MS 2007-0081-RR
Booil Jo

Supplemental Materials
(to be posted on the Web)

Table 2 – Mplus Input

```
title:  Table 2
        Monte Carlo simulation using externally generated data.
        One level CACE analysis based on eqs. (17) & (18), ignoring clustering.

data:   file is treplist.dat; ! 500 data sets generated based eqs. (15) & (16).
        type=montecarlo;

variable: names are u y x1 z x2 c cluster;
          usev are u y z x1 x2;
          classes= c(2); ! analysis is done assuming 2 classes.
          categorical = u; ! u is a binary indicator of compliance.
          missing = u (999); ! missing is coded 999.

analysis: type = mixture missing; ! unobserved u in the control condition
                                  ! is treated as missing data.

model:
        %overall%

          [C#1*0]; ! eq.(18) logit intercept (50% compliance).
           C#1 on x1*0.7; ! logit coefficient in eq. (18).
           C#1 on x2*0.7; ! logit coefficient in eq. (18).

           y on z*0.6; ! intervention effect on the outcome.
           y on x1*-0.2; ! regression coefficient in eq. (17).
           y on x2*0.2;  ! regression coefficient in eq. (17).

          [y*1]; ! outcome intercept.
           y*1;  ! outcome residual variance.

        %c#1% ! noncompliers

          [u$1@15]; ! probability of being a complier = 0.
           y on z*-0.2; ! intervention effect for noncompliers.
           y on x1*-0.1; ! noncomplier regression coefficient (y on x1).
           y on x2*0.1;  ! noncomplier regression coefficient (y on x2).

          [y*1.0]; ! noncomplier outcome intercept in eq. (17).
           y*1.0;  ! noncomplier residual variance.

        %c#2% ! compliers

          [u$1@-15]; ! probability of being a complier = 1.
           y on z*0.6; ! intervention effect for compliers.
           y on x1*-0.2; ! complier regression coefficient (y on x1).
           y on x2*0.2;  ! complier regression coefficient (y on x2).

          [y*2]; ! complier outcome intercept in eq. (17).
           y*1.0; ! complier residual variance.
```

Table 3 – Mplus Input

```
title: Table 3
       Data generation and analysis according to eqs. (15) and (16).
       CACE analysis considering both clustering and noncompliance.
       Warning: this run may take long time (i.e., several days).
       Try it with one replication (nrep=1) and see how long
       multiple replications will take on your computer.

montecarlo:
       names are u y z x1 x2; ! y is a continuous outcome.
       nobservations = 2000; ! 100*20 = 2000.
       ncsizes = 1; ! cluster size is the same across clusters.
       csizes = 100 (20); ! 100 clusters with cluster size=20.
       genclasses = c(2); ! 2 compliance classes are generated.
       classes = c(2); ! analysis is done assuming 2 classes.
       between = z x2; ! x2 is a level 2 covariate.
                       ! intervention assigned (z) at the cluster level.
       within = x1; ! x1 is a level 1 covariate.
       generate = u(1); ! a binary variable u is generated
                        ! using a logistic model (1=logistic).
       categorical = u; ! u is a compliance indicator.
                        ! 0/1 for noncomplier/complier.
       missing = u; ! u is unobserved in the control condition
       cutpoints = z(0); ! z is the intervention assignment.
                         ! split at zero in the normal distribution.
                         ! 50% control, 50% intervention.
       seed = 4985107;
       nrep = 500;  ! 500 data sets generated.
       repsave = all; ! all 500 data sets will be saved.
       save = trep*.dat;

analysis:
       type = twolevel mixture missing;
                  ! considers both
                  ! clustering (twolevel) and noncompliance (mixture).
                  ! unobserved u in the control condition is treated
                  ! as missing data.

model missing:
        %overall%

     u on z@-30; ! model missing uses logistic regression.
                 ! probability of having missing compliance
                 ! information is zero for the intervention
                 ! condition individuals (z=1).
    [u@15];      ! otherwise (z=0),
                 ! compliance information is completely missing.

model population: ! data generation model.

        %within%
        %overall%

     x1*1;   ! generate a within-cluster covariate.
    [x1*0];  ! normally distributed with mean=0 and variance=1.
```

```
[c#1*0]; ! logit intercept in eq. (16).
 c#1 on x1*0.7; ! logit coefficient.

 y on x1*-0.2; ! within-cluster regression coefficient.
 y*0.8; ! within-cluster residual variance.

    %c#1% !noncompliers
 y on x1*-0.1; ! noncomplier within-cluster regression coefficient.
 y*0.9; ! noncomplier within-cluster residual variance.

    %c#2% !compliers
 y on x1*-0.2; ! complier within-cluster regression coefficient.
 y*0.8; ! complier within-cluster residual variance.

    %between%
    %overall%

 z*1;  ! generate a cluster level intervention assignment variable.
[z*0]; ! normally distributed with mean=0 and variance=1.
       ! this variable is dichotomized (see cutpoints command).
 x2*1;  ! generate a between-cluster covariate.
[x2*0]; ! normally distributed with mean=0 and variance=1.

 c#1*2.191; ! between cluster residual variance of compliance.
            ! according to eq. (13), ICCc=0.4.
 c#1 on x2*0.7; ! logistic regression coefficient of compliance on
                ! the between-cluster covariate x2 (c on x2).

 y on x2*0.2; ! between-cluster regression coefficient (y on x2).
 y on z*0.6;  ! intervention effect on the outcome.

 en by y*1; y@0;
[en@0];
 ec by y*1; y@0;
[ec@0];

 en*0.1; ! noncomplier between-cluster residual variance.
 ec*0.2; ! complier between-cluster residual variance.
 en with ec*0.0; ! covariance between macro-level residuals = 0.

    %c#1%  !noncompliers

[u$1@15]; ! probability of being a complier = 0.
 y on z*-0.2; ! intervention effect for noncompliers.
 y on x2*0.1; ! noncomplier between-cluster regression coeff.
[y*1];      ! noncomplier outcome intercept in eq. (15).
 en by y@1; ! en is between-cluster noncomplier residual variance.
 ec by y@0;

    %c#2% !compliers

[u$1@-15]; ! probability of being a complier = 1.
 y on z*0.6; ! intervention effect for compliers (CACE).
 y on x2*0.2; ! complier between-cluster regression coeff.
[y*2];      ! complier outcome intercept in eq. (15).
 en by y@0;
 ec by y@1; ! ec is between-cluster complier residual variance.
```

```
model: ! data analysis model.

        %within%
        %overall%

    [c#1*0];
     c#1 on x1*0.7;

     y on x1*-0.2;
     y*0.8;

        %c#1% !noncompliers
     y on x1*-0.1;
     y*0.9;

        %c#2% !compliers
     y on x1*-0.2;
     y*0.8;

        %between%
        %overall%

     c#1*2.191;
     c#1 on x2*0.7;

     y on x2*0.2;
     y on z*0.6;

     en by y*1; y@0;
    [en@0];
     ec by y*1; y@0;
    [ec@0];
     en*0.1;
     ec*0.2;
     en with ec*0.0;

        %c#1%  !noncompliers

    [u$1@15];
     y on z*-0.2;
     y on x2*0.1;
    [y*1];
     en by y@1;
     ec by y@0;

        %c#2% !compliers

    [u$1@-15];
     y on z*0.6;
     y on x2*0.2;
    [y*2];
     en by y@0;
     ec by y@1;
```

A Real Data Application – Mplus input for 2-level CACE estimation

```
title:    Real Data Analysis
          Two level CACE analysis based on eqs. (15) & (16), considering
          clustering & noncompliance.
data:     file is real.dat; ! the name of the real data set is “real.dat”.

variable: names are u y z x1 x2 teacher;
usev are u y z x1 x2 teacher;
classes= c(2); ! analysis is done assuming 2 classes.

categorical = u; ! u is a binary indicator of compliance (0 = noncomplier,
                 ! 1 = complier, 999 = compliance status unknown).
missing = u (999); ! missing is coded 999.

between = z x2; ! x2 is a level 2 covariate.
                ! intervention assigned (z) at the cluster (teacher) level.
within = x1;    ! x1 is a level 1 covariate.
cluster = teacher; ! students are nested within teacher
```

analysis: type = twolevel mixture missing;

```
model: ! data analysis model.

        %within%
        %overall%

    [c#1]; ! logit intercept in eq. (16).
     c#1 on x1; ! logistic regression coefficient of compliance on
                ! the within-cluster covariate x1 (c on x1).
     y on x1;   ! within-cluster regression coefficient.
     y;         ! within-level outcome residual variance.

     ! any level-1 parameters that are allowed to vary across compliers and
     ! noncompliers are specified under each compliance class below.

        %c#1% !noncompliers
     y on x1;
     y;

        %c#2% !compliers
     y on x1;
     y;

        %between%
        %overall%

     c#1*2.191; ! between cluster residual variance of compliance
                !(need a nonzero starting value)
     c#1 on x2; ! logistic regression coefficient of compliance on
                ! the between-cluster covariate x2 (c on x2).

     y on x2;   ! between-cluster regression coefficient.
     y on z;    ! intervention effect on the outcome.
     [y]; ! intercept.

     en by y@1; y@0;
    [en@0];
     ec by y@1; y@0;
```

```
[ec@0];
 en; ! en is between-cluster noncomplier residual variance.
 ec; ! ec is between-cluster complier residual variance.
 en with ec; ! covariance of the noncomplier and complier outcome residuals.
 c#1 with ec; ! covariance of the complier outcome and compliance residuals.
 c#1 with en; ! covariance of the noncomplier outcome and compliance
              ! residuals.

 ! any level-2 parameters that are allowed to vary across compliers and
 ! noncompliers are specified under each compliance class below.

    %c#1%  !noncompliers

[u$1@15]; ! probability of being a complier = 0.
 y on z;
 y on x2;
[y];
 en by y@1;
 ec by y@0;

    %c#2% !compliers

[u$1@-15]; ! probability of being a complier = 1.
 y on z;
 y on x2;
[y];
 en by y@0;
 ec by y@1;
```

A Real Data Application: Simpler Analyses
Mplus input for 2-level single class analysis to get ICCyc

```
title:    Real Data Analysis
          Two level analysis

data:     file is real.dat; ! the name of the real data set is “real.dat”.

variable: names are u y z x1 x2 teacher;
usev are y teacher;
useobs = (z==1 and u==1); ! select only compliers
                          ! (to select noncompliers use u==0).
cluster = teacher; ! students are nested within teacher
analysis: type = twolevel;

model:
        %within%

     y;        ! within-level outcome variance.

        %between%

     [y];      ! mean.
     y;        ! between-level outcome variance.
```

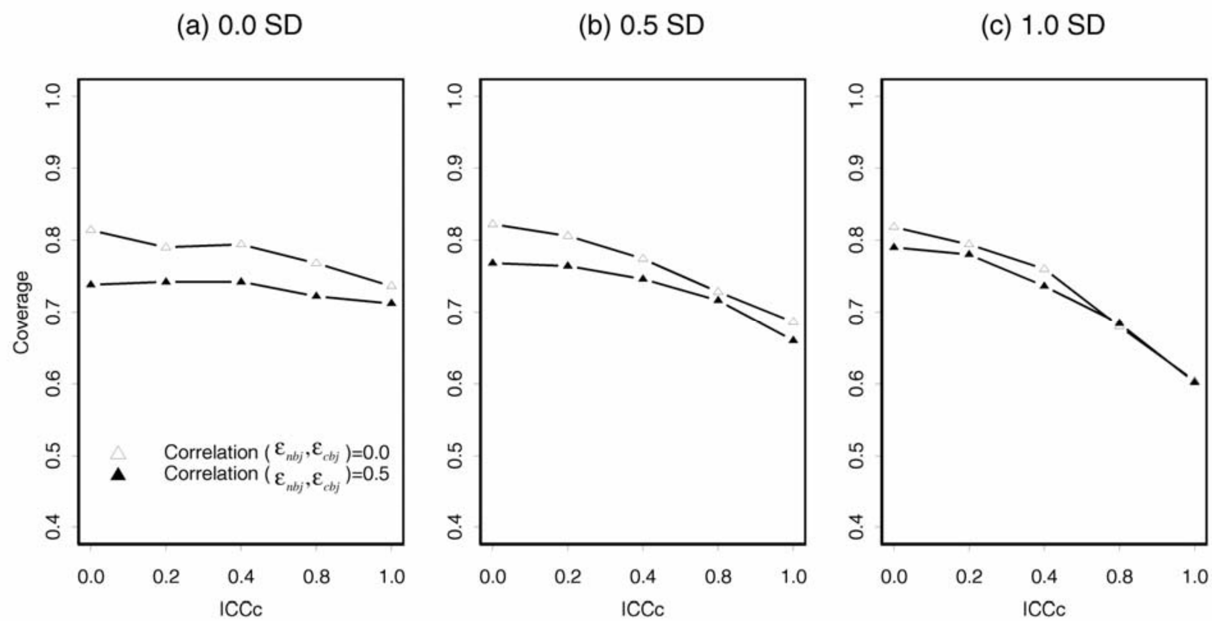
A Real Data Application: Simpler Analyses
Mplus input for 2-level logistic regression to get ICCc

```
title:   Real Data Analysis
         Two level analysis

data:    file is real.dat; ! the name of the real data set is “real.dat”.

variable: names are u y z x1 x2 teacher;
usev are u teacher;
categorical = u;
useobs = (z==1); ! select only treatment group

cluster = teacher; ! students are nested within teacher
analysis: type = twolevel;

model:
        %within%

        %between%
     [u$1];      ! threshold or intercept in the empty logistic regression.
      u;         ! between-level compliance variance.
```

A Real Data Application: Simpler Analyses
Mplus input for 1-level CACE estimation

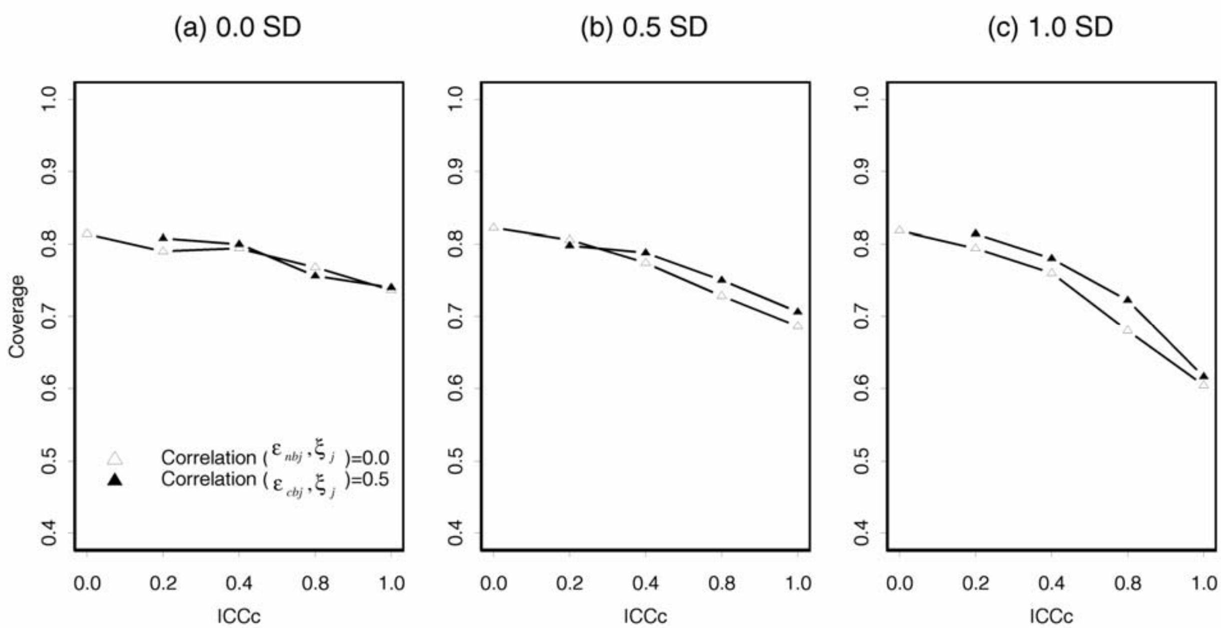
```
title:    Real Data Analysis
          One level CACE analysis ignoring clustering.
data:     file is real.dat; ! the name of the real data set is “real.dat”.

variable: names are u y z x1 x2 teacher;
usev are u y z x1 x2;
classes= c(2); ! analysis is done assuming 2 classes.

categorical = u; ! u is a binary indicator of compliance (0 = noncomplier,
                 ! 1 = complier, 999 = compliance status unknown).
missing = u (999); ! missing is coded 999.

analysis: type = mixture missing;


model:

  %overall%

    [C#1]; ! logit intercept.
     C#1 on x1; ! logit coefficient.
     C#1 on x2; ! logit coefficient.

     y on z;    ! intervention effect on the outcome.
     y on x1;   ! regression coefficient.
     y on x2;    ! regression coefficient.

    [y]; ! outcome intercept.
     y;  ! outcome residual variance.

  %c#1% ! noncompliers

    [u$1@15];   ! probability of being a complier = 0.
     y on z;    ! intervention effect for noncompliers.
     y on x1;   ! noncomplier regression coefficient.
     y on x2;   ! noncomplier regression coefficient.

    [y]; ! noncomplier outcome intercept.
     y;  ! noncomplier residual variance.

  %c#2% ! compliers

    [u$1@-15];  ! probability of being a complier = 1.
     y on z;    ! intervention effect for compliers.
     y on x1;   ! complier regression coefficient.
     y on x2;   ! complier regression coefficient.

    [y];        ! complier outcome intercept.
     y;         ! complier residual variance.
```

Figure 5: Impact of ICC_y and ICC_c on variance misestimation when the correlation between the macro-unit outcome residuals (ε_{nbj} and ε_{nbj}) is 0.0 vs. when the correlation between the macro-unit outcome residuals (ε_{nbj} and ε_{nbj}) is 0.5. Each cluster consists of 20 individuals. $\text{ICC}_{yn} = \text{ICC}_{yc} = 0.1$. Complier and noncomplier outcome means are (a) 0.0, (b) 0.5, and (c) 1.0 standard deviation apart given treatment assignment. The y-axis represents the coverage rate for nominal 95% confidence intervals.



In the simulations reported in Figure 3, data were generated and analyzed assuming that the correlation between the two between-cluster residuals is zero. In practice, these residuals may be correlated. That is, outcome variation across classrooms given treatment assignment may be similar or very different for compliers and noncompliers. As this correlation deviates from zero, the results can be somewhat different from those reported in Figure 3. The simulation results reported in **Figure 5** demonstrate how deviation from zero correlation affects variance estimation, focusing on the setting where ICCyn = ICCyc = 0.1. It is shown that the coverage of the CACE estimate is somewhat lower when the correlation between the two between-cluster residuals is 0.5 than when the correlation is zero. However, both settings lead to similar results as complier and noncomplier means have a substantial distance (i.e., 1.0 SD apart).

Figure 6: Impact of ICC_y and ICC_c on variance misestimation when the correlation between the macro-unit outcome and compliance residuals is 0.0 (between ε_{nbj} and ξ_j is zero, between ε_{cbj} and ξ_j is zero) vs. when the correlation between the macro-unit outcome and compliance residuals is 0.5 (between ε_{nbj} and ξ_j is 0.5, between ε_{cbj} and ξ_j is 0.5). Each cluster consists of 20 individuals. $ICC_{yn} = ICC_{yc} = 0.1$. Complier and noncomplier outcome means are (a) 0.0, (b) 0.5, and (c) 1.0 standard deviation apart given treatment assignment. The y-axis represents the coverage rate for nominal 95% confidence intervals.



In the simulations reported in Figure 3, data were also generated and analyzed assuming that the correlation between compliance and outcome at the cluster level is zero. This correlation may increase in some trials, for example, where clusters with higher proportions of compliers tend to have better outcomes given treatment assignment. The simulation results reported in **Figure 6** demonstrate how deviation from zero correlation affects variance estimation, focusing on the setting where ICCyn = ICCyc = 0.1. In general, the coverage of the CACE estimate shows little difference between the setting where the true correlation between compliance and outcome at the cluster level is 0.5 and the setting where the correlation is zero. The coverage is slightly higher in the setting where the correlation is 0.5 as complier and noncomplier means have a substantial distance.