

## Adjusting for Unequal Selection Probability in Multilevel Models: A Comparison of Software Packages

Kim Chantala

Chirayath Suchindran

Carolina Population Center, UNC at Chapel Hill

Carolina Population Center, UNC at Chapel Hill

### Abstract

Most surveys collect data using complex sampling plans that involve selecting both clusters and individuals with unequal probability of selection. Research in using multilevel modeling techniques to analyze such data is relatively new. Often sampling weights based on probabilities of selecting individuals are used to estimate population-based models. However, sampling weights used for estimating multilevel models (MLM) need to be constructed differently than weights used for population-average models. This paper compares the capabilities of MLWIN, MPLUS, LISREL, PROC MIXED (SAS), and gllamm (Stata) for estimating MLM using data collected with a complex sampling plan. We illustrate how sampling weights need to be constructed for estimating MLM with these software packages. Finally, we contrast the results from these packages using data collected with a complex sampling plan.

**Keywords:** multilevel models, sample weights, software

### 1. Introduction

Population surveys commonly use complex sampling plans to collect information on individuals. These sampling plans often involve sampling both clusters and individuals with unequal probability of selection. Special analysis techniques are needed to compute unbiased point estimates and variances when analyzing data collected with these methods. The purpose of this paper is to demonstrate the use of recent advances in statistical methods to easily adjust for the sampling design characteristics by using appropriate sampling weights to estimate multilevel models.

Extensive research in estimating single-level (population-average or marginal) models from complex survey data has resulted in the availability of several software packages (SUDAAN, svy commands in Stata, and SURVEYFREQ,

SURVEYREG, etc. in SAS) that use appropriate design-based analysis techniques for complex survey data. However, research in analysis techniques for estimating multilevel models from complex survey data is quite recent (Pfefferman (1998), Stapleton (2002), Asparouhov (2006)). Not only has this research resulted in new methods for incorporating sampling weights into multilevel models, but has emphasized an important point often overlooked by both analysts and providers of the survey data: the sampling weights used for multilevel analysis need to be constructed differently than the sampling weights used for single-level analysis. The sampling weight used in estimating single-level models is computed as the inverse of the probability that the individual was selected from the population and represents the number of individuals in the population that are likely to answer the survey in a manner similar to the individual interviewed. This type of sampling weight is commonly distributed with data from population surveys. Ideally, estimation of the multilevel models requires scaling weights at each level of sampling. Public use data may not provide this information.

We first review available software packages for estimating multilevel models that incorporate sampling weights in analysis. Next, we discuss how the sampling weights for multilevel models need to be constructed for these packages. We conclude by using data from the National Longitudinal Study of Adolescent Health (Add Health) to provide examples of estimating a multilevel model with a few of these packages.

### 2. MLM Software for Complex Surveys

Several of the commercially available software packages allow analysts to use sampling weights when estimating Structural Equation Models (SEM) and Multilevel Models (MLM). As shown in Table 1, the SEM software packages include MPLUS, LISREL, and the Stata user written program gllamm. Except for MIXED and NLMIXED from SAS, all of these packages have been designed to analyze data collected with a complex sampling plan.

Table 1. General Information about Software Packages.

	SEM Analysis	MLM Analysis	Adjust for Clustering	Adjust for Stratification	Subpopulation Analysis
MPLUS 4.1	•	•	•	•	•
LISREL 8.8	•	•	•	•	
GLLAMM (Stata 9)	•	•	•		
MLWIN 2.02		•	•		
HLM 6.0		•	•		
MIXED (SAS 9.1)		•	•		
NLMIXED (SAS 9.1)		•	•		

In addition to allowing sampling weights for estimating single-level models, some of these software packages also allow users to specify sampling weights designed for estimating multilevel models (table 2). Because these weights need to be constructed differently than sampling weights used for single-level models, users should make sure the weights they are using are scaled properly for the particular software package being used for MLM analysis.

While MIXED (SAS 9.1) does allow users to specify a single weight, the weight is not expected to be a sampling weight, but a weight designed to be inversely proportional to the variability of the observations. Hence, users must be quite cautious in using MIXED when analyzing data collected with a complex sampling plan. SAS also provides a separate package, NLMIXED, for estimating nonlinear multilevel models. Although there is no weight statement available with NLMIXED, special weighting procedures have been implemented through a SAS macro to adjust for the sampling design (Grilli and Pratesi, 2004).

Table 3 lists the types of MLM analyses available from these packages that allow users to use multilevel sampling weights. The vendors of MPLUS, MLWIN, LISREL, and HLM report that the most recent versions of their software packages all produce comparable results when estimating models from complex survey data.

### 3. Data used in examples

Examples in this paper use data from the National Longitudinal Study of Adolescents (Add Health). Add Health is a longitudinal study of adolescents listed on grade 7-12 enrollment rosters for the 1994-1995 academic year. A sample of 80 high schools and 52 middle schools was chosen with unequal probability of selection. Incorporating systematic sampling methods and implicit stratification in the study design ensured that these schools were representative of US schools with respect to region of country, location (urban, suburban, rural), school type (private, public, parochial), percentage of students who were white, and school size. Administrators at each school were asked to fill out a

Table 2. Implementation of Sampling Weights

	Allow MLM Sampling Weights	Method for Scaling MLM Sampling Weights	Responsibility for Scaling MLM Sampling Weights
MPLUS 4.1	•	Asparouhov (2006)	User
LISREL 8.8	•	Pfefferman, et al (1998)	LISREL
GLLAMM (Stata 9)	•	Pfefferman, et al (1998)	User
MLWIN 2.02	•	Pfefferman, et al (1998)	User or MLWIN
HLM 6.0	•	Pfefferman, et al (1998)	HLM
MIXED (SAS 9.1)		Unknown	User
NLMIXED (SAS 9.1)		Grilli, and Pratesi, (2004)	User

Table 3. Types of Multilevel analysis that allow use of sampling weights during estimation.

Software Package	Outcome Variable					
	Normal	Binary	Poisson	Multinomial	Categorical	Ordered Categorical
MPLUS 4.1	•	•	•		•	•
LISREL 8.7	•					
GLLAMM (Stata 9)	•	•	•		•	•
MLWIN 2.02	•	•	•		•	•
HLM 6.0	•	•	•		•	•
MIXED (SAS 8.2)	•					
NLMIXED (SAS 8.2)		•	•			

special survey that captured characteristics of the school. Add Health has collected four panels of data on adolescents: In-School (1994), the Wave I In-home Survey (1995), the Wave II In-home Survey (1996), and the Wave III In-home Survey (2001). The In-School survey included all students from sampled schools who were in attendance on the day the survey was administered. The Wave I In-home survey selected students from the enrollment rosters of the 132 schools with unequal probability of selection. Several special over-sampled groups were also recruited for the Wave I interview. These include the core sample (roughly equal-sized samples), purposively selected schools (all students selected from certain schools), non-genetic supplements (Black adolescents whose parents were college graduates, adolescents whose race was Cuban, Puerto Rican, or Chinese.), the disabled sample, and the genetic supplement (biologically related adolescents, non-related adolescents living together). The Wave II and Wave III samples were selected from the Wave I respondents. In addition to providing sampling weights that are designed for estimating population-average models, weights for schools and adolescents are available for each wave of the Add Health data. The descriptive statistics for these weight components are shown in table 4.

**4. Sampling Weights used in Multilevel Analysis**

If the probability of being selected is related to the outcome variable even after conditioning on the model covariates, the sampling process is informative and it will be important to adjust the estimates for the sampling process. Some analysts decide to adjust for the sampling design by adding covariates to the model rather than using sampling weights to adjust for the sampling process. Because of the large number of variables involved in the sampling

process, this can add unwanted complexity to the model and interfere with the scientific reasons for conducting the analysis.

Because no one way of scaling the weights for multilevel modeling has been widely accepted, the method of scaling the weights can be different for different MLM software packages. First we describe the weight information that is needed to construct multilevel sampling weights and then describe two of the methods most commonly implemented by developers of the MLM software packages.

**4.1 Constructing Multilevel Sampling Weights**

Multilevel sampling weights can be constructed using the weight components from each level of sampling. The weight component for a specific level of sampling provides information about how many individuals within a sampling unit are represented by each selected individual. In general, a set of weight components is typically computed for each stage of sampling in a population survey and then used in calculating the weights for fitting both population-average and multilevel models. For example, the weight components computed for each wave of Add Health are:

*Primary Sampling Unit (PSU) or level 2 weight component.* Each school in the sample has a weight equal to the number of schools in the sampling frame represented by that school. It is computed as the inverse of the probability that school j was selected from the sampling frame:

$$psu\_wt_j = 1/\{\text{Pr}(\text{school } j \text{ selected})\}.$$

Table 4. Available weight components designed for use with the Wave I, II, or III In-Home Data.

Statistical Measures	Level-2 (School) weight component ( <i>psu_wt<sub>ij</sub></i> )	Level-1 (Adolescent) weight component ( <i>fsu_wt<sub>ij</sub></i> )			
		Wave I In-Home <sup>1</sup>	Wave II In-Home <sup>2</sup>	Wave III In-Home <sup>1</sup>	Wave III In-Home <sup>2</sup>
Variable Name	schwt1	w1_wc	w2_wc	w3_2_wc	w3_wc
N	132	18,924	13,568	14,322	10,828
Minimum	35.89	1.00	1.00	1.00	1.00
Median	151.80	3.31	4.05	4.12	4.77
Maximum	4,170.13	119.98	119.98	119.98	119.98
Mean	243.80	6.17	7.18	8.12	8.96
Variance	171,803.92	54.88	70.86	102.20	119.40
Sum	32,181.22	116,706.97	97,361.25	116,266.04	97,018.23

<sup>1</sup> The Target population is adolescents enrolled in grades 7-12 for the 1994-1995 academic year.

<sup>2</sup> The Target population is adolescents enrolled in grades 7-11 for the 1994-1995 academic year.

*Final Sampling Unit (FSU) or level 1 weight component.* Each student selected from school *j* has a sampling weight that is equal to the number of students within school *j* represented by that student. It is computed as the inverse of the probability that student *i* was selected given that school *j* was selected:

$$fsu\_wt_{ij} = 1/\{\text{Pr}(\text{student } i \text{ selected}|\text{school } j \text{ selected})\}.$$

Non-response and post-stratification adjustments are often made to the weight components to ensure that they give the best estimate possible. Analysts wanting to construct sampling weights for estimating multilevel models need to have the weight components from each stage of sampling available. If only weights designed for estimating population-average models are available for each stage of sampling, the analyst can estimate the level 1 weight component for a two-stage design as follows. Since the commonly distributed sampling weight used in estimating population-average or single-level models is computed as:

$$W_{i,j} = psu\_wt_j \times fsu\_wt_{ij}$$

analysts can estimate *fsu\_wt<sub>ij</sub>* by dividing *W<sub>i,j</sub>* by *psu\_wt<sub>j</sub>*.

#### 4.2 Scaling Weights For Two-Level Analysis

Scaling the weights for analysis reduces small sample bias while retaining consistency. The parameter estimates are invariant to scalar multiples of the weights. Pfefferman et al (1998) describes the following method of scaling sampling weights for use

in two-level models. The method we discuss here is recommended for scaling the level 1 weight for the MLM analysis when informative sampling methods are used for selecting units at both levels of sampling (Method 2 from Pfefferman, 1998). The scaled level 1 weight for each unit *i* sampled from cluster *j* is computed by dividing each level 1 weight (*fsu\_wt<sub>ij</sub>*) by the average of all level 1 weight components in cluster *j*:

$$pw2lev1_{ij} = \frac{fsu\_wt_{ij}}{\left( \frac{\sum_i^{n_j} fsu\_wt_{ij}}{n_j} \right)}$$

Here *n<sub>j</sub>* is the number of units selected from cluster *j*. MLWIN, LISREL and HLM will automatically do Method 2 scaling for the user. In MLWIN, the weights are assumed to be independent of random effects. The level 2 weight requires no scaling. The developers of gllamm recommended that the analyst select the best performing method from Pfefferman et al (1998) that most closely matches the sample design used to collect the survey data being analyzed. Method 2 is the best choice for the Add Health data and also allows comparison of estimates from these software packages.

MPLUS uses weights at both levels of sampling to construct one composite sampling weight for the two-level analysis. Sampling weights for used with MPLUS to estimate our two-level model were constructed using MPML Method A (Asparouhov, 2004) recommended when both informative sampling methods are used for selecting units at both levels of

sampling. Method A weight construction involves dividing the product of the level 1 and level 2 weight components by the average of the level 1 weight components for units sampled from cluster j:

$$mp\_wt_{i,j} = \frac{fsu\_wt_{i,j} * psu\_wt_j}{\left( \frac{\sum_i^{n_j} fsu\_wt_{i,j}}{n_j} \right)} = pw2lev1_{i,j} * psu\_wt_j$$

This is just the product of the Method 2 scaled level 1 weight and the level 2 weight component. The analyst must create this weight for MPLUS. During estimation MPLUS separates the level 1 and level 2 weights internally to produce scaling method 2.

**5. Two-level Model with Continuous Outcome**

Data for the example used to illustrate the SEM and MLM software packages comes from the School Administrator Survey and the Wave I In-home survey of the Add Health data. This example will estimate body mass index of the students in a school from the hours spent watching TV or using computers and availability of a school recreation center. Information on the availability of an on-site school recreation center (variable RC\_S) was provided by each school. Each adolescent answered questions used to compute percentile body mass index (BMIPCT) and hours watching TV or playing video or computer games during the past week (HR\_WATCH). Our example will fit a MLM with a level for the school and a level for the adolescent. The algebraic formulas describing the model and assumptions appear below.

*Student-level model (Within or Level 1):*

$$(BMIPCT)_{ij} = \{ \beta_{0j} + \beta_{1j}(HR\_WATCH_{ij}) \} + e_{ij}$$

where:

$$E(e_{ij}) = 0 \text{ and } Var(e_{ij}) = \sigma^2$$

*School-level Model (Between or Level 2):*

$$\beta_{0j} = \gamma_{00} + \gamma_{01}(RC\_S)_j + \delta_{0j}$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11}(RC\_S)_j + \delta_{1j}$$

where:

$$E(\delta_{0j}) = E(\delta_{1j}) = 0,$$

$$Var(\delta_{0j}) = \sigma^2_{\delta 0}, \text{ Var}(\delta_{1j}) = \sigma^2_{\delta 1},$$

$$Cov(\delta_{0j}, \delta_{1j}) = \sigma_{\delta 01}$$

These two equations can be combined to be expressed as:

*Combined Model:*

$$BMIPCT_{ij} = \{ \gamma_{00} + \gamma_{01} * RC\_S_j + \gamma_{10} * HR\_WATCH_{ij} + \gamma_{11} * RC\_S_j * HR\_WATCH_{ij} + \{ \delta_{0j} + \delta_{1j} * HR\_WATCH_{ij} \} + e_{ij}$$

In this example, we incorporate the sample design by using the sampling weights to adjust for unequal probability of selection. We followed Method 2 from Pfefferman et al to scale the weights for level 1 in the MLM analysis. These weights were used for the gllamm run. Because MLWIN and LISREL automatically scale the level 1 weight with the Method 2 technique, we used the unscaled level 1 weight in our analyses.

MPLUS uses weights at both levels of sampling to construct one scaled sampling weight for the two-level analysis. Sampling weights for use with MPLUS two-level model were constructed using MPML Method A.

If the sampling weights are ignored, these packages produce the same results. The results of the estimation using sampling weights with each package are given in Table 5. MPLUS and gllamm produce nearly identical results on both parameter estimates and standard errors. Except for the standard error of  $\sigma^2$ , MLWIN gave results close to those of MPLUS and gllamm. Estimates from LISREL are surprisingly different for both parameters and standard errors. We contacted the developers at SSI Scientific concerning these differences and they were able to identify a problem with the implementation of the sampling weights with LISREL 8.8 and earlier versions. We received a special updated version, and the estimates from this version matched the estimates from MLWIN 2.02. The developers of LISREL are working to determine why the standard error estimates differ from the standard errors of gllamm and MPLUS are correct.

Table 5. Results from estimation of 2-level model estimated with sampling weights.

Parameter in 2-Level Model	MPLUS 4.1 Estimate (S.E)	LISREL 8.8 Estimate (S.E.)	MLWIN 2.02 Estimate (S.E.)	GLLAMM Estimate (S.E.)
<i>Fixed Effects</i>				
$\gamma_{00}$ (Intercept for $\beta_{0j}$ )	60.22 (1.10)	59.27 (0.75)	60.28 (1.17)	60.22 (1.10)
$\gamma_{01}$ (Slope for $\beta_{0j}$ )	-5.48 (1.50)	-3.02 (1.08)	-5.62 (1.65)	-5.48 (1.50)
$\gamma_{10}$ (Intercept for $\beta_{1j}$ )	0.032 (0.022)	0.041 (0.019)	0.030 (0.023)	0.032 (0.022)
$\gamma_{11}$ (Slope for $\beta_{1j}$ )	0.13 (0.031)	0.11 (0.026)	0.130 (0.032)	0.13 (0.031)
<i>Random Effects</i>				
$\sigma^2_{\delta_0}$ (Var ( $\delta_{0j}$ ))	19.31 (6.97)	8.96 (1.87)	20.18 (6.04)	19.32 (6.97)
$\sigma^2_{\delta_1}$ (Var ( $\delta_{1j}$ ))	0.003 (0.002)	0.003 (0.001)	0.003 (0.001)	0.003 (0.002)
$\sigma_{12}$ (Cov ( $\delta_{0j}, \delta_{1j}$ ))	-0.081 (0.098)	-0.057 (0.036)	-0.091 (0.071)	-0.079 (0.097)
$\sigma^2$ (Var ( $e_{ij}$ ))	788.79 (17.02)	793.87 (75.64)	786.37 (86.62)	788.81 (17.02)

### 6. Subpopulation Analysis

Analyzing subpopulations from a sample survey should be done by analyzing observations from everyone who completed the survey rather than by deleting the observations outside the subpopulation of interest. If the population was resampled using the same sampling plan, a different number of people in the subpopulation would be obtained each time. Thus, the subpopulation sample size is a random variable that affects variance estimation. An appropriate contribution due to the random nature of the sample size should be included when estimating variance and standard errors than those obtained by analyzing a subset of the data.

Software packages designed for estimating population-average models from complex survey data usually provide analysts special options to define the subpopulation of interest but still analyze the full data set so necessary design information can be determined for variance estimation. For survey software packages estimating marginal models that do not have a subpopulation statement, an analyst can set the sample weight for observations outside the subpopulation to have a value of zero and estimate the model using the full set of data (Sribney, 1999). This removes the contribution of the observations outside the subpopulation to the point estimates, but leaves the sample design structure intact for proper estimation of variances and standard errors. Some packages become numerically unstable when sampling weights have a value of zero while others do not allow a sampling weight of zero. For these packages, setting the sampling weight for members

outside the subpopulation to a very small value should allow parameter estimates to be very close to those that would be computed with a zero weight and allow the variance and standard errors to be adjusted for sample size variability. MLWIN documentation also suggests setting the weight to zero for observations that the analyst desires to remove from the analysis.

Of the software packages we are testing, only MPLUS provides a subpopulation statement allowing users to define the population of interest. The MPLUS developers report that the zero-weight approach would work when the subpopulations were at the primary sampling unit level (level 2 in our model), but would produce biased standard errors when at lower levels. The developers of MPLUS and LISREL also suggest using a multiple-group analysis for analyzing subpopulations defined at any level to produced unbiased parameter estimates and standard errors. This method simultaneously fits separate models for the desired subpopulation and the group outside the subpopulation. This allows means, slopes, variance, and covariance parameters to differ for the two groups.

Because of the simplicity, we used the zero-weight technique to explore subpopulation analysis to provide an example showing how well this technique will work for multilevel methods and provide analysts with a simple method of analyzing subpopulations with software packages that do not provide special options for subpopulation analysis. We also included an example of the multiple-group

analysis for subpopulations defined at level 1. Because of the previously mentioned problems with weighted estimation using LISREL 8.8, we have omitted these results pending further investigation. We ran tests with subpopulations defined first at level 2 and then at level 1. LISREL, MLWIN, and gllamm all allow the analyst to control the level 1 and level 2 weights. Since MPLUS requires one composite weight used in analysis rather than a separate weight at each level of sampling, the analyst does not have control over the defining the subpopulation at each level with the zero-weight technique.

**6.1 Subpopulation Analysis at Level 2**

We defined the subpopulation to be adolescents attending urban schools and re-fit the continuous model from the first example. Only 40 of the 132 schools are classified as being located in an urban area. Table 6 shows the results of our analysis. First, we subset the data to include only the observations from participants sampled from urban schools. Analysis of this reduced data set is expected to give correct parameter estimates, but incorrect estimates

of the standard errors. Because the level 1 sampling weights are conditional on the level 2 weights, the level 1 sampling weight should be set to 0 when the level 2 weight is 0. We refit the model setting both level 1 and level 2 sampling weights to 0.00001 for MLWIN and gllamm. Since MPLUS uses a combined weight we cannot independently control setting the level 1 and 2 weights to zero. We set the combined weight to 0.00001 for MPLUS one of the MPLUS runs. We also ran MPLUS using the subpopulation statement. The parameter estimates for MPLUS, MLWIN, and gllamm are all quite close to the results from analyzing the subset of the data. The standard errors from MPLUS and MLWIN using the zero-weight approach are all slightly larger than for the estimates for the subset data. The results computed with the subpopulation statement in MPLUS had identical parameter estimates but larger standard errors than the results from the zero-weight approach. However, gllamm computed smaller standard errors for estimates using the near-zero weight subpopulation method than for the subset data.

Table 6. Subpopulation Analysis on Urban schools (level 2).

	NOT RECOMMENDED: Delete non-urban schools			Subpopulation Statement	Alter Sampling weight for non-urban subpopulation		
	Subset Data to only Urban Schools				Composite weight =0.00001	Level 2 weight=0.00001 Level 1 weight=0.00001	
	MPLUS 4.0 Estimate (S.E)	MLWIN 2.02 Estimate (S.E.)	GLLAMM Estimate (S.E.)	MPLUS 4.1 Estimate (S.E.)	MPLUS 4.1 Estimate (S.E)	MLWIN 2.02 Estimate (S.E.)	GLLAMM Estimate (S.E.)
<i>Fixed Effects</i>							
$\gamma_{00}$	59.648 (2.412)	59.786 (2.165)	59.669 (2.443)	59.678 (2.438)	59.678 (2.428)	59.910 (2.260)	59.669 (2.421)
$\gamma_{01}$	-4.117 (2.660)	-4.234 (2.447)	-4.133 (2.694)	-4.147 (2.685)	-4.147 (2.674)	-4.310 (2.534)	-4.133 (2.670)
$\gamma_{10}$	-0.004 (0.067)	-0.006 (0.058)	-0.004 (.067)	-0.005 (0.067)	-0.005 (0.067)	-0.011 (0.062)	-0.004 (0.067)
$\gamma_{11}$	0.112 (0.075)	0.112 (0.066)	0.112 (.076)	0.113 (0.076)	0.113 (0.075)	0.114 (0.070)	0.112 (0.075)
<i>Random Effects</i>							
$\sigma^2_{\delta_0}$	15.427 (6.627)	14.752 (5.059)	15.547 (6.782)	15.607 (6.797)	15.606 (6.770)	15.434 (5.679)	15.548 (6.723)
$\sigma^2_{\delta_1}$	0.006 (0.005)	0.005 (0.002)	0.006 (.006)	0.007 (0.006)	0.007 (0.006)	0.007 (0.003)	0.006 (0.006)
$\sigma_{12}$	-0.135 (0.218)	-0.101 (0.120)	-0.136 (0.224)	-0.142 (0.225)	-0.142 (0.224)	-0.137 (0.139)	-0.136 (0.221)
$\sigma^2$	764.007 (24.586)	762.352 (127.285)	763.944 (24.882)	763.761 (24.682)	763.761 (24.587)	761.322 (166.449)	763.945 (24.66)

### 6.3 Subpopulation Analysis at Level 1

The level 1 subpopulation was defined to be boys. First, we subset the data to include only the boys sampled from schools and ran our model from example 1. All schools remained in the analysis. Because of this, we expect the standard errors for the subset data to be quite close to the standard errors for the runs testing near-zero weights to define subpopulations. Estimates from these models are shown in table 7. The estimates from gllamm using only the level 1 near-zero weights showed almost no difference from the results using subset data. All other runs showed variability in estimates for the random effects and the slope for the School-level intercept model ( $\gamma_{01}$ ). MLWIN shows the differences in the fixed parameter estimates. This could be caused by MLWIN automatically standardizing the level 1 sampling weights while gllamm and MPLUS allow the user to control scaling. Thus MLWIN

would have used a different weighting for the boys in the subset analysis versus the zero-weight analysis.

An alternate method to analyze subpopulations is to use a multiple-group analysis to produce unbiased parameter estimates and standard errors. This method simultaneously fits separate models for the desired subpopulation and the group outside the subpopulation. To implement this method, the model is parameterized with separate covariates for the subpopulation of interest and the remainder of the sample. For the boys' subpopulation analysis at level 1, we define two dummy variables to denote each population:

B=0 for girls, B=1 for boys  
G=1 for girls, G=0 for boys

Table 7. Subpopulation Analysis on Level 1.

Parameter	NOT RECOMMENDED Delete girls			Subpopulation Statement	Alter Sampling weight for girls		
	Subset Data to only Boys				Composite weight =0.00001	Level 1 weight=0.00001 for girls, boys weight is unchanged	
	MPLUS 4.1 Estimate (S.E)	MLWIN 2.02 Estimate (S.E.)	GLLAMM Estimate (S.E.)		MPLUS 4.1 Estimate (S.E)	MLWIN 2.02 Estimate (S.E.)	GLLAMM Estimate (S.E.)
<i>Fixed Effects</i>							
$\gamma_{00}$	60.942 (1.364)	61.321 (1.608)	61.032 (1.394)	61.087 (2.098)	61.087 (2.090)	61.439 (2.231)	61.033 (1.395)
$\gamma_{01}$	-5.013 (1.711)	-5.529 (2.018)	-5.144 (1.735)	-5.482 (2.325)	-5.482 (2.316)	-6.136 (2.673)	-5.144 (1.735)
$\gamma_{10}$	0.020 (0.029)	0.011 (0.032)	0.0184 (0.030)	0.015 (0.042)	0.015 (0.042)	0.007 (0.046)	0.0183 (0.030)
$\gamma_{11}$	0.127 (0.042)	0.135 (0.044)	0.129 (0.042)	0.125 (0.050)	0.125 (0.049)	0.134 (0.055)	0.129 (0.042)
<i>Random Effects</i>							
$\sigma^2_{\delta_0}$	18.173 (15.296)	20.693 (11.023)	19.653 (15.876)	56.221 (38.235)	56.221 (38.088)	59.669 (28.122)	19.647 (15.876)
$\sigma^2_{\delta_1}$	.007 (.005)	0.007 (0.003)	0.008 (0.005)	0.027 (0.011)	0.027 (0.011)	0.027 (0.007)	0.008 (0.005)
$\sigma_{12}$	-0.146 (0.220)	-0.182 (0.130)	-0.175 (.229)	-0.875 (0.610)	-0.875 (0.608)	-0.923 (0.422)	-0.175 (0.229)
$\sigma^2$	819.862 (22.871)	814.703 (95.358)	818.573 (22.907)	803.001 (22.268)	803.001 (22.182)	797.602 (90.107)	818.578 (22.903)



For the Boys subpopulation analysis our model becomes:

$$\begin{aligned}
 \text{BMIPCT}_{ij} = & \{ \gamma_{00} * B_{ij} + \gamma_{01} * B_{ij} * RC_{S_j} \\
 & + \gamma_{10} * B_{ij} * HR\_WATCH_{ij} \\
 & + \gamma_{11} * B_{ij} * RC_{S_j} * HR\_WATCH_{ij} \} \\
 & + \{ \delta_{0j} + \delta_{1j} * B_{ij} * HR\_WATCH_{ij} \} + e_{ij} \\
 & + \{ \tau_{00} * G_{ij} + \tau_{01} * G_{ij} * RC_{S_j} \\
 & + \tau_{10} * G_{ij} * HR\_WATCH_{ij} \\
 & + \tau_{11} * G_{ij} * RC_{S_j} * HR\_WATCH_{ij} \} \\
 & + \{ \eta_{0j} + \eta_{1j} * G_{ij} * HR\_WATCH_{ij} \} + \epsilon_{ij}
 \end{aligned}$$

where:

$$E(e_{ij}) = 0, \text{Var}(e_{ij}) = \sigma^2 e_{ij},$$

$$\begin{aligned}
 E(\epsilon_{ij}) &= 0, \text{Var}(\epsilon_{ij}) = \sigma^2 \epsilon_{ij} \\
 E(\delta_{0j}) &= E(\delta_{1j}) = 0, \\
 \text{Var}(\delta_{0j}) &= \sigma^2_{\delta 0}, \text{Var}(\delta_{1j}) = \sigma^2_{\delta 1}, \text{Cov}(\delta_{0j}, \delta_{1j}) = \sigma_{\delta 0} \\
 E(\eta_{0j}) &= E(\eta_{1j}) = 0, \\
 \text{Var}(\eta_{0j}) &= \sigma^2_{\eta 0}, \text{Var}(\eta_{1j}) = \sigma^2_{\eta 1}, \text{Cov}(\eta_{0j}, \eta_{1j}) = \sigma_{\eta 01} \\
 \text{Cov}(\eta_{0j}, \delta_{0j}) &= \text{Cov}(\eta_{1j}, \delta_{0j}) = \text{Cov}(\eta_{0j}, \delta_{1j}) = \text{Cov}(\eta_{1j}, \delta_{1j}) = 0
 \end{aligned}$$

This method was suggested by the developers of both LISREL and MPLUS. However, the developers of LISREL specified that the covariance across the two groups be constrained to be zero. If we wanted to contrast boys and girls, they we would not constrain the covariance terms to zero. However, for our problem estimating parameters for girls is a nuisance and we wish to focus only on the boys. The estimates from this model appear in table 8.

Table 8. Subpopulation Analysis using Multiple Group Method using MLWIN.

Parameter	Subset Data to Only Boys	Boys Results from Two-group Analysis
	MLWIN 2.02 β (S.E)	MLWIN 2.02 β (S.E)
<i>Fixed Effects</i>		
γ <sub>00</sub>	61.321 (1.608)	61.123 (1.575)
γ <sub>01</sub>	-5.529 (2.018)	-5.303 (1.992)
γ <sub>10</sub>	0.011 (0.032)	0.017 (0.031)
γ <sub>11</sub>	0.135 (0.044)	0.129 (0.043)
<i>Random Effects</i>		
σ <sup>2</sup> <sub>δ0</sub>	20.693 (11.023)	20.139 (10.623)
σ <sup>2</sup> <sub>δ1</sub>	0.007 (0.003)	0.007 (0.003)
σ <sub>12</sub>	-0.182 (0.130)	-0.164 (0.123)
σ <sup>2</sup>	814.703 (95.358)	841.662 (97.506)

The two-group analysis results produced parameter estimates identical to the subset analysis, but MLWIN has slightly smaller errors. Overall, these estimates are quite reasonable. We ran these models with the special version of LISREL we received that had fixed the problem with the weights in estimating the parameters in a two-level model and obtained the same estimates as we did with MLWIN 2.02.

### Conclusion

Several software packages have recently incorporated use of sampling weights to adjust for non-response and the design characteristics of complex survey data when estimating structural estimation models and multilevel models. This provides analysts with a

simple method for obtaining unbiased estimates from complex survey data. However, the collectors and distributors of complex survey data need to be aware that sampling weights must be provided for every level of sampling for these weights to be constructed.

Our tests showed that MLWIN, MPLUS and gllamm produce similar results when analyzing the full sample. MLWIN, MPLUS, and LISREL all reached convergence in less than one minute of CPU time. However, the same model took over 6 hours of cpu time with gllamm. This time advantage could be due to approximations that work well for the normal model. Additional models with non-normal outcomes should be run on all packages to determine if the consistency in results remains in these packages

and to monitor differential convergence times for the multilevel software packages.

We explored ways of using near-zero sampling weights to define sample members who are outside of the desired subpopulation as a way of obtaining valid point estimates and standard errors when analyzing survey data with multilevel models. Although this method works well for estimating marginal models, it showed limited promise when estimating subpopulations in multilevel models. Software packages that allow the user to impose constraints on the covariance terms allow users to implement the multiple-group method for subpopulation analysis. Analysts should consider using the multiple-group methods for subpopulation analysis until further development has been to verify methods of estimating multilevel models using subpopulations from complex survey data.

### References

- Asparouhov, T. (2005). Sampling weights in latent variable modeling. *Structural Equation Modeling*, 12, 411-434.
- Asparouhov, T. (2006). General Multilevel Modeling with Sampling Weights. *Communications in Statistics: Theory and Methods*, Volume 35, Number 3, pp. 439-460(22).
- Du Toit, S. et al (2006), LISREL for Windows: MULTILEV User's Guide, Lincolnwood, IL: SSI Scientific Software International.
- Muthen, L.K. and Muthen, B.O. (2006), MPLUS Statistical Analysis With Latent Variables User's Guide Fourth Edition and MPLUS software, Los Angeles, CA: Muthen & Muthen.
- Pfeffermann, D., The Role of Sampling Weights when Modeling Survey Data, (1993) *International Statistical Review*, p 317-337.
- Pfeffermann, D., Skinner, C. J., Holmes D. J, and H. Goldstein, Rasbash, J(1998). Weighting for Unequal Selection Probabilities in Multilevel Models. *JRSS, Series B*, 60, 123-40.
- Rabe-Hesketh, S., Skrondall A., and Pickles, (2004) GLLAMM Manual and GLLAMM software. UC Berkeley Division of Biostatistics Working Paper Series, [www.gllamm.org](http://www.gllamm.org).
- Rasbash J., et al. (2002), A Users Guide to MLWIN and MLWIN software, Centre for Multilevel Modeling Institute of Education University of London.
- SAS Institute. 2004 SAS/STAT 9.1, User's Guide, Cary, NC: SAS Institute Inc.
- Stapleton, L. M., (2002) the Incorporation of Sample Weights Into Multilevel Structural Equation Models, *Structural Equation Modeling*, 9(4), 475-502.
- Sribney, W. (1999), "Comparing RE and PA models", <http://www.stata.com/support/faqs/stat/rep.html> StataCorpwebsite.