

Using PROC MIXED in Hierarchical Linear Models: Examples from two- and three-level school-effect analysis, and meta-analysis research

Sawako Suzuki, DePaul University, Chicago
Ching-Fan Sheu, DePaul University, Chicago

ABSTRACT

The study presents useful examples of fitting hierarchical linear models using the PROC MIXED statistical procedure in the SAS system. Hierarchical linear models are quite common in social science studies, in particular educational research, due to naturally occurring hierarchies or clusters (e.g., students belong to classes which are nested in schools). Despite their prevalence, the SAS PROC MIXED does not seem to be fully recognized of its usefulness in analyzing these models. The current paper discusses the advantages of fitting the hierarchical linear models to multilevel data sets and the convenience of conducting such analysis with PROC MIXED. Examples from two- and three-level school-effects analysis, and meta-analysis research are introduced. Particular focus will be on practical usage of the program: how the program scripts are constructed in relation to the model, and how to interpret the output in the context of the research question.

INTRODUCTION

Hierarchical linear models are common in social science research. In educational studies, for example, students belong to classrooms nested in schools, which are in turn clustered within school districts, and so forth. Similarly, clinical trials are hierarchical in nature, with repeated measures of patients being the first level and each individual being the second. Meta-analysis can be considered multilevel as well (Kalaian & Raudenbush, 1996). The observations (first level) are nested within studies (second level).

Despite the prevalence of hierarchical data structure, classical analysis ignored such structure for many years, partly due to the underdevelopment of statistical models (Plewis, 1997). The recently developed multilevel linear models offer researchers methods to increase accuracy and flexibility in analyzing multilevel data.

There are several advantages of fitting multilevel linear models to hierarchically structured data (Raudenbush, 1993). First, both continuous and categorical variables can be specified to have random effects. Variability can be partitioned at each level, which becomes an important process when accounting for dependency due to clustering effects. In addition, independent variables or covariates can be included in the model at different levels. For example, predictors pertaining to the client (e.g., age, gender, previous medical history) as well as information regarding the clinic in which clients are nested can be included in the model at each level. Moreover, the collected data can be unbalanced at any level, and theoretically, higher levels can be added without limit.

The present tutorial demonstrates fitting hierarchical linear models using the MIXED procedure in SAS. Unfortunately, SAS PROC MIXED does not seem to be fully recognized of its usefulness in analyzing these models (for example, Kreft, de Leeuw, and van der Leeden, 1994). Our attempt is to provide the social scientists with an alternative choice to some computer software programs, such as BMDP-5V,

GENMOD, HLM, ML3, VARCL, when analyzing hierarchical data. Because the SAS system is a generalized statistical environment available to many institutions, using SAS PROC MIXED is a convenient solution to many researchers. Moreover, as Singer (1998) points out, SAS PROC MIXED is especially attractive for its ability to run various data management procedures and mixed-effects analysis, all in one single statistical package.

The current paper presents useful examples of fitting hierarchical linear models using SAS PROC MIXED. Examples from three common social science research are introduced: two- and three-level school-effect analysis, and meta-analysis on dichotomous data. The emphasis of this tutorial is on the practical usage of the program, such as the way SAS codes are constructed in relation to the model. The interpretation of the output in the context of the research question is illustrated as well.

TWO-LEVEL SCHOOL-EFFECT ANALYSIS

THE DATA

The data were collected from the Television School and Family Smoking Prevention and Cessation Project which tested independent and combined effects of various programs designed to promote smoking resistance and cessation (Flay et al., 1989). For illustrating purposes, Hedeker Gibbons and Flay (1994)¹ focused on a subset of the full data set; specifically, data from 28 Los Angeles schools which were randomly assigned to one of the four program conditions: (a) a social-resistance classroom curriculum (CC), (b) a television intervention (TV), (c) both CC and TV curriculums, and (d) a no treatment control group. Namely, the subset data consist of three levels: 1,600 students (level 1) from 135 classrooms (level 2) nested within 28 schools (level 3). The predictors at each level are: pretest scores (PRETEST) at level 1 (individual level), and CC, TV at level 3 (school level). Moreover, the number of observations within each group is not equal, with a range of 1 to 13 classrooms per school and 1 to 28 students per classroom.

The students were pretested in January 1986 and were given a posttest in April of the same year, immediately following the intervention. The test, administered twice before and after the intervention, was a seven-item questionnaire used to assess student knowledge about tobacco use and related health issues. The main research question is to investigate whether the various program conditions and the pretest scores can successfully predict the postintervention test scores.

Hedeker et al. (1994) illustrate a random-effects regression model analysis using SAS IML. The syntax for SAS PROC IML used in the article added up to multiple pages of SAS codes. Therefore, we will replicate Hedeker's (1994) findings using PROC MIXED, which is a less costly syntax to develop and run. We begin our analysis with two-level

¹ Raw data are available on the web at <http://www.uic.edu/~hedeker/mix.html>.

models – the pupils nested in classrooms – before adding the third level (i.e., schools).

A. UNCONDITIONAL MEANS MODEL

THE MODEL

The unconditional means model expresses the student-level outcome Y_{ij} by combining two linked models: one at the student level (level 1) and another at the classroom level (level 2). The model at level 1 expresses a student's outcome as the sum of the intercept for the student's classroom and a random error term associated with each individual. At level 2, the classroom intercept is expressed as a sum of the grand mean and sequences of random deviations from such mean. Combined together, this multilevel model becomes:

$$Y_{ij} = \gamma_{00} + u_{0j} + r_{ij}$$

where

$$u_{0j} \sim N(0, \tau_{00}) \text{ and } r_{ij} \sim N(0, \sigma^2)$$

Y_{ijk} is the i th student in the j th classroom

THE SYNTAX

```
PROC MIXED NOCLPRINT NOITPRINT COVTEST;
  CLASS classrm;
  MODEL posttest = / SOLUTION;
  RANDOM intercept / SUBJECT=classrm;
RUN;
```

The PROC MIXED statement includes three options, NOCLPRINT, NOITPRINT, and COVTEST. NOCLPRINT and NOITPRINT suppress the printing of information at the CLASS level and of the iteration history, respectively. COVTEST provides you with the hypothesis testing of the variance and covariance components. NOCLPRINT and NOITPRINT options are included here merely for space-saving reasons. Moreover, the variable, *classrm*, is declared in the CLASS statement because it does not contain quantitative information.

The MODEL and RANDOM statements together specify the model we are running. Whereas the MODEL statement includes the fixed-effect components, the RANDOM statement contains the random effects. The above syntax expresses that the outcome, *posttest*, is modeled by a fixed intercept (which is implied in the MODEL statement), a random intercept clustered by classrooms ("SUBJECT=classrm"), and a random error (which is implied in the RANDOM statement). Furthermore, the SOLUTION option in the MODEL statement is a way to ask SAS to print the estimates for the fixed effects.

THE OUTPUT

Covariance Parameter Estimates (REML)					
Cov Parm	Subject	Estimate	Std Error	Z	Pr> Z
INTERCEPT	CLASSRM	0.1972	0.0458	4.31	0.0001
Residual		1.7253	0.0638	27.05	0.0001

Model Fitting Information for POSTTEST

Description	Value
Observations	1600.000
Res Log Likelihood	-2762.97

Akaike's Information Criterion	-2764.97
Schwarz's Bayesian Criterion	-2770.35
-2 Res Log Likelihood	5525.938
Solution for Fixed Effects	

Effect	Estimate	Std Error	DF	t	Pr> t
INTERCEPT	2.6178	0.0523	134	50.08	0.0001

The 'Covariance Parameter Estimates (REML)' section in the outcome presents the random effects in the model. For this model, the estimated τ_{00} is 0.1972 and the estimated σ^2 is 1.7253. Hypothesis testing of these estimates reveals that both of these values significantly differ from zero ($p < .001$). Therefore, the results suggest that the classrooms do differ in their posttest scores and that there are even more variation among students within classrooms.

The next 'Model Fitting Information for POSTTEST' portion provides values which can be used to examine the model's goodness of fit. It is useful in comparing multiple models with identical fixed effects but different random effects (Littell et al., 1996). The two criteria most likely to be useful are the AIC (Akaike's Information Criterion) and the SBC (Schwarz's Bayesian Criterion). Larger values of these criteria suggest a better fitting model.

The last 'Solution for Fixed Effects' section includes the fixed-effects portion of the model. The estimated classroom effect of 2.6178 refers to the average classroom-level posttest scores within the sampled classroom pool. All of these results will prove useful as a baseline for latter comparisons with other models.

B. INCLUDING PREDICTORS

We will now include the classroom level predictors, CC, TV, and CCTV. These experimental conditions were randomly assigned to schools; however, we will nonetheless consider them as classroom-level predictors here because they were administered at the classroom level. These variables were dummy coded as 0 or 1 depending upon whether the treatment was absent or present. For example, the control group would be coded as 0 in both CC and TV, whereas the group receiving both treatments would be coded as 1 under both variables. Moreover, CCTV is the interaction term of CC and TV.

By including the classroom predictors, we are now expressing the individual outcome as a function of the treatment to which the classroom was assigned. Compared to the previous unconditional model, this model is conditional on the fixed effects of the treatments. It can be written as:

$$Y_{ij} = \gamma_{00} + \gamma_{01} CC_j + \gamma_{02} TV_j + \gamma_{03} CCTV_j + u_{0j} + r_{ij}$$

$$\text{where } u_{0j} \sim N(0, \tau_{00}) \text{ and } r_{ij} \sim N(0, \sigma^2)$$

THE SYNTAX

The only difference from the earlier syntax is the addition of the fixed effects, cc, tv, and cctv (interaction term) in the MODEL statement. In addition, the DDFM=BW option in the MODEL statement requests SAS to use the "between/within" method in computing the denominator degrees of freedom for tests of fixed effects.

```
PROC MIXED NOCLPRINT NOITPRINT COVTEST;
  CLASS classrm;
  MODEL posttest = cc tv cctv / SOLUTION
    DDFM=BW;
  RANDOM intercept / SUBJECT=classrm;
RUN;
```

THE OUTPUT

Covariance Parameter Estimates (REML)					
Cov Parm	Subject	Estimate	Std Error	Z	Pr> Z
INTERCEPT	CLASSRM	0.1437	0.0389	3.69	0.0002
Residual		1.7261	0.0638	27.07	0.0001

Model Fitting Information for POSTTEST

Description	Value
Observations	1600.000
Res Log Likelihood	-2754.82
Akaike's Information Criterion	-2756.82
Schwarz's Bayesian Criterion	-2762.19
-2 Res Log Likelihood	5509.636

Solution for Fixed Effects

Effect	Estimate	Std Error	DF	t	Pr> t
INTERCEPT	2.3406	0.0939	131	24.92	0.0001
CC	0.5881	0.1357	131	4.34	0.0001
TV	0.1173	0.1337	131	0.88	0.3820
CCTV	-0.2434	0.1921	131	-1.27	0.2073

Tests of Fixed Effects

Source	NDF	DDF	Type III F	Pr > F
CC	1	131	18.79	0.0001
TV	1	131	0.77	0.3820
CCTV	1	131	1.61	0.2073

The additional 'Tests of Fixed Effects' portion of the outcome provides hypothesis testing for the fixed effects. This section can be suppressed by including a NOTEST option in the MODEL statement. For space-saving purposes, we will not print this portion for the following models.

The estimated intercept value of 2.3406 in the 'Solution for Fixed Effects' section refers to γ_{00} , the classroom mean posttest scores in the control group. The estimates for other experimental conditions refer to γ_{01} , γ_{02} , and γ_{03} , and each present the relationship between mean posttest scores and the experimental conditions. For example, the estimated value of 0.5881 for the CC condition implies that, on average, the students in the CC-conditioned classrooms score 0.5881 points higher than the control group. The standard error of 0.14 for this value yields an observed t -statistic of 4.34 ($p < .001$), revealing the significant effect of the CC condition on the average posttest scores. Moreover, the hypothesis testing suggests that neither the TV condition nor the interaction term had a significant effect on the mean posttest scores.

Finally, we can look at the 'Covariance Parameter Estimates (REML)' section in comparison with the previous unconditional model. Since the current model is conditional on the predictors, the variance components presented here have different meanings than those in the earlier unconditional model. We can see that, whereas the residual component (variance within classrooms) remained almost unchanged, the classroom intercepts component (variance between classrooms) decreased notably. The reduced

value indicates that some of the variance between classrooms in the mean posttest scores was accounted for the predictors (CC, TV, CCTV).

C. RANDOM INTERCEPT AND SLOPE

THE MODEL

The student level predictor is the *pretest*. By adding this level-1 predictor, not only are we predicting the outcome as a function of the individuals' pretest scores, but also specifying that the relationship between the outcome and the pretest scores may vary across classrooms. In other words, we are adding both fixed and random effects. The model now has intercepts and slopes that vary across classrooms.

$$Y_{ij} = \gamma_{00} + \gamma_{01}CC_j + \gamma_{02}TV_j + \gamma_{03}CCTV_j + \gamma_{10}PRETEST_{ij} + \gamma_{11}CC_jPRETEST_{ij} + \gamma_{12}TV_jPRETEST_{ij} + \gamma_{13}CCTV_jPRETEST_{ij} + u_{0j} + u_{1j}PRETEST_{ij} + r_{ij}$$

$$\text{where } r_{ij} \sim N(0, \sigma^2) \text{ and } \begin{matrix} u_{0j} & 0 & \tau_{00} & \tau_{01} \\ u_{1j} & \sim N & 0 & , \tau_{10} & \tau_{11} \end{matrix}$$

THE SYNTAX

Note that the *pretest* variable is included in both MODEL and RANDOM statements. The MODEL statement contains five fixed effects (i.e., an intercept and fixed slopes for pretest, cc, tv, and cctv). Moreover, there are three random effects expressed under the RANDOM statement (i.e., an intercept, a slope for pretest, and r_{ij} , the variation within-classroom across students.) Furthermore, the TYPE=UN option in the RANDOM statement specifies an *unstructured* variance-covariance matrix for the intercepts and slopes.

```
PROC MIXED NOCLPRINT COVTEST NOITPRINT;
  CLASS classrm;
  MODEL posttest = pretest cc tv cctv /
    SOLUTION DDFM=BW NOTEST;
  RANDOM intercept pretest /
    TYPE=UN SUBJECT=classrm;
RUN;
```

THE OUTPUT

Covariance Parameter Estimates (REML)					
Cov Parm	Subject	Estimate	Std Error	Z	Pr> Z
UN(1,1)	CLASSRM	0.0179	0.0596	0.30	0.7639
UN(2,1)	CLASSRM	0.0133	0.0226	0.59	0.5543
UN(2,2)	CLASSRM	0.0062	0.0107	0.58	0.5623
Residual		1.5926	0.0605	26.33	0.0001

Model Fitting Information for POSTTEST

Description	Value
Observations	1600.000
Res Log Likelihood	-2686.53
Akaike's Information Criterion	-2690.53
Schwarz's Bayesian Criterion	-2701.28
-2 Res Log Likelihood	5373.060
Null Model LRT Chi-Square	24.8491
Null Model LRT DF	3.0000
Null Model LRT P-Value	0.0000

Solution for Fixed Effects					
Effect	Estimate	Std Error	DF	t	Pr> t
INTERCEPT	1.6907	0.0972	131	17.40	0.0001
PRETEST	0.2983	0.0271	1464	11.00	0.0001
CC	0.6196	0.1200	131	5.16	0.0001
TV	0.1474	0.1183	131	1.25	0.2149
CCTV	-0.2092	0.1690	131	-1.24	0.2178

The outcome reveals three fixed effects (intercept, pretest, cc), which significantly differ from zero ($p < .001$). As with the previous model, this suggests that the students in the CC-conditioned classroom report higher average posttest scores. Since the TV and CCTV estimates do not significantly differ from zero, we can summarize the fixed-effects portion of the model as:

Posttest scores (control group)
 $= 1.6907 + 0.2983 \times (\text{Pretest Score})$
 Posttest scores (CC condition)
 $= 2.3103 + 0.2983 \times (\text{Pretest Score})$

The estimated values of the random effects in the REML section indicate that the random slopes do not significantly differ from each other. The variance component for slopes is only 0.0062, which does not differ from zero ($p = .56$). Moreover, the covariance component for intercepts and slopes is also very small (0.0133) ($p = .55$). Therefore, a reduced model that does not contain slopes varying across classrooms may be suggested.

THE SYNTAX

The reduced model includes the same fixed effects as above, but the random effect is reduced to contain only the intercept.

```
PROC MIXED NOCLPRINT COVTEST NOITPRINT;
  CLASS classrm;
  MODEL posttest = pretest cc tv cctv /
    SOLUTION DDFM=BW NOTEST;
  RANDOM intercept / SUBJECT=classrm;
RUN;
```

THE OUTPUT

Covariance Parameter Estimates (REML)					
Cov Parm	Subject	Estimate	Std Error	Z	Pr> Z
INTERCEPT	CLASSRM	0.0950	0.0307	3.09	0.0020
Residual		1.6036	0.0592	27.08	0.0001

Model Fitting Information for POSTTEST	
Description	Value
Observations	1600.000
Res Log Likelihood	-2688.92
Akaike's Information Criterion	-2690.92
Schwarz's Bayesian Criterion	-2696.29
-2 Res Log Likelihood	5377.841

Solution for Fixed Effects					
Effect	Estimate	Std Error	DF	t	Pr> t
INTERCEPT	1.6788	0.1002	131	16.76	0.0001
PRETEST	0.3108	0.0258	1464	12.03	0.0001
CC	0.6323	0.1209	131	5.23	0.0001
TV	0.1570	0.1189	131	1.32	0.1892
CCTV	-0.2715	0.1710	131	-1.59	0.1147

Referring to the model fitting information provided in the two outcomes, we can compare the AIC, SBC, and the -2LL (-2 Res Log Likelihood) values.

	AIC	SBC	-2LL
random intercepts and slopes	-2690.53	-2701.28	5373.060
random intercepts	-2690.92	-2696.29	5377.841

As discussed earlier, larger values of AIC and SBC suggest a better fitting model. However, in the above case, the AIC and SBC values suggest opposite directions. The difference in the -2LL values can test the null hypothesis that the two models do not differ from each other using the χ^2 distribution. The observed difference of 4.781 on 4 degrees of freedom fails to reject the null hypothesis. Therefore, we can safely conclude that adding the random slopes do not significantly improve the model.

THREE-LEVEL SCHOOL-EFFECT ANALYSIS

THE MODEL

We will extend the previous model to include a third level using the same data set.

(a) Fixed Effects

The level-1 predictor (PRETEST) and the level-3 predictors (CC, TV, CCTV) are included in the model. The experimental conditions are predictors at the school level, because each school was randomly assigned to one of the four conditions: control, CC (classroom curriculum), TV (television program), both CC and TV. We are now expressing the student outcome as a function of the individual's pretest score and of the treatment to which his or her school was assigned.

(b) Random Effects

This 3-level model expresses the student-level outcome by combining three linked models: one at the student level (level 1), one at the classroom level (level 2), and one at the school level (level 3). At level 1, the individual's postintervention scores are expressed as a sum of the student's classroom intercept and a random error term associated with each individual. At level 2, the classroom intercept is expressed as a sum of the student's school intercept and random deviations among classrooms. Finally, at level 3, the school intercept is expressed as a sum of the grand mean and sequences of random deviations from such mean.

(c) Mixed Effects

Combined together, this multilevel model becomes:

$$Y_{ijk} = \beta_0 \text{ (grand average)} + \beta_1 \text{ PRETEST}_i + \beta_2 \text{ CC}_k + \beta_3 \text{ TV}_k + \beta_4 \text{ CCTV}_k + \varepsilon_k + \varepsilon_{j(k)} + \varepsilon_{i(j(k))}$$

where

Y_{ijk} is the i th student in the j th classroom of the k th

school,

$\varepsilon_{i(j(k))}$ is the random individual variance within

classrooms nested in schools,

$\varepsilon_{j(k)}$ is the random classroom variance nested in

schools, and

ε_k is the random school variance.

We will not include random slopes for each of the four predictors, because our preliminary analysis indicated that the 'goodness of fit' is better without.

THE SYNTAX

```
PROC MIXED NOCLPRINT COVTEST NOITPRINT;
  CLASS classrm school;
  MODEL posttest = pretest cc tv cctv /
    SOLUTION DDFM=BW NOTEST;
  RANDOM intercept / SUBJECT=school;
  RANDOM intercept / SUBJECT=classrm(school);
RUN;
```

THE OUTPUT

Covariance Parameter Estimates (REML)

Cov Parm	Subject	Estimate	Std Error	Z	Pr> Z
INTERCEPT	SCHOOL	0.0386	0.0253	1.52	0.127
INTERCEPT	CLASSRM(SCHOOL)	0.0647	0.0286	2.26	0.024
Residual		1.6023	0.0591	27.10	0.000

Model Fitting Information for POSTTEST

Description	Value
Observations	1600.000
Res Log Likelihood	-2686.67
Akaike's Information Criterion	-2689.67
Schwarz's Bayesian Criterion	-2697.73
-2 Res Log Likelihood	5373.335

Solution for Fixed Effects

Effect	Estimate	Std Error	DF	t	Pr> t
INTERCEPT	1.7020	0.1254	24	13.57	0.0001
PRETEST	0.3054	0.0259	1571	11.79	0.0001
CC	0.6413	0.1609	24	3.99	0.0005
TV	0.1821	0.1572	24	1.16	0.2582
CCTV	-0.3309	0.2245	24	-1.47	0.1535

(a) Fixed Effects

The fixed-effects component of the outcome ("Solution for Fixed Effects") reveals that INTERCEPT, PRETEST, and CC differ significantly from zero ($p < .001$). This suggests that the students in the CC-conditioned schools, on average, report higher posttest scores. Since the TV and CCTV estimates do not significantly differ from zero, we can summarize the fixed-effects portion of the model as:

Posttest scores (TV, CCTV, or control group)
 $= 1.7020 + 0.3054 * (\text{pretest score})$
 Posttest scores (CC group)
 $= 1.7020 + 0.6413 + 0.3054 * (\text{pretest score})$

(b) Random Effects

The first parameter estimate, INTERCEPT SCHOOL (0.0386, s.e.=0.0253), under the "Covariance Parameter Estimates" represents the variance component between schools. The following INTERCEPT CLASSRM(SCHOOL) estimate (0.0647, s.e.=0.0286) indicates the variance between classrooms nested in schools. Lastly, the Residual (1.6023, s.e.=0.0591) is the random individual differences within classrooms nested in schools.

While there are significant differences in the mean posttest scores across classrooms ($p < .05$), the differences between schools on the postintervention test scores are negligible

($p = .13$) after the previous classroom variances have been accounted for.

The SAS system uses the REML (Restricted Maximum Likelihood) method by default. Other methods can be specified with the METHOD option under the PROC statement. (For further details, refer to Littell et al., 1996.)

SUMMARY

The above random-effects regression model is capable of looking at individual characteristics taking into account the effects of clustering. In other words, the current model fits the data better compared to the ordinary regression analysis, because the multilevel model incorporates the individual level information and attends to its dependency to higher-level groupings as well. Were we to run an ordinary regression analysis at the individual level, it may over- or underestimate the effects of experimental conditions due to its negligence of clustering effects. Moreover, an ordinary regression analysis run at the cluster level (classroom or school in the present case) will also be insensitive to the nature of the data, because it will fail to incorporate individual level information. It is clear that fitting hierarchical linear models to data with naturally occurring hierarchies has many advantages.

META-ANALYSIS

SAS PROC MIXED is also useful for analyzing data for meta-analytical research. The data structure can be considered as multilevel, where the responses are the first level unit nested in studies. However, the usefulness of the MIXED procedure is only recently beginning to be recognized in this area (Wang and Bushman, 1999).

The current tutorial examines meta-analysis of dichotomous data. Haddock, Rindskopf, and Shadish (1998) contend that many researchers inappropriately employ correlations or standardized mean difference statistics to estimate effect sizes for meta-analytic research on dichotomous data. Alternatively, they propose the use of odds ratios (or the logarithm thereof) to compute proper effect sizes in such cases. While this method has been common among other disciplines such as epidemiology and medicine, its use among psychological and educational research has been minimal. Therefore, we are motivated to illustrate the new technique – the application of mixed-effects models (including both fixed and random effects) on odds ratio – using the MIXED procedure.

THE ORIGINAL DATA

Twenty-four (24) studies on addiction treatment (Haddock et al., 1998) were entered into the meta-analysis. The studies were categorized into three groups, depending on the type of addiction they surveyed: alcohol ($n=12$), substance abuse ($n=5$), or smoking cessation ($n=7$). The data structure of the studies were fourfold tables; it involved treatment and control group and the response measures were the number of subjects who succeeded (or failed) to overcome the addiction with (or without) treatment. Hence, the raw data appear as below:²

² See "Using odds ratios as effect sizes for meta-analysis of dichotomous data: A primer on methods and issues". *Psychological Methods*, 3 (3), 339-353, for the full data set.

Study	Treatment		Control	
	Success	Failure	Success	Failure
1	4	3	1	4
2	5	14	6	12
3	5	3	3	6
.	.	.	.	.
.	.	.	.	.

In their analyses, Haddock et al. (1998) use the odds ratio as the dependent measure. They reason that using odds ratio is statistically convenient because the normal assumption can be met. In a few words, the odds ratio combines a row of information into a single number, and can be calculated as below:

$$\text{Odds Ratio} = \frac{(\text{Treatment} \& \text{Success}) \times (\text{Control} \& \text{Failure})}{(\text{Treatment} \& \text{Failure}) \times (\text{Control} \& \text{Success})}$$

Moreover, the variance of an odds ratio can be obtained by taking the sum of the reciprocals of the four frequencies.

However, the normal approximation of odds ratio does not occur without limitation. The normality assumption may be violated in cases with small sample size or when zero (0) counts are common in the collected data. For these reasons, we propose that a general linear mixed model, which does not rely on the normality assumption, is a more appropriate model to fit the data.

Therefore, we omit the replication of Haddock et al.'s (1998) earlier models, and instead focus on the demonstration of the random-effects logistic regression model – a model discussed but not illustrated in the original study.

GENERALIZED LINEAR MIXED MODEL

Just as generalized linear models extend linear models to non-normal data, generalized linear mixed models extend linear mixed models to non-normal data. In SAS environment, the GLIMMIX macro³ able PROC MIXED to fit various generalized linear mixed models to the available data.

In our current example, we are modeling the logit of success probability within each (treatment or control) group of a study. Therefore, our model is referred to a linear logistic model with random effects. Such model can be expressed as (Collett, 1991):

$$\text{logit}(\vartheta_i) = \gamma_0 + \gamma_1 X_i + \delta_i$$

where

- ϑ_i , the true response probability, is a random variable with an expected value of p_i ,
- δ_i is the random effect.

Since the response measures from each study will consist of two probabilities – one from the treatment group and another from the control group – nested within each study, the original data have to be rearranged as follows:

study	addctn	trt	favor	unfavor
1	alch	trt	4	3
1	alch	cntl	1	4
2	alch	trt	5	14
2	alch	cntl	6	12
3	alch	trt	5	3
3	alch	cntl	3	6
.	.	.	.	.
.	.	.	.	.

Our example can be modeled as the following:

$$\text{logit}[\pi_{ij} / (1 - \pi_{ij})] = [\gamma_0 + \gamma_1(\text{Treatment})_{ij} + \gamma_2(\text{Alcohol})_i + \gamma_3(\text{Smoking})_i + \gamma_4(\text{Alcohol})_i * (\text{Treatment})_{ij} + \gamma_5(\text{Smoking})_i * (\text{Treatment})_{ij}] + [u_i]$$

where

- π_{ij} is the number of favorable outcomes within the i th group in the j th study,
- Treatment is coded 0 for control group and 1 for treatment group.

As mentioned earlier, the response measure is the logit of a ratio of two variables, the number of favorable outcomes within a treatment (or control) group and the total number of subjects within the same group. Notice that the third addiction type, substance abuse, is omitted from the model, because it is linearly dependent on two other categories and intercept. For illustration purposes, we embraced the fixed effects with the first bracket and the random effects with the second.

THE SYNTAX

In correspondence to the aforementioned data arrangement, the SAS codes should begin with an INPUT statement similar to the following:⁴

```
INPUT study drug $ trt $ favor unfavor;
      n = favor + unfavor;
%INCLUDE 'glimm612.sas';
%GLIMMIX(DATA=meta,
  PROCOPT=METHOD=REML,
  STMTS=%STR(
    CLASS study addctn trt;
    MODEL y/n = trt addctn addctn*trt /
      SOLUTION;
    RANDOM intercept / SUBJECT=study SOLUTION;
  ),
  ERROR=BINOMIAL,
  LINK=LOGIT
);
RUN;
```

The %INCLUDE statement specifies the location of and the file name containing the GLIMMIX macro. The subsequent %GLIMMIX command initiates the procedure and includes statements between the parentheses which specify the procedure. The PROC MIXED statements (e.g., CLASS, MODEL, and RANDOM statements) belong in the parentheses under STMTS=%STR. These commands are quite similar to the PROC MIXED statements we used in our

³ The GLIMMIX macro is offered on the web at <http://www.sas.com/techsup/download/stat/>. GLIMMIX macro for versions up to 8 are available.

⁴ When the response measure is the logit of a ratio of two variables, the convergence of the algorithms may become difficult. A more consistent convergence can be obtained by reexpressing the data to contain 1's (favorable) and 0's (unfavorable), and then using this single response variable (Littell et al., 1996, *SAS system for Mixed Models*, p. 440). With this procedure, we obtained results that were very similar to those presented herein.

earlier examples, with one major difference being that, for binomial data, the response variable must be given as a ratio of two variables. As discussed earlier, this ratio is the number of successes (numerator) divided by the total number of observations (denominator). In our specific case, the variable y stands for the number of subjects who successfully overcame addiction, and n refers to the total number of subjects within the given treatment or control group.

The PROC OPT, ERROR, and LINK statements can specify the variance component estimation procedure, the error distribution, and the link function, respectively. Further information regarding statement options in GLIMMIX are given in the GLIMMIX macro available in SAS Online Samples or on the web.

THE OUTPUT

Class Level Information

Class	Levels	Values
STUDY	24	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24
ADDCTN	3	alch smok subs
TRT	2	cntl trt

Covariance Parameter Estimates

Cov Parm	Subject	Estimate
INTERCEPT	STUDY	0.69552506

GLIMMIX Model Statistics

Description	Value
Deviance	43.9908
Scaled Deviance	25.6150
Pearson Chi-Square	42.4792
Scaled Pearson Chi-Square	24.7348
Extra-Dispersion Scale	1.7174

Parameter Estimates

Effect	ADDCTN	TRT	Estimate	Std Error	t	Pr> t
INTERCEPT			0.7943	0.4265	1.86	0.0766
TRT		cntl	-1.1338	0.3313	-3.42	0.0026
TRT		trt	0.0000	.	.	.
ADDCTN	alch		-0.9043	0.5182	-1.75	0.0956
ADDCTN	smok		-1.4859	0.5411	-2.75	0.0121
ADDCTN	subs		0.0000	.	.	.
ADDCTN*TRT	alch	cntl	1.1063	0.3949	2.80	0.0107
ADDCTN*TRT	alch	trt	0.0000	.	.	.
ADDCTN*TRT	smok	cntl	0.7828	0.3640	2.15	0.0433
ADDCTN*TRT	smok	trt	0.0000	.	.	.
ADDCTN*TRT	subs	cntl	0.0000	.	.	.
ADDCTN*TRT	subs	trt	0.0000	.	.	.

Tests of Fixed Effects

Source	NDF	DDF	Type III F	Pr > F
TRT	1	21	12.81	0.0018
ADDCTN	2	21	2.55	0.1023
ADDCTN*TRT	2	21	3.93	0.0356

The variance component between studies is 0.6955. Furthermore, according to the given parameter estimates, the fixed-effect portion of the model can be described as:

Substance Abuse

(Control) = 0.7940 - 1.1326

(Treated) = 0.7940

(Average Effect) = (Treated) - (Control) = 1.1326

Alcohol

(Control) = 0.7940 - 1.1326 - 0.9039 + 1.1051

(Treated) = 0.7940 - 0.9039

(Average Effect) = (Treated) - (Control) = 0.0275

Smoking Cessation

(Control) = 0.794 - 1.1326 - 1.4855 + 0.7817

(Treated) = 0.794 - 1.4855

(Average Effect) = (Treated) - (Control) = 0.3509

Our findings are similar to those reported by Haddock et al. (1998). The figures are not identical, because of the difference in model formulation. Whereas the original study modeled on log odds ratio, we modeled ours on binary data. In addition, different computer software was used – Haddock et al. used HLM.

The result suggests that, the effect size improves under the treatment condition in substance abuse studies more than in other types of studies. Overall, the effects of the three study categories alone could not explain the different outcomes between studies ($p=ns$); however, the treatment conditions could be accounted for the difference among the two treatment groups ($p<.05$). Further, a significant interaction effect between the type of studies and the treatment condition was observed ($p<.05$).

SUMMARY

The current example presented meta-analysis on dichotomous data, using SAS PROC MIXED. As Haddock et al. (1998) assert, many meta-analysts are not familiar with statistical methods appropriate for dichotomous data. Furthermore, fitting random-effects model to dichotomous data is still new in the field of psychology and education.

The above procedure – fitting general linear mixed models (logistic linear mixed model, in our case) – can be easily carried out in SAS. The GLIMMIX macro, which is available on the web, able PROC MIXED to fit generalized linear mixed models. Hence, we believe that this rare tutorial would prove useful among meta-analysts using SAS.

CONCLUSION

Fitting multilevel linear models using SAS PROC MIXED was illustrated using three examples: two-level and three-level school-effect analysis, and meta-analysis research. In the school-effect analysis, we began with two-level analysis (pupil and classroom) and then added a third level (schools). The example showed the advantages of being able to partition variance at different levels – one of the strongest benefits of fitting hierarchical linear models. Unlike ordinary regression models, hierarchical linear models agree with the data structure and can account for the dependency due to clustering effects. For the meta-analysis of dichotomous data, the GLIMMIX macro was used to enable PROC MIXED to fit the generalized linear mixed model. Specifically, we demonstrated to fit the linear logistic model with random-effects. In either case, the merits of fitting multilevel linear models were apparent. SAS PROC MIXED proved to be a useful and simple procedure which facilitates researchers to fit hierarchical linear models to multilevel data.

REFERENCES

Collett, D. (1991). *Modelling Binary Data*. London: Chapman & Hall.

Flay, B. R., Brannon, B. R., Johnson, C. A., Hansen, W., B., Ulene, A. L., Whitney-Saltiel, D. A., Gleason, L. R., Sussman, S., Gavin, M., Glowacz, K. M., Sobol, D. F., & Spiegel, D. C. (1989). The Television, School and Family Smoking Cessation and Prevention Project: I. Theoretical basis and program development. *Preventive Medicine*, 76, 585-607.

Haddock, C. K., Rindskopf, D., and Shadish, W. R. (1998). Using odds ratios as effect sizes for meta-analysis of dichotomous data: A primer on methods and issues. *Psychological Methods*, 3 (3), 339-353.

Hedeker, D., Gibbons R. D., & Flay, B. R. (1994). Random-effects regression models for clustered data with an example from smoking prevention research. *Journal of Consulting and Clinical Psychology*, 62 (4), 757-765.

Kalaian, H. A., & Raudenbush, S. W. (1996). A multivariate mixed linear model for meta-analysis. *Psychological Methods*, 1, 227-235.

Kreft, I., de Leeuw, J., & van der Leeden, R. (1994). Review of five multilevel analysis programs: BMDP-5V, GENMOD, HLM, ML3, VARCL. *The American Statistician*, 48 (4), 324-335.

Littell, R. C., Milliken, G. A., Stroup, W. W., & Wolfinger, R. D. (1996). *SAS System for Mixed Models*. Cary, NC: SAS Institute, Inc.

Plewis, I. (1997). *Statistics in Education*. London: Arnold.

Raudenbush, S. W. (1993). Hierarchical linear models and experimental design. In Lynne, E. K. (ed.) *Applied Analysis of Variance in Behavioral Science*. New York: M. Dekker.

Singer, J. D. (1998). Using SAS PROC MIXED to fit multilevel models, hierarchical models, and individual growth models. *Journal of Educational and Behavioral Statistics*, 23(4), 323-355.

Wang, M. C., & Bushman, B. J. (1999). *Integrating Results through Meta-Analytic Review Using SAS® Software*. Cary, NC: SAS Institute, Inc.

Ching-Fan Sheu
Department of Psychology
DePaul University
2219 N. Kenmore Ave.
Chicago, IL 60614-3522
E-mail: csheu@condor.depaul.edu

ACKNOWLEDGMENTS

We thank Rebecca White for her comments on the preliminary draft of this paper.

CONTACT INFORMATION

Sawako Suzuki
Graduate School of Education
University of California, Berkeley
1600 Tolman Hall
Berkeley, CA 94720-1670