair Effects Model | .37 | .35 | .53 | .52 |

[a] 1,000 replications. All models are *with* covariance-adjustment.

[b] ESV: Effect Size Variability across pair or cluster, represented by the variance of the Effect Size adopted.

**Table 4**

*Average Estimated $R^2$ of various study designs from simulation*[a] *(J=40, n=20, ICC=.1, ES=.3,* $R_1^2 = R_2^2 = .5$*)*

| Study Design | Analytic Model | ESV[b]=0 | | | ESV[b]=.01 | | |
|---|---|---|---|---|---|---|---|
| | | Total $\hat{R}^2$ | $\hat{R}_2^2$ | $\hat{R}_1^2$ | Total $\hat{R}^2$ | $\hat{R}_2^2$ | $\hat{R}_1^2$ |
| Completely group-randomized design | 2-level HLM | .50 | .48 | .50 | .50 | .45 | .50 |
| Matched-pair Randomized Block Design (Nearest neighbor matching) | Random Pair Effects Model | .50 | *.49*[c] | .50 | .50 | *.51*[c] | .50 |
| | Fixed Pair Effects Model | .50 | .49 | .50 | .50 | .45 | .50 |
| Matched-pair Randomized Block Design (Random matching) | Random Pair Effects Model | .50 | *.48*[c] | .50 | .50 | *.50*[c] | .50 |
| | Fixed Pair Effects Model | .50 | .48 | .50 | .50 | .44 | .50 |

[a]1,000 replications. All models are *with* covariance-adjustment.

[b]ESV: Effect Size Variability across pair or cluster, represented by the variance of the Effect Size adopted.

[b]Representing the proportion of the sum variances at level 2 and level 3 that are explained by the covariate $W$. It was calculated for the purpose of comparison with the other models.

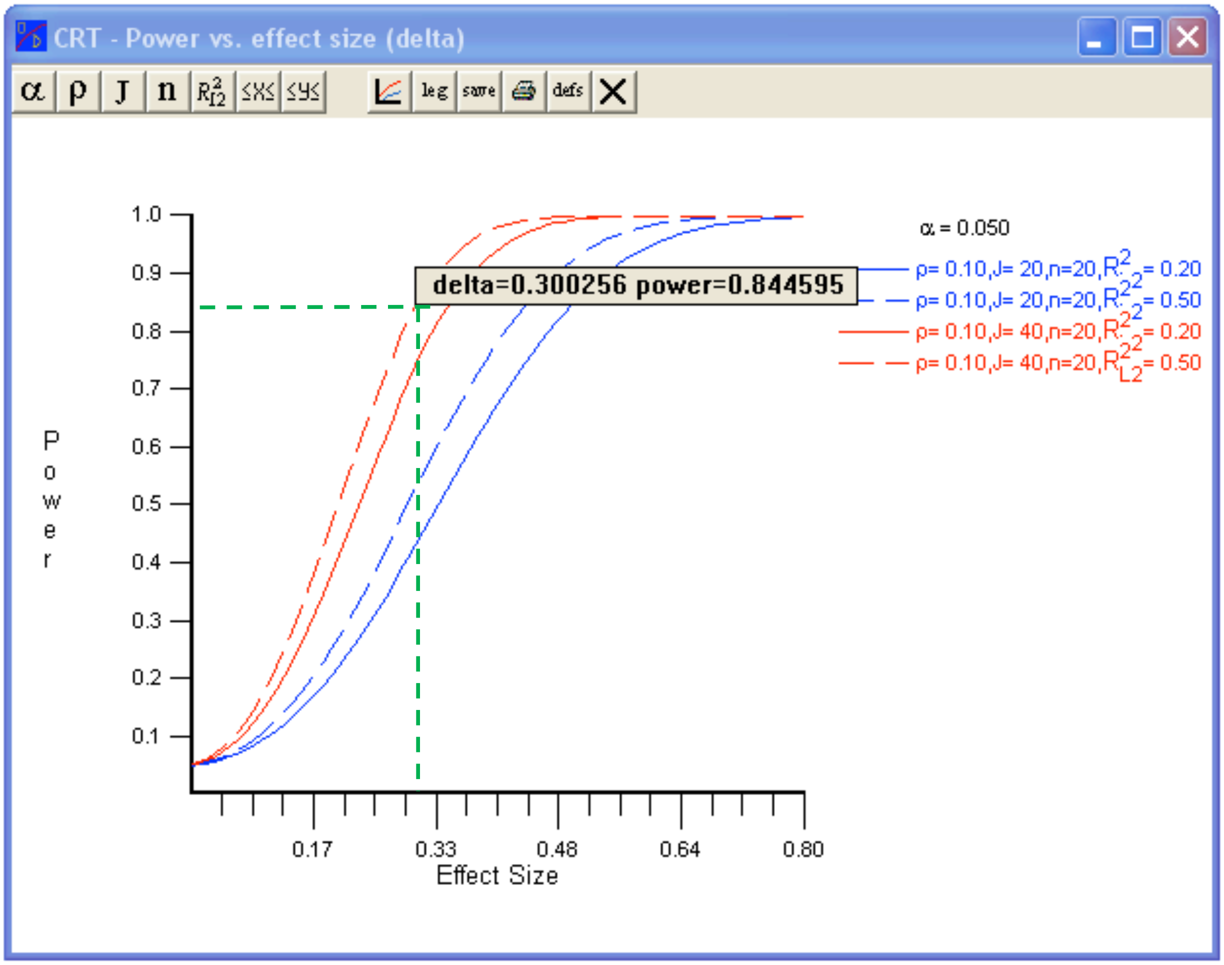


**Figure 1. Power vs. Effect Size by Optimal Design**

## Appendix C. Models Used to Generate Data at Stage 1

*Not included in page count.*

Unconditional Model:

(1) $Y_{ij} = \gamma_{00} + \mu_0 + r_{ij}$, $\mu_0 \sim N(0, \tau_0)$, $r_{ij} \sim N(0, \sigma_0^2)$

Where:

$Y_{ij}$ = Outcome for individual, *i* in cluster, *j*

$\gamma_{00}$ = Grand mean of the outcome

$\mu_0$ = Random error term for cluster, *j*

$r_{ij}$ = Random error term for individual, *i* in cluster, *j*

Unconditional ICC = $\dfrac{\tau_0}{\tau_0 + \sigma_0^2}$

Pooled standard deviation: $\sqrt{\tau_0 + \sigma_0^2}$

Conditional Model:

Level 1 (individual)

(2) $Y_{ij} = \beta_{0j} + \beta_{1j} X_{ij} + r_{ij}$ $r_{ij} \sim N(0, \sigma_1^2)$

Level 2 (cluster)

(3) $\beta_{0j} = \gamma_{00} + \gamma_{01} W_j + \mu_0$ $\mu_0 \sim N(0, \tau_1)$

(4) $\beta_{1j} = \gamma_{10}$

The reduced model is:

(5) $Y_{ij} = \gamma_{00} + \gamma_{01} W_j + \gamma_{10} X_{ij} + \mu_0 + r_{ij}$

Where:

$Y_{ij}$ = Outcome for individual, *i* in cluster, *j*

$X_{ij}$ = Covariate for individual, *i* in cluster, *j* (normally distributed)

$W_j$ = Cluster-level covariate for cluster, *j* (normally distributed)

$\gamma_{00}$ = Grand mean of the outcome

$\gamma_{01}$ = Coefficient of $X_j$

$\gamma_{10}$ = Coefficient of $C_{ij}$

$\mu_0$ = Random error term for cluster, *j*

$r_{ij}$ = Random error term for individual, *i* in cluster, *j*

Individual level covariate $R_1^2 = \dfrac{\sigma_0^2 - \sigma_1^2}{\sigma_0^2}$

Cluster level covariate $R_2^2 = \dfrac{\tau_0 - \tau_1}{\tau_0}$