

Analyzing Data from Experiments in which the Treatment Groups Have Different Hierarchical Structures

Sharon L. Lohr, Westat; Peter Z. Schochet, Mathematica Policy Research

ABSTRACT

In randomized experiments, it is generally assumed that the hierarchical structures and variances are the same in the treatment and control groups. In some situations, however, these structures and variance components can differ. Consider a randomized experiment in which individuals randomized to the treatment condition are further assigned to clusters in which the intervention is administered, but no such clustering occurs in the control condition. Such a structure can occur, for example, when the individuals in the treatment condition are randomly assigned to group therapy sessions or to mathematics tutoring groups: individuals in the control condition do not receive group therapy or mathematics tutoring and therefore do not have that level of clustering. In this example, individuals in the treatment condition have a hierarchical structure but individuals in the control condition do not. If the therapists or tutors differ in efficacy, the clustering in the treatment condition induces an extra source of variability in the data that needs to be accounted for in the analysis. We show how special features of SAS[®] PROC MIXED and PROC GLIMMIX can be used to analyze data in which one or more treatment groups have a hierarchical structure that differs from that in the control group. We also discuss how to code variables to increase the computational efficiency for estimating parameters from these designs.

INTRODUCTION

Randomized controlled trials (RCTs) are commonly recommended for studies in which it is desired to establish whether an experimental intervention causes changes in outcomes.

The simplest type of RCT randomizes *individuals* to treatment groups. A population of individuals is available for the experiment. In a balanced design, half of the individuals are randomly assigned to treatment A and the remaining half are assigned to treatment B. For example, suppose a pool of students is available for an educational experiment. The students are randomly assigned to treatment A (the control group) or treatment B (the intervention). The randomization act allows us to make the assumption that all observations are independent. The treatment effect may then be assessed using a procedure for analyzing data from two independent samples such as TTEST or NPAR1WAY. If the outcome is binary (for example, whether the student graduates), the FREQ procedure may be used to analyze the data. If there are covariates, the REG, GLM, or LOGISTIC procedures may be used.

In a cluster-randomized experiment, *clusters* of individual units, rather than the individual units themselves, are randomly assigned to treatment groups. An example of a clustered RCT is an educational experiment in which schools are randomly assigned to either treatment A or treatment B. All students in a particular school will be in the same treatment group. If the experiment were run again, a particular school might be in either treatment A or treatment B, but all of the students will be together receiving the same treatment. Students in the same school have the same teachers and school environment, and therefore are expected to have outcomes that are more similar to each other than students who are selected at random from the set of all schools. This similarity means that students in the same school often have positively correlated outcomes.

While randomization in a clustered RCT is done at the cluster level, the outcomes are often measured and analyzed at the individual level. Hierarchical models, also referred to as multilevel models or mixed models, are commonly used to analyze data from a clustered RCT (Donner and Klar, 2004). These models explicitly account for the variability that occurs at each level of the experiment. In a clustered RCT with schools randomized to treatment groups, there are two sources of variability: the variability among school means induced by the school environmental factors and peer effects, and the variability from student to student within each school. It is generally assumed that the variability at each level is the *same* in each treatment group. There is a complete nesting structure: students are nested in schools for both treatment A and treatment B.

In this paper we discuss how to analyze experiments in which the hierarchical structure *differs* in the two treatment groups. This structure arises when the protocol for treatment B involves placing individuals in clusters, while the protocol for treatment A does not involve clustering. The experiment is *partially nested* because the individuals in treatment B are nested in clusters but the individuals in treatment A are not. We give three examples of partially nested experiments.

Example 1. A pool of students is available for an experiment evaluating the effects of a summer tutoring program. Half of the students are randomly assigned to the treatment group, and the other half are assigned to the control group. The students in the control group receive no tutoring. The students in the treatment group are divided into

small groups, and each group receives tutoring over the summer from a different tutor. It is expected that the tutor and peer effects in the treatment group will cause the students in the same tutor group to have more similar outcomes. Thus, there is clustering in the treatment group but not in the control group.

Example 2. Persons with mild depression are randomly assigned to one of two treatments. In treatment A, persons are given instructions for an exercise regimen, which they are to carry out individually. In treatment B, persons are assigned to one of six group therapy sessions, where each therapy group has 8 persons. The therapist effects and group dynamics may induce clustering in treatment B that does not occur in treatment A.

Example 3. Fifteen schools are available for a study of a small-group intervention intended to improve reading in second grade students. In each school, the students who need improvement in reading are randomly assigned to treatment A (individualized tutoring) or treatment B (tutoring in a group setting with groups of four students). Each school has the same basic design structure described in Example 1: in each school, the treatment B students receive the protocol in a group setting, while the treatment A students receive the protocol individually. The basic design structure in Example 1 is replicated in each of the fifteen schools. In this example, the schools are a *blocking factor* for the design, and this is a *blocked partially nested design*.

Example 4 is a clustered RCT, but has features of partially nested designs.

Example 4. Classrooms are randomly assigned to treatment A or treatment B. In the classrooms assigned to treatment A, students at risk for difficulties in mathematics receive their usual classroom instruction. In the classrooms assigned to treatment B, at-risk students receive the usual classroom instruction; in addition, they receive tutoring in groups of five students outside of school. This is a clustered RCT, because classrooms of students are assigned to the treatment groups. Within each classroom, however, the students in the treatment A classrooms have a different dependence structure than the students in the treatment B classrooms. The scores of students in treatment B have an additional source of clustering (the tutor groups) that is not found in treatment A.

All of these examples have the common feature that the individuals assigned to treatment A (often the control condition) have a different hierarchical structure than the individuals in treatment B. For examples 1 and 2, the individuals in treatment A have a structure similar to that in an individual RCT, while the individuals in treatment B have a structure similar to that in a clustered RCT. In example 3, the basic partially nested design, with clustering in treatment B but not in treatment A, is replicated across different blocking units. In example 4, the students in both treatments A and B are clustered, but treatment B has an additional level of variability (the tutor groups) not found in treatment A.

In this paper, we illustrate how to analyze data from experiments such as these in SAS[®] software. Options of the MIXED and GLIMMIX procedures allow fitting different hierarchical structures for multiple treatment groups. We also discuss how to code the variables to achieve greater computational efficiency.

PARTIALLY NESTED DESIGNS

In this section we discuss the basic partially nested design used for Examples 1 and 2. The key features of the simplest form of a partially nested design are as follows:

- Individuals are randomly assigned to treatment A or treatment B. Treatment A is often a control, or status quo, condition, in which individuals receive no intervention.
- After randomization, the individuals in treatment B are randomly assigned to clusters. The shared environment in each cluster induces dependence in the outcomes of individuals in the cluster.
- The individuals in treatment A are not subject to clustering. They may therefore be treated as independent observations.
- The variance at the cluster level therefore differs for treatments A and B: treatment B has variability at the cluster level, while treatment A does not.
- The variance at the individual level may also differ for treatments A and B. This may occur if the treatment B clustering causes the outcomes of individuals in the same cluster to be more or less homogeneous than outcomes of individuals in treatment A.

MIXED MODEL FOR A PARTIALLY NESTED RCT

The outcomes of individuals in treatment B may be described by the following model. Let y_{ij} be the outcome of individual j in cluster i , for $i = 1, 2, \dots$ (number of clusters in treatment B) and $j = 1, 2, \dots$ (number of individuals in cluster i). Then a hierarchical model that captures both levels of variability is

$$y_{ij} = \mu_B + v_i + e_{ij},$$

where v_i has mean 0 and variance σ_{vB}^2 and e_{ij} has mean 0 and variance σ_{eB}^2 .

To analyze the data in PROC MIXED, we need to specify a cluster-level variable for the observations in treatment A as well, even though that treatment group has no clustering. We therefore construct “fictional clusters” for the individuals in treatment A—clusters that have mean 0 and variance 0 so that they induce no additional variability. We may do this by assigning all individuals in treatment A to a cluster labeled as 0, or assigning each individual to his or her own cluster. In treatment A, then, we also write

$$y_{ij} = \mu_A + v_i + e_{ij},$$

but now v_i has mean 0 and variance $\sigma_{vA}^2 = 0$ and e_{ij} has mean 0 and variance σ_{eA}^2 . Because we force the random effects v_i to equal 0 in treatment A, the model for the treatment A individuals may be equivalently written as

$$y_{ij} = \mu_A + e_{ij}.$$

We can combine the models for the individuals in treatments A and B into a unified model by defining the treatment indicator $T_{ij} = 1$ if individual (ij) is in treatment B and $T_{ij} = 0$ if individual (ij) is in treatment A. Then, the model for all individuals in the experiment is

$$y_{ij} = \alpha + \beta T_{ij} + v_i + e_{ij}, \quad (1)$$

where $\alpha = \mu_A$, $\beta = \mu_B - \mu_A$, v_i has mean 0 and variance σ_{vB}^2 if $T_{ij} = 1$, v_i has mean 0 and variance $\sigma_{vA}^2 = 0$ if $T_{ij} = 0$, e_{ij} has mean 0 and variance σ_{eB}^2 if $T_{ij} = 1$, and e_{ij} has mean 0 and variance σ_{eA}^2 if $T_{ij} = 0$. The parameter β gives the effect of the treatment.

If desired, the simplifying assumption that $\sigma_{eA}^2 = \sigma_{eB}^2$ may be made. In some experiments, however, the nature of the intervention and the clustering in treatment B might cause the residual variance in treatment B to differ from that in treatment A. For example, it is possible that the intervention causes the outcomes of individuals in the same group to be more similar than outcomes of individuals in the control group. This might occur because the intervention might induce conformity, or because peer effects might cause individuals in the same group to have similar opinions after the intervention. For the analysis, no harm is done by allowing the residual variances to differ: the only cost is the loss of one degree of freedom for the residual variance.

To express the partially nested design using the conventions in Goldstein (2011), treatment B has level-1 units (students in Example 1; patients in Example 2) and level-2 units (tutor groups in Example 1, therapy groups in Example 2). Treatment A has only level-1 units. Level-2 units are not formed in treatment A. To analyze the data, however, we use the fiction that there are level-2 units in treatment A, but the random effects for those units have zero mean and variance.

Although we define clusters for both treatment A and treatment B participants, only the clusters for treatment B are “real.” The “clusters” defined for treatment A are merely a notational convenience that allow us to fit the model. This is in contrast to the erroneous approach sometimes taken for analyzing data from partially nested designs (Bauer et al., 2008, describe the pitfalls of this approach) in which the individuals in treatment A are treated as clusters having the same variability as in treatment B.

USING PROC MIXED FOR PARTIALLY NESTED RCTS

The MIXED procedure can fit different hierarchical structures for the two treatment groups with the GROUP = option. According to SAS Institute Inc. (2011), the GROUP = option “defines an effect specifying heterogeneity in the covariance structure of \mathbf{G} .” We can fix $\sigma_{vA}^2 = 0$ by using the PARMS statement, holding that variance component to the initial value through the HOLD option. The PARMS statement requires initial values for all of the variance components. These can be set to values that seem reasonable, or initial values can be estimated by fitting models separately for the two treatment groups.

The following code fits the model in equation (1).

```
proc mixed data=pnrct;
  class student trtname cluster;
  model y=trt / solution ddfm=sat;
  random intercept / group=trtname subject=cluster(trtname);
  repeated student / group=trtname;
  parms (0) (50) (200) (200) / hold = 1;
```

The generated data set *PNRCT* used for this example has 100 students in each treatment group. The students in treatment B are assigned to one of 20 tutor groups with 5 students each. The variable *TRT* has the value 0 for students in treatment A, and the value 1 for students in treatment B. The variable *TRTNAME* is a categorical variable

taking on values “Control” and “Treatment.” For this example, the variable *CLUSTER* has 21 levels: the real clusters (tutor groups) in treatment B are labeled from 1 to 20, and all the students in treatment A (the control group) are placed in an artificial cluster labeled as 0.

The RANDOM statement fits separate cluster-level variances for the treatment groups, and the PARMs statement specifies that the cluster-level variance for the first treatment group is fixed at 0. The REPEATED statement specifies fitting separate individual-level variances for the two treatment groups. The REPEATED statement may be omitted if it is assumed that the individual-level variances are the same for both groups (i.e., $\sigma_{eA}^2 = \sigma_{eB}^2$).

The LOG file after running this code gives the message in Display 1. This message does *not* describe an error. The PARMs statement fixes the first variance component at zero. The **G** matrix, therefore, is singular and consequently cannot be positive definite. We set up the code deliberately so that the **G** matrix would be singular.

NOTE: Convergence criteria met.
 NOTE: Estimated G matrix is not positive definite.
 NOTE: Asymptotic variance matrix of covariance parameter estimates has been found to be singular and a generalized inverse was used. Covariance parameters with zero variance do not contribute to degrees of freedom computed by DDFM=SATTERTH.

Display 1. LOG File for Partially Nested Model Analysis.

Output 1 contains selected output from this model. This analysis uses restricted maximum likelihood (the default method in PROC MIXED), but maximum likelihood can be used if it is desired to conduct likelihood ratio tests of the parameters. The Class Level Information confirms that there are 21 levels for the *CLUSTER* variable. All of the treatment A students are placed in cluster 0. The covariance parameter estimates are $\hat{\sigma}_{vB}^2 = 73.1008$, $\hat{\sigma}_{eA}^2 = 286.96$, and $\hat{\sigma}_{eB}^2 = 252.75$. The value of σ_{vA}^2 is fixed at zero as desired.

The Satterthwaite option is used to calculate the degrees of freedom for testing the effect of *TRT*. For this example, there are 100 independent units (students) for treatment A, but there are 20 independent units (clusters) for treatment B. Because the hierarchical structures differ for treatments A and B, the variance of the independent units in treatment A does not equal the variance of the independent units in treatment B. The Satterthwaite method adjusts the degrees of freedom for the unequal variances of the independent units. For this example, 39.1 degrees of freedom are used, and the estimated treatment effect is $\hat{\beta} = 6.685$. For large studies, the degrees of freedom should make little difference to the inferences, but for studies with small numbers of clusters, the default method for calculating degrees of freedom often overstates them, and we recommend using the Satterthwaite method as a correction.

BINARY RESPONSES

Binary responses from a partially nested RCT may be analyzed similarly, using PROC GLIMMIX. The GROUP = option and PARMs statement are used to fix the cluster-level variability at 0 for the treatment A students. The following code may be used to estimate the treatment effect when the response of interest, *YBIN*, takes on values 0 and 1:

```
proc glimmix data=pnrct method=quad;
  class trtname;
  model ybin (event='1') = trt / solution dist=binary link=probit;
  random intercept / group=trt subject=cluster(trtname);
  parms (0) (.2) / hold = 1;
```

This example uses probit regression, but logistic regression (LINK = LOGIT) could alternatively be used for this analysis. Note that the PARMs statement for PROC GLIMMIX contains only two values, while the PROC MIXED code for a continuous response has four variance parameters. Only two parameters are needed because the residual variance is fixed at 1 in a probit-normal model.

The default computational method in PROC GLIMMIX is a pseudo-likelihood method. In our simulation studies, the pseudo-likelihood methods in PROC GLIMMIX tended to underestimate the cluster-level variance component. We recommend using the Laplace approximation (METHOD = LAPLACE) or adaptive quadrature (METHOD = QUAD) to estimate the parameters. For small cluster sizes, METHOD = QUAD is preferred.

The Satterthwaite method for adjusting degrees of freedom is not currently available in PROC GLIMMIX with the quadrature method. If desired, the degrees of freedom may be entered manually using the DDF option to the MODEL statement.

Class Level Information		
Class	Levels	Values
student	200	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200
trtname	2	Control Treatment
cluster	21	0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20

Covariance Parameter Estimates			
Cov Parm	Subject	Group	Estimate
Intercept	cluster(trtname)	trtname Control	0
Intercept	cluster(trtname)	trtname Treatment	73.1008
student		trtname Control	286.96
student		trtname Treatment	252.75

Solution for Fixed Effects					
Effect	Estimate	Standard Error	DF	t Value	Pr > t
Intercept	98.9067	1.6940	99	58.39	<.0001
trt	6.6850	3.0087	39.1	2.22	0.0321

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
trt	1	39.1	4.94	0.0321

Output 1. Output for Partially Nested Model Analysis.

IMPROVING COMPUTATIONAL EFFICIENCY FOR PROC MIXED

For small or moderately sized data sets, the code presented earlier works well and requires minimal pre-processing of the data. For larger data sets, however, the asymmetric structure of the variance components in a partially nested design can result in long computation times or even a failure to converge. Kiernan et al. (2012) discuss methods that may be used to obtain greater computational efficiency from PROC MIXED. One of their suggestions is to use a nested SUBJECT = effect. Instead of having the variable CLUSTER take on different values for every cluster in the experiment, create a variable CLUSTER_NEST that takes on values 1, 2, ... within each treatment group. For treatment B, CLUSTER_NEST is coded as 1, 2, ... (number of clusters) (1 to 20 in the above example) and is the same as the variable CLUSTER. For treatment A, CLUSTER_NEST is coded as 1, 2, ... for the individual students in treatment A (1 to 100 in the above example). Similarly, the variable STUDENT, which has a unique value for every student, can be recoded as variable STUDENT_NEST, which repeats the values 1, 2,... within each treatment group. In the above example, STUDENT_NEST would be coded as 1 to 100 for Treatment B students and 1 to 100 for Treatment A students. Thus, STUDENT_NEST is the same as CLUSTER_NEST for the students in Treatment A.

Another option for speeding computations is to use a continuous, rather than categorical, variables for the GROUP = and SUBJECT = options. SAS Institute Inc. (2011) states: "Using a continuous variable decreases execution time for models with a large number of subjects or groups and also prevents the production of a large 'Class Level Information' table." However, "PROC MIXED does not sort by the values of the continuous variable; rather, it considers the data to be from a new subject or group whenever the value of the continuous variable changes from the previous observation." If continuous variables are used for the GROUP = or SUBJECT = options, the data *must* be sorted before using PROC MIXED.

The following code uses continuous variables *TRT* and *CLUSTER_NEST* for the GROUP = and SUBJECT = options. These variables are treated as continuous because they are omitted from the CLASS statement. The continuous variable *CLUSTER_NEST* is nested in the categorical variable *TRTNAME*; nesting in continuous variables is not allowed.

```
proc sort data=pnrct;
  by trt cluster_nest;
proc mixed data=pnrct noclprint noitprint plots=none;
  class student_nest trtname;
  model y=trt / solution ddfm=sat;
  random intercept / group=trt subject=cluster_nest(trtname);
  repeated student_nest / group=trt;
  parms (0) (50) (200) (200) / hold = 1;
```

We performed a factorial experiment to study which settings result in the fastest computations. The factors studied were: (1) number of clusters in treatment group and total number of students, (2) treating GROUP = as continuous or class variable, (3) treating SUBJECT = variable as continuous or class variable, and (4) using duplicated variable names for clusters and students. All runs used the options NOCLPRINT, NOITPRINT, and PLOTS=NONE.

Table 1 gives the CPU time for the different runs. It is clear that using the duplicated level names resulted in the greatest increases in speed. For large data sets, using a continuous variable for the SUBJECT = variable also speeds up the computations. The code with duplicated level names and with continuous variable *CLUSTER_NEST* produced the fastest run times for PROC MIXED.

Total number of students	Number of clusters in treatment group	GROUP =: continuous or categorical variable	SUBJECT = variable: continuous or categorical variable	Use duplicated level names for clusters and students	CPU time (seconds)
2000	100	categorical	categorical	no	5.88
2000	100	categorical	categorical	yes	0.12
2000	100	categorical	continuous	no	5.89
2000	100	categorical	continuous	yes	0.09
2000	100	continuous	categorical	no	5.91
2000	100	continuous	categorical	yes	0.10
2000	100	continuous	continuous	no	6.05
2000	100	continuous	continuous	yes	0.09
200000	10000	categorical	categorical	no	Insufficient memory
200000	10000	categorical	categorical	yes	32.87
200000	10000	categorical	continuous	no	Insufficient memory
200000	10000	categorical	continuous	yes	19.73
200000	10000	continuous	categorical	no	Insufficient memory
200000	10000	continuous	categorical	yes	33.21
200000	10000	continuous	continuous	no	Insufficient memory
200000	10000	continuous	continuous	yes	19.75

Table 1. CPU Time for Run Options with Partially Nested Design in PROC MIXED.

INCLUDING COVARIATES OR MULTIPLE TREATMENT GROUPS

Covariates at the individual level (level 1) are included in the usual way, through the MODEL statement. These covariates may be desired in the model to control for baseline characteristics of the students, or to adjust for missing data in treatments A and B.

Sometimes there are covariates at the cluster level as well. For example, suppose that half of the tutors are experienced and the other half of the tutors are inexperienced. Including tutor experience as a covariate can decrease the variability at the cluster level (σ_{vB}^2) and hence increase the precision of the estimated treatment effect. That covariate only appears for treatment B, however, and it is important to code the covariate carefully so that the main effect of the treatment can still be estimated cleanly. In a balanced design, where every cluster has the same size, this is done by defining cluster level covariates so that they are orthogonal to the main treatment effect. Thus, suppose a covariate for tutor experience has two levels: experienced and inexperienced. Define the covariate *Experience* for the treatment B students only:

$$Experience_{ij} = \begin{cases} 1 & \text{if tutor of student } (ij) \text{ is experienced} \\ 0 & \text{if tutor of student } (ij) \text{ is inexperienced} \end{cases}$$

Then define a centered version of the covariate:

$$CenterExp_{ij} = \begin{cases} Experience_{ij} - \overline{Experience} & \text{if student } (ij) \text{ is in treatment B} \\ 0 & \text{if student } (ij) \text{ is in treatment A} \end{cases}$$

where $\overline{Experience}$ is the average of the values of $Experience_{ij}$ for the students in treatment B. Then the covariate *CENTEREXP* is orthogonal to the treatment indicator which takes on the value 0 in treatment A and 1 in treatment B.

The following code will fit a model with a cluster-level covariate. If the design is balanced, the coefficient of *TRT* is the average score of the treatment students minus the average score of the control students, and that coefficient is the same whether the covariate *CENTEREXP* is included in the model or not. With unbalanced data, the estimated parameters depend on the variance components and the estimate of the treatment effect may change slightly if the cluster-level covariate is added to the model.

```
proc mixed data=pnrct2;
  class student trtname cluster;
  model y=trt centerexp / solution ddfm=sat;
  random intercept / group=trtname subject=cluster(trtname);
  repeated student / group=trtname;
  parms (0) (50) (200) (200) / hold = 1;
```

An alternative analysis could be performed if the cluster-level covariate is categorical. In that case, each value of the covariate could be considered to be a different treatment. In the example above, then, there would be three treatment groups: the control group (*TRTGRP* = 0), treatment group 1 with inexperienced tutors (*TRTGRP* = 1), and treatment group 2 with experienced tutors (*TRTGRP* = 2).

The following code provides an analysis of data with three treatment groups, where the students with *TRTGRP* = 0 are unclustered and the students in *TRTGRP* 1 and 2 are formed in clusters after randomization. The cluster-level variances and the residual-level variances are allowed to differ in the three treatment groups. The experiment needs to have at least two clusters within *TRTGRP* 1 and at least two clusters within *TRTGRP* 2 for the separate variance components to be estimable. If desired, a common cluster-level variance component can be assumed for *TRTGRP* 1 and 2 by replacing *GROUP* = *TRTGRP* in the *RANDOM* statement by *GROUP* = *TRT*, where *TRT* = 0 if *TRTGRP* = 0 and *TRT* = 1 if *TRTGRP* = 1 or 2; the *PARMS* statement would then contain only five values. The *ESTIMATE* statements compare the outcomes in *TRTGRP* 1 and 2 with the outcomes in the control group.

```
proc mixed data=pnrct2;
  class student trtgrp cluster;
  model y=trtgrp / solution ddfm=sat;
  estimate 'treatment 1 minus control' trtgrp -1 1 0;
  estimate 'treatment 2 minus control' trtgrp -1 0 1;
  random intercept / group=trtgrp subject=cluster(trtgrp);
  repeated student / group=trtgrp;
  parms (0) (50) (50) (200) (200) (200) / hold = 1;
```

KEY POINTS FOR ANALYZING DATA FROM A PARTIALLY NESTED RCT

In summary, here are the key points for using PROC MIXED or PROC GLIMMIX to analyze partially nested data:

- Use the GROUP = option to the RANDOM statement to specify that the variances differ in the treatment groups.
- Use the PARMS statement to fix the cluster-level variance at 0 for treatment A. The PARMS statement assigns values to the parameters in the order in which they are entered into the model, so be sure to verify that the correct variance component is held at 0.
- In PROC MIXED, use the REPEATED statement to fit different residual variances for the treatments. The REPEATED statement can be omitted if the residual variances are assumed to be the same for the treatment groups. Do not include a RANDOM statement for the residual variances in PROC GLIMMIX if the response is binary.
- In PROC MIXED, use DDFM = SATTERTHWAIT to calculate the appropriate degrees of freedom for the test of the treatment effect.
- In PROC GLIMMIX, use METHOD = QUAD or METHOD = LAPLACE to fit the model. METHOD = QUAD is preferred.

BLOCKED PARTIALLY NESTED RCTS

Example 3 describes a blocked partially nested RCT, in which the basic partially nested RCT is replicated at different sites. The site, therefore, adds an additional variance component to the model. The key features of a blocked partially nested RCT are as follows:

- Potential participants in the experiment are arranged in K blocks. These blocks can be naturally occurring units such as schools, cities, or clinics. Alternatively, blocks may be deliberately formed by the researcher prior to randomization, for example if students are assigned to blocks by their scores on a pretest.
- Randomly assign participants to treatment A or treatment B within each block. The randomization is performed separately for each block.
- The students in treatment B receive the intervention in a group setting, forming clusters within each block.
- The students in treatment A receive the treatment A protocol (often the status quo) individually within the block.
- For each block k , the students in treatment A within block k serve as the comparison group for the students in treatment B within block k .

Each block contains its own basic partially nested RCT. We now present an example in which there are $K = 15$ sites (= blocks). In each site, 20 students are available for the study: 10 of them are randomly assigned to treatment A and the other 10 are assigned to treatment B. The 10 treatment B students in each site are randomly assigned to two tutor groups (clusters), each with five students. In this design, every site has 10 students in treatment A and 10 students in treatment B. In the analysis, a treatment effect is computed separately for each site, and then the treatment effects from the different sites are combined to compute an overall treatment effect.

A random coefficient regression model allows us to have a different treatment effect at each site. The model may be expressed mathematically as follows. Let y_{kij} represent the outcome for student j in cluster i of site k : as with the basic partially nested RCT model, we set $i = 0$ for students in treatment A group at each site. Let $T_{kij} = 1$ if student (kij) is assigned to treatment B, and $T_{kij} = 0$ if student (kij) assigned to treatment A. The model is:

$$y_{kij} = \alpha + \beta T_{kij} + [a_k + b_k T_{kij}] + v_{ki} + e_{kij}. \quad (2)$$

The terms outside of the brackets are essentially the same as in equation (1) for the basic partially nested design. The parameter β is the overall treatment effect, and v_{ki} and e_{kij} are the random terms at the cluster and student levels. As before, we assume that v_{ki} has mean 0 and variance σ_{vB}^2 if $T_{kij} = 1$, v_{ki} has mean 0 and variance $\sigma_{vA}^2 = 0$ if $T_{kij} = 0$, e_{kij} has mean 0 and variance σ_{eB}^2 if $T_{kij} = 1$, and e_{kij} has mean 0 and variance σ_{eA}^2 if $T_{kij} = 0$.

The terms in brackets in equation (2) account for the site effects, and allow the treatment impact to differ across sites. We assume that a_k and b_k , which give the site-specific intercept and slope, follow a bivariate normal distribution with mean 0 and covariance matrix $\begin{bmatrix} \tau_1^2 & \tau_{12} \\ \tau_{12} & \tau_2^2 \end{bmatrix}$.

The following code allows the treatment effect to vary across blocks, using a random coefficient regression model as specified in equation (2). As in the basic partially nested design, the REPEATED statement specifies estimating separate residual variances for the two treatment groups. The REPEATED statement can be omitted if the residual variances are assumed to be the same for the two groups.

```
proc mixed data=blocked noclprint;
  class trtname block cluster student;
  model y = trt / ddfm = sat solution cl;
  random intercept trt / subject=block type=un;
  random intercept / group=trtname subject=cluster(trtname block);
  repeated student / group=trtname;
  parms (15) (10) (50) (0) (50) (200) (200)/ hold = 4;
```

Two RANDOM statements are used for this model. The first RANDOM statement fits the random coefficient regression model for the treatment effect. The TYPE=UN structure for the variance components specifies that three variance parameters [$\tau_1^2 = \text{Var}(a_k)$; $\tau_2^2 = \text{Var}(b_k)$; $\tau_{12} = \text{Cov}(a_k, b_k)$] are to be estimated. The second RANDOM statement specifies the structure of the cluster-level variance components. The fourth variance parameter is the term for σ_{vA}^2 , and the PARMS statement fixes that variance component at 0.

Output 2 shows the covariance parameter and fixed effects estimates from this code. Note that because there are 15 blocks for the random coefficient regression model, the estimated treatment effect has 14 degrees of freedom. The LOG file associated with Output 2 gives the same warning found in Display 1, stating that the G matrix is not positive definite. This message is always provided for an analysis of a partially nested RCT, because the PARMS statement

fixes σ_{vA}^2 at 0. For the blocked design, however, the estimate of the covariance matrix $\begin{bmatrix} \tau_1^2 & \tau_{12} \\ \tau_{12} & \tau_2^2 \end{bmatrix}$ should be positive

definite, and this can be checked by calculating the determinant of the estimated matrix. For our example, the determinant is $\hat{\tau}_1^2 \hat{\tau}_2^2 - \hat{\tau}_{12}^2 = (14.7989)(17.0641) - (14.7973)^2 = 33.57$, so the estimated matrix is indeed positive definite. For other data sets, however, the estimated matrix might not be positive definite. If that occurs, a different model for the site-specific treatment effects should be considered. A simpler model for the site-specific treatment effects might have different intercepts (a_k) but specify the same treatment effect at all sites ($b_k = 0$). The RANDOM statement for the site-specific effects in this simpler model would therefore be

```
random intercept / subject=block;
```

Covariance Parameter Estimates			
Cov Parm	Subject	Group	Estimate
UN(1,1)	block		14.7989
UN(1,2)	block		14.7973
UN(2,2)	block		17.0641
Intercept	cluster(trtname*block)	trtname Control	0
Intercept	cluster(trtname*block)	trtname Treatment	33.3690
student		trtname Control	164.56
student		trtname Treatment	169.59

Solution for Fixed Effects								
Effect	Estimate	Standard Error	DF	t Value	Pr > t	Alpha	Lower	Upper
Intercept	96.3891	1.4435	14	66.78	<.0001	0.05	93.2932	99.4851
trt	8.0314	2.1160	14	3.80	0.0020	0.05	3.4930	12.5698

Output 2. Output for Blocked Partially Nested Design.

CLUSTERED RCTS WITH DIFFERENT HIERARCHICAL STRUCTURES

In a clustered RCT, clusters are randomly assigned to the treatment conditions. Consider Example 4, in which classrooms are randomly assigned to treatment A or treatment B. All of the students in a particular classroom receive the same treatment under this randomization, so that students are nested in classrooms for both treatment A and treatment B. In the schools assigned to treatment B, however, the participating students receive the intervention in a small group setting. In treatment A, therefore, students are directly nested in classrooms. Treatment B, however, has an additional layer of nesting, with students nested in tutor groups which are in turn nested in classrooms.

The key features of the design are as follows:

- Potential participants in the experiment are clustered in units such as schools, classrooms, or hospitals. To prevent confusion with the multiple levels of clustering, we refer to these clusters as the level-3 units.
- Randomly assign half of the level-3 units to treatment A and the other half to treatment B. As in a regular clustered RCT, all individuals in a level-3 unit receive the same treatment.
- The protocol in treatment B is administered in a group setting. Within each level-3 unit that is randomized to treatment B, the individuals (level-1 units) are assigned to sub-clusters (level-2 units) where the intervention is performed. For example, the students in the classrooms in treatment B receive tutoring in small groups, and the tutor groups form the level-2 units. In treatment B, level-1 units are nested in level-2 units, which are nested in level-3 units.
- The individuals in treatment A have no additional clustering. In treatment A, the level-1 units are nested directly in the level-3 units.

The model can be written as follows. Let y_{kij} represent the outcome for student j in cluster i of classroom k : as before, we set $i = 0$ for all students receiving treatment A. Let $T_{kij} = 1$ if student (kij) is assigned to treatment B, and $T_{kij} = 0$ if student (kij) is assigned to treatment A. Note that because of the cluster randomization of the level-3 units, T_{kij} takes on the same value (0 or 1) for all students in the same classroom. The model is:

$$y_{kij} = \alpha + \beta T_{kij} + c_k + v_{ki}T_{kij} + e_{kij}. \quad (3)$$

The parameter β is the overall treatment effect, and c_k is the random term with mean 0 for the effect of classroom (level-3 unit) k . We allow the variance of c_k to differ in treatments A and B: $Var(c_k) = \sigma_{cA}^2$ for classrooms randomized to treatment A, and $Var(c_k) = \sigma_{cB}^2$ for classrooms randomized to treatment B. The random variables v_{ki} and e_{kij} are the random terms at the cluster and student levels. As before, we assume that v_{ki} has mean 0 and variance σ_{vB}^2 if $T_{kij} = 1$, v_{ki} has mean 0 and variance $\sigma_{vA}^2 = 0$ if $T_{kij} = 0$, e_{kij} has mean 0 and variance σ_{eB}^2 if $T_{kij} = 1$, and e_{kij} has mean 0 and variance σ_{eA}^2 if $T_{kij} = 0$.

The following code can be used to estimate the parameters in this model:

```
proc mixed data=classroom_rand;
  class trtname class student;
  model y = trt / ddfm = sat solution;
  random intercept / group=trtname subject=class(trtname);
  random intercept / group=trtname subject=cluster(class trtname);
  repeated student / group=trtname;
  parms (20) (20) (0) (50) (200) (200)/ hold = 3;
```

The first RANDOM statement estimates the level-3 variance components. These are allowed to differ for the two treatment groups, because the asymmetric treatment structure may also affect the classroom-level variance component. The second RANDOM statement estimates the level-2 variance component in treatment B only. The level-2 variance component in treatment A is fixed at 0 in the PARMS statement. Note that the code used in this section can be easily adapted to a situation in which there are level-2 units in both treatment groups, but these have different variances. For example, tutor groups of size 2 might be formed for both treatments, but the tutor groups might meet once a week in treatment A and daily in treatment B. One might expect higher homogeneity in the tutor groups that meet daily. This variant can be analyzed in PROC MIXED by using the code above, but removing the HOLD option in the PARMS statement so that the level-2 variability is estimated in both treatment groups.

Output 3 presents the output for this example. There are 30 classrooms randomly assigned to each treatment condition (60 classrooms in all), and 12 eligible students in each classroom. In treatment B, the students are assigned to 3 tutor groups, each with 4 students.

Covariance Parameter Estimates			
Cov Parm	Subject	Group	Estimate
Intercept	class(trtname)	trtname Control	72.5516
Intercept	class(trtname)	trtname Treatment	31.3900
Intercept	cluster(trtname*class)	trtname Control	0
Intercept	cluster(trtname*class)	trtname Treatment	44.7733
student		trtname Control	229.23
student		trtname Treatment	178.79

Solution for Fixed Effects					
Effect	Estimate	Standard Error	DF	t Value	Pr > t
Intercept	101.72	1.7479	29	58.19	<.0001
trt	0.8078	2.2573	55.8	0.36	0.7218

Output 3. Output for Clustered RCT with Different Hierarchical Structures.

Note that for this design, the observed variability across the classroom means in treatment B will absorb the additional variability caused by the presence of the tutor groups. Thus, a similar estimate of the treatment effect can be obtained by using a model that contains only the highest level of clustering (in this example, the classrooms). This result occurs because in multi-stage clustered designs, the variance of the treatment effect is driven by the highest level of clustering. The estimated treatment effect is 0.8078 and its standard error is 2.2573 for this example—exactly the same as in the full analysis above—if the following code is used:

```
proc mixed data= classroom_rand;
  class trtname class subjid;
  model y = trt / ddfm = sat solution;
  random intercept / group=trtname subject=class(trtname);
```

The two analyses produce the same treatment effect and standard error for this example because each treatment, classroom, and tutor group has the same number of students. With unbalanced data, the results may differ slightly for the two analyses. If only the highest level of clustering is included in the model, however, it is important to allow the variance among the level-3 units to differ for the treatment groups, so that the variance in the treatment B group will include the additional variability caused by the level-2 clustering. The full analysis, with all variance components, is needed if information about the sources of variability is desired.

CONCLUSION

In this paper, we have used PROC MIXED and PROC GLIMMIX to analyze data from experiments where individuals are grouped into clusters as part of the protocol for one of the treatments, but individuals are not grouped into clusters for the other treatment. The RANDOM statement with the GROUP = option can be used to account for the different hierarchical structures in the treatment groups. The HOLD option of the PARMS statement then forces specified variance components to be zero.

In this paper, we have concentrated on fitting different hierarchical structures in the context of having two treatment groups, but all of the models and code are easily extended to multiple treatment groups. This can be done by declaring the variable describing treatment membership to be a categorical variable in the CLASS statement, as discussed in the section “Including Covariates or Multiple Treatment Groups.”

All of the examples in this paper have been constructed using balanced data, but the code is general and works for unbalanced as well as balanced data.

REFERENCES

- Bauer, D.J., Sterba, S.K., and Hallfors, D.D. (2008). Evaluating Group-Based Interventions when Control Participants are Ungrouped. *Multivariate Behavioral Research*, 43: 210-236.
- Donner, A. and Klar, N. (2004). *Design and Analysis of Cluster Randomized Trials in Health Research*. London: Arnold.
- Goldstein, H. (2011). *Multilevel Statistical Models*, 4th edition. Hoboken, NJ: Wiley.
- Kiernan, K., Tao, J., and Gibbs, P. (2012). Tips and Strategies for Mixed Modeling with SAS/STAT® Procedures. *Proceedings of the SAS® Global Forum 2012 Conference*, Cary, NC: SAS Institute Inc. Available at support.sas.com/resources/papers/proceedings12/332-2012.pdf.
- SAS Institute, Inc. (2011). *SAS/STAT® 9.3 User's Guide: The MIXED Procedure (Chapter)*. Cary, NC: SAS Institute, Inc.

ACKNOWLEDGMENTS

This research was partially supported by the Institute of Education Sciences under contract number ED-IES-12-D-0015.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Name: Sharon L. Lohr
Organization: Westat
Address: 1600 Research Blvd.
City, State ZIP: Rockville, MD 20850
Work Phone: (301) 738-3512
Fax: (301) 294-2034
Email: sharonlohr@westat.com
Web: www.westat.com

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.