HIERARCHICAL MODELS APPLIED TO THE NATIONAL HOUSEHOLD SURVEY ON DRUG ABUSE (NHSDA)

Doug Wright, SAMHSA and Zhiwei Zhang, NORC
Doug Wright, SAMHSA, Rm. 16C-06, 5600 Fishers Lane, Rockville, MD 20857

Key Words: Hierarchical Models, Multilevel Models

I. Background. A fairly comprehensive discussion of the methodology in hierarchical modeling was introduced in Hierarchical Linear Models (1992) by Bryk and Raudenbush. A significant part of that book focused on hierarchical structures in the field of education and the necessity of using an estimation methodology that reflects the structure of the data. A natural question is how that methodology might apply to the drug field. Much of the drug analysis research over the years has utilized simple logistic regression to estimate relationships between drug use and a variety of person-level variables. We wanted to explore how hierarchical models could be applied to the NHSDA data and what the consequences would be of ignoring the hierarchy.

Since its inception, the NHSDA survey has been a multi-phase stratified sample of primary sampling units (counties or groups of counties), segments (blocks or block groups), households, and individuals. The result of this complex nested design is that using traditional techniques of variance estimation that ignore the clustering of sample cases and the use of sample weights tends to overstate the significance of many findings. A number of statistical packages have been developed using Taylor series methods or replication methods, such as SUDAAN (RTI) and WESTVAR (WESTAT), that take this structure into account. There was no widely used software for hierarchical mixed models, however, until HLM. The last few years have seen improvements in the front-ends of the two most popular packages -- HLM (Bryk, Raudenbush, & Congdon, 1996) and Mln (Prosser, Rabash, & Goldstein, 1996). HLM software has been expanded to include nonlinear models (HGLM), and this is helpful since many of the variables of interest in NHSDA are dichotomous, e.g. use of a specific drug. In 1992, SAS Institute also introduced a multilevel analysis routine -- PROC MIXED -- into their statistical package (SAS Institute (1992)).

This paper sets out to explore some of the special circumstances in applying hierarchical models to the NHSDA. We decided to use the data from six oversampled cities from 1991-93 NHSDA. With a sample of approximately 2500 persons per city for each of the three years, we had a total sample of about 45,000 persons. Each of the six cities (Washington DC, Chicago, Miami, New York City, Denver, and Los Angeles) was selected with certainty, and within each city, segments were sampled at the first stage. Within a segment, usually about 10-12 persons age 12 and older were selected. If we wanted to make inference to the collection of the six large cities as typical large cities, then we could consider four levels of hierarchy: city, segment, family, and person. Two issues arose immediately. Were there sufficient number of observations at each hierarchical level to support the analysis and was it necessary to include all four levels of hierarchy in the estimation? By the latter, what we mean is was there sufficient variability at each stage to be necessary for inclusion?

A. **Levels of Hierarchical modeling**. In 1991-93, the NHSDA design typically resulted in one person selected per household, although 2 were selected in a small nonrandom subset of households. With usually only a single observation per household, proper estimates could not be made at that level, and the household level of the hierarchy could not be modeled.

When a stage of variation is omitted from multi-level analysis, its variability is attributed to the next lower level. Since the variability at the household level could not be measured well with the 1991-93 NHSDA, all of the household variability was attributed to the person level. It should be noted that in 1997, the NHSDA design was modified so that every pair of persons in the sample had a probability of selection. This should enhance the possibility of measuring family effects in 1997 and later, although the small number of observations at the household level may still present an estimation concern.

Therefore, for the years 1991-93, we were limited to a 3-level model of cities, segments, and persons. Furthermore, in analyzing the between-level variance contributions for selected variables, it became apparent that the proportion of total variation contributed by the between-city variance was quite small, generally about 5% of the total variation or less. In addition, the number of metropolitan areas -- here, 6 -- was probably too small to obtain valid inferences in the 3-level modeling. For these reasons, we decided to eliminate the city stage of the hierarchy in the analysis and to utilize a 2-level hierarchical model of segments and persons. This analysis treated the segments as the first stage of selection in the combined 6-city population.

B. **Sampling Weights.** The question of whether or not to use the sampling weights in the analysis did not

appear to be a major issue in the modeling for the 6 city data. Since each city was selected with certainty, all the differential selection probabilities took place at the segment and person levels. NHSDA surveys have traditionally oversampled youth age 12-17, blacks, and Hispanics; therefore, the weights for individuals in these groups were smaller than for the rest of the sample. The result of ignoring the weights leads to population estimates that gave relatively more weight to the oversampled groups than they should have - in other words, a city population that reflected proportionally more youth, blacks, and Hispanics than really existed.

One empirical way of determining whether the weights would make a significant difference in the estimation was to compare the weighted and unweighted estimates to see if they differed greatly. We conducted simple multi-level analysis using the final person-level weight and compared this to an unweighted analysis, and obtained similar results. Based on our initial analysis that showed few differences between the weighted person-level results and unweighted estimates, we chose to use the unweighted modeling in our preliminary estimation.

Recent research (Pfeffermann et al) has demonstrated the importance of using weights that represent each stage (both person level and segment level in our data) of the hierarchical sampling structure. However, it didn't appear that the currently-available software incorporated the Pfeffermann weight components correctly to handle the NHSDA data, so we stayed with our original decision not to use the weights in the analysis.

C. **Linear versus non-linear modeling.** While much of the NHSDA data was dichotomous data, analyzing noncontinuous data invited certain difficulties in interpretation. In HLM with continuous data, the components of variance at each stage of the hierarchy had a straight intuitive interpretation that was very useful. The total variance of the dependent variable could be partitioned into a component at each level. By looking at these proportions, one could tell whether the exclusion of a given level of the hierarchy would have a significant effect on the results. In the dichotomous situation, the equations would have needed to be transformed in order to satisfy the usual assumptions of Normality of the residuals and homogeneity of variance at level 1. In this situation, the variance components are no longer in the original scale, but in a log odds scale, so that it becomes more difficult to interpret when a variance component is proportionately "large." For this reason, we focused our initial analysis on some constructed linear scales that did not have this problem.

II. File Construction, Variables, and Sample Sizes. We considered two different focuses: One was on youth, since this is often an interest in prevention activities. The other was on all persons 12 and older. The two different focuses were based on practical considerations of small average number of persons per level-2 unit. Even though the expected segment sample size was 10-12 for persons 12 and older, there was no control on the sample size for youth. Initially, when we explored dichotomous variables like the use of marijuana, many segments did not include any youth, or, if they did contain youth, there were no marijuana users among them. In order to improve the chances of measurement, we considered other more prevalent behaviors, such as smoking cigarettes and drinking alcohol. We also considered expanding the prevalence period from past month to past year in order to increase the probability of finding a user.

Since the sample sizes for youth were so small, we also decided to look at broader areas of geography such as Census tracts. There were, however, both advantages and disadvantages to using the tract level of hierarchy rather than the segment level. One advantage was that there was a variety of data available from the Census at the tract level - more than was available at the segment level. Since tracts are larger than segments, there was a chance that the tract would have larger sample sizes than the segment level. Another advantage was that we could estimate level-2 effects with very little sampling error because of the large Census samples. A disadvantage, however, was that tracts were somewhat larger than segments: the typical tract size is about 100 blocks, or a square of about 10 blocks on each side. If one were interested in immediate neighborhood effects, the tract may be too large in area. Also, since the sample was not selected with tracts as a level of sampling, the likely outcome was that there would be some tracts with larger sample sizes than the 10-12 segment size, but others with relatively small sample sizes.

We calculated the distribution of sample at the tract level for each of the 6 cities. As anticipated, there were many tracts with small samples (306 tracts with 4 or fewer respondents, accounting for about 11% of the total number of tracts. For youth, many of the tracts had no sample at all. The two smallest MSAs, Miami and Denver, had larger than average sample sizes. Therefore, we decided to focus on these two cities for youth. Still, this left a number of tracts with small sample sizes. We decided that since our main interest was in trying to study the phenomenon of multilevel effects in the NHSDA, that we would not unduly compromise this if we were to eliminate all tracts with 7 or fewer youth. Deleting these small tracts resulted in median tract sizes for Miami and Denver of 15 and 12, respectively.

For the 12 and older analysis, we decided that a segment file for the 6 cities combined would suffice. As

with the youth, we eliminated segments with 7 or fewer persons. This resulted in median sample sizes of 12 persons for all six cities and 14 for Miami.

We constructed two continuous dependent variables. The first variable was SUMRKMJ, the risk of using marijuana, based on a summed scale based of 3 indicators: RSKTRYMJ - risk of trying marijuana once or twice, RSKMJOCC - risk of smoking marijuana occasionally, and RSKMJREG - risk of smoking marijuana regularly. Each of these measures was on a Likert scale of 1-4, with a 1 indicating little perceived risk and 4 indicating a great perceived risk. The second variable constructed was SUMRKDIF, the difficulty of obtaining drugs. This variable was equal to the sum of the difficulty of obtaining marijuana and the difficulty of obtaining cocaine. A value of 1 indicated that obtaining the drug was probably impossible, while a value of 5 indicated that obtaining the drug was very easy.

At level 1, the person level, the variables used in the youth analysis were past year marijuana use, lifetime cigarette use, race/ethnicity, age, gender, enrolled in school, moved in past 5 years, and living with both parents. The patterns of (weighted) drug use were quite different for the Miami and Denver. For example, past year marijuana use among youth was 5% in Miami, but 19% in Denver. A number of the other variables (e.g., age, gender, enrollment, mobility, and family status) were similar for the two cities.

For the level-2 analyses in which the tract was the level of analysis, the following variables were included: percent black, percent under 18, percent in poverty, percent of households with a child under 18 and with a female head of household, median housing value, median rent, and median household income.

For all persons 12 and older, at level 1, the variables were the same as those for adolescents with the exception of enrolled in school and living with both parents. At level-2, the segment level, we used percent in poverty, percent of people with associate degrees, percent of persons 16-64 with a work disability, percent of males who were separated or divorced, and percent black.

III. Analysis

A. **Youth** The first model estimated was the model with only a random intercept. Using the simple random effects ANOVA with a main effect and error term at the person level and the tract group level, we modeled SUMRKMJ. Here, $(SUMRKMJ)_{ij} = \gamma_{00} + U_{0j} + R_{ij}$. The estimated intercept was 10.5 (the range of possible scores is from 3 to 12), indicating a fairly high estimated average perception among youth. For Miami, only 5% of the total variation was between tracts and the remainder was within tracts. In Denver, less than .9% of the total variation was between tracts. For this variable, therefore, it appeared that there was not much variation between tracts, and that this level of hierarchy could probably be ignored in the modeling if the primary purpose was to develop predictors of the perceived risk.

Looking at SUMRKDIF for Miami revealed that 10% of the total variation was between tracts. So, relative to SUMRKMJ, there was a lot more variation between neighborhoods. Therefore, the perception of the difficulty of obtaining drugs differed more across neighborhoods than did the perceived risk of using marijuana.

In Denver, the total variance on SUMRKMJ accounted for by the between tract group differences was only 0.87% and the total variance on SUMRKDIF accounted by the between tract differences was 5.0%. Since the numbers for Denver were much smaller than those for Miami, this indicates that Denver was relatively more homogeneous at the tract level than Miami in youth perception of drug use risks and the perception of the difficulties of obtaining drugs.

The average score for Miami was 5.9 (the range was 2 to 10), indicating that, on average, youth thought it was fairly difficult to obtain these drugs.

We decided to focus on the difficulty of obtaining drugs SUMRKDIF for the remaining analyses since a significant portion of the variation was between tracts for this variable.

We plotted SUMRKDIF as a function of a few variables of interest. When we plotted age by lifetime cigarette use for Miami, it was clear that the ease of getting drugs increases with each succeeding year, and for those who have ever smoked a cigarette, the line was higher than for those who have never smoked. We also noticed that for 3 out of 4 lines, the perceived ease of getting drugs dipped at age 13. This could have been due to various factors, and may partly reflect the transition from middle school (grade 8) to high school (grade 9).

Entering age at level-1 in our model, the level 1 model was

$$Y_{ij} = \beta_{oj} + \beta_{1j}(AGE) + r_{ij.}$$

The level 2 model was

$$\beta_{oj} = \gamma_{00} + u_{oj}$$
$$\beta_{1j} = \gamma_{10}$$

Based on our analysis, we constrained the random component around the age slope to 0. A preliminary analysis specified a model where both β_{oj} and β_{1j} were random. Since the variance component u_{1j} for the age slope was only 0.00309 (χ^2 =102.492 with df=98), this suggested that the age effect was largely invariant across the level 2 units; therefore, we assumed a fixed effect for the age covariate.

The choice of the specification of the level 1 coefficients (i.e., as fixed, random, or nonrandomly varying) could have also been based on the reliability estimates from the random-coefficient regression model (Bryk and Raudenbush, 1992:110). In the preliminary analysis specifying a model where both β_{oj} and β_{1j} were random, the reliabilities for β_0 and β_1 were 0.613 and 0.017, respectively. These numbers suggested that a considerable amount of the observed variation in the β_{oj} was potentially explainable. At the same time, since the reliability of the random level 1 coefficient β_{1j} was close to 0.02, this coefficient was treated as fixed.

The combined model was:

$$Y_{ij} = \gamma_{00} + u_{oj} + \gamma_{10}(AGE) + r_{ij}$$

This is different than the usual single level model in that it has random terms at both level 1 and level 2. Both u_{oj} and r_{ij} are random quantities, with expected means equal to zero.

The estimated coefficient for age was .5, indicating that for each year of age, the ease of obtaining drugs was .5 higher. Over the 5 years from 12 to 17, the ease of obtaining drugs increased from 5.9 (fairly difficult) to 8.4 (fairly easy). Age explained 10.0% of the level-1 variance.

The other demographic variables on the file that proved to be significant but explained little of the level-1 variation were whether the person was Hispanic or not and whether the person was black or not. Variables that didn't prove significant were gender and total family income.

Next, we introduced lifetime cigarette use. It's coefficient was 1.1, indicating that use of cigarettes increases the ease of obtaining drugs (the predicted value of SUMRKDIF) by .4 standard deviations. Use of cigarettes explained an additional 2.0% of the level 1 variation - 12.0% of the level-1 variation for both age and cigarette use. This suggested another possible way of looking at the level-1 variation, namely, focusing on the variation that remains once demographic variation has been eliminated since the demographic variables (age, race, gender, etc.) are not subject to intervention, but the cigarette behavior is.

We also wanted to introduce a level-2 predictor for the level-2 intercept. The variable that seemed most related was the percent black. When we introduced this variable, it accounted for approximately 65% of the variation at the tract level (i.e., the between tract variance was reduced from .97 for the model including Age and Cigarette Use to .34 with the addition of Percent Black at level 2). The addition of other variables did not explain much of the remaining variation. A possible explanation of this is that the other tract variables were mainly related to socioeconomic status. What was needed to explain the perceived difficulty of obtaining drugs were variables that were more drug-related.

B. **Adults.** For the dependent variable SUMRKDIF, 13% of the total variation was between segments for the combined 6 cities. The average perceived difficulty of obtaining drug for adults, 6.2 across all 6 cities, was in the 'fairly difficult' range.

Does Age play the same role as it does for youth? A plot of the data was helpful. Two trends were present: the increased ease of obtaining drugs for those age 12-17 and the increased difficulty of getting drugs starting sometime after age 17 and decreasing monotonically as adults age. The plot suggested entering both age and square root of age as independent variables.

When we introduced age, the age/difficulty slope was -.020 and significant, indicating that persons in this age range found it increasingly difficult to obtain drugs. Age explained about 1.0% of the level 1 variation. When we introduced square root of age as well, the total variation explained by both of the age variables was 5.8%.

We constrained the random effects for both slopes to be 0 since there was not sufficient variation left once the effect of the slopes was accounted for. The other demographic variable on the file was 'gender.' When we added this variable and constrained its random effect to 0, the slope was significant and slightly more of the level 1 variation was explained. The male coefficient was .38, which says that a male's perceived difficulty of obtaining drugs was .38 less than a female's.

The addition of 'ever smoked cigarettes' at level 1 was highly significant with a coefficient of .87. Thus, having smoked cigarettes increased the ease of obtaining drugs by almost 1, and explained another 3% of the level-1 variation. The total variation explained by these variables was 8.5%.

Next, we added in a variable at level 2 to explain some of the level 2 segment variation. All of the variables that we had on the file were SES-related variables, and the strongest relationship appeared to be with percent black. The coefficient was .29 and was significant, showing that neighborhoods (tracts) with a high percentage of blacks found it easier to obtain drugs than areas with a low percent black. Percent black explained 17% of the level-2 variation.

IV. Validation of Assumptions Our next step was to take our HLM results and verify them with the use of MLwiN. MLwiN has built in two algorithms that make better estimates of the between-segment and between-tract variation when the level 2 sample sizes are small and when the distribution is not well-behaved, as is true for dichotomous data. (One of the reasons that we had used the HLM software was that it had been

interwoven with examples from the excellent book by Bryk and Raudenbush, and the clear exposition of hierarchical models in that book was very helpful.) Data results showed that the results of the two programs were quite similar.

A. **MCMC estimates.** In small samples, there is uncertainty associated with the estimation of the random parameters. The maximum likelihood methods tend to overestimate precision because they ignore the prior distributions of random parameters by treating them as known for purpose of inference. We thus used the Bayesian modeling using two Markov Chain Monte Carlo (MCMC) procedures, Gibbs sampling and Metropolis-Hastings, to get exact measures of uncertainty associated with the random parameters.

The following analytic results for adolescents in Miami using tract groups as the level 2 unit of analysis were based on both the HLM and MlwiN analyses for a simple ANOVA with random effects, using SUMRKDIF as the dependent variable, in which the level 1 model was: SUMRKDIF = $\beta_{0j} + r_{ij}$, and the level-2 Model was $\beta_{0j} = \gamma_{00} + u_{ij}$.

Comparing the MCMC's Gibbs sampling estimation with HLM's restricted maximum likelihood estimation, we found that level 2 variance shrunk a bit from 0.842 to 0.821, and the level 1 variance was a bit larger for the Gibbs sampling (7.947 vs 7.858), which resulted in a smaller intraclass correlation (0.0936) for the Gibbs sampling estimate. Despite this, the estimation magnitudes were still quite similar between the HLM's maximum likelihood estimate and the MCMC estimates.

B. **Test of homogeneity of level-1 variance.** To assess whether the assumption that the errors in the Level 1 model have equal variance across the level 2 units, we tested the random coefficient model on adolescents' perception of the difficulties of obtaining drugs with the adolescents' age and the lifetime use of cigarettes as the level 1 predictors. The results indicted that the Chi-square statistic was 41.874 with 72 degrees of freedom and a P-value larger than .50. This showed that no significant heterogeneity of variance was found and that our specification of the level 1 model was appropriate.

V. Comparison between HLM and Ordinary Least Squares (OLS). For the data that we analyzed, we found that HLM and OLS gave different or similar results depending on the degree of clustering at level 2. As we have indicated, the degree of clustering varies by dependent variable (less for risk of using marijuana than for difficulty of obtaining drugs) and by city (less for Denver youth than for Miami youth). Not shown is the fact that the cluster varies in predictable ways by size of the level 2 unit. For example, while 13% of the total variation in the difficulty in obtaining drugs for the 12 and older group was explained at the segment level, only 10.6% was explained when the Census tract was used as the level 2 unit. Similarly, for all persons 12 and older in Miami, 17.8% of the variation was at the segment level, while only 13.2% existed at the broader tract level.

The results indicated that parameter estimates associated with the fixed effects were similar across the HLM and the OLS when either the intra-unit correlations

HLM and MlwiN estimates

	HLM	MLwiN		
	Restricted ML	Restricted Iterated GLS	Gibbs Sampling	Metropolis - Hastings (MH)
Miami				
Intercept	5.819 (0.112)	5.869 (0.121)	5.871 (0.127)	5.862 (0.132)
Level 2 Variance	0.842	0.757	0.821	0.822
Level 1 Variance	7.858	7.926	7.947	7.945
Intracorrelation	0.0968	0.0872	0.0936	0.0938

are relatively small or the level-2 sample sizes are small, or both. The relationship between the standard error of the estimated parameter using the multilevel approach relative to OLS is (Goldstein, 1995): $S.E._{HLM} = \{1 + \rho_y\rho_x[n-1]\}^{1/2} * S.E._{OLS}$. Thus, for all persons 12 and older in Miami, where $\rho_y = \rho_x = 0.18$ and there were 14 level 1 units per level 2 unit, the standard error estimated through HLM would be about 1.19 times the standard error estimated through OLS. In this situation, OLS would not give adequate estimates for the fixed coefficients. Therefore, in general, before conducting an OLS analysis, one should determine the size of the intraclass correlation and apply this information to the Goldstein formula above to see whether use of OLS is warranted.

VI. Future. Starting in 1999, the number of sample persons in the NHSDA has been increased in order to make possible the calculation of state level estimates. States will be divided into field interviewer regions (approximately 10-15 in 42 of the smaller states and more in the 8 largest states), and segments will be the first stage of selection. This type of design also should facilitate modeling state and lower level hierarchical effects at the neighborhood level.

In the above analysis, a significant portion of the variation (about 80% or more) was attributed to person level variation. We know that the person-level variation really includes the variation between families as well. Since we know from basic research that family-level variables such as family drug use, parental attitudes about drug use, and family management practices and family conflict have been identified as risk factors for adolescent drug use, it is possible that a significant amount of what we have labeled as person-level variation is really between family variation. These components of variance need to be estimated.

The issue of small sample sizes at the family level in particular needs to be explored. In the 1997 and later NHSDAs, we have attempted to collect data on a random sample of pairs of persons within a sample of households. Using Goldstein's formula for the impact of this vis-a-vis OLS, the (n-1) would be equal to 2, and the inflation factor above and beyond simple random sampling would be (1+rho). For a large rho, this might make a considerable difference. If the level-2 sample sizes were larger, then even a small rho could make a difference. So, it is important to understand that any conclusions with respect to the use of HLM versus OLS apply only to the NHSDA and only with the current NHSDA sample sizes. The impact on other data sets or on the NHSDA with different sample sizes could be quite different.

While the variables that have been analyzed are interesting, there is a need to consider other drug-related variable scales and to extend the analysis to dichotomous data, like past year use of marijuana or any illicit drug.

Another widely-used software in this area is MLwiN. MLwiN can handle multiple levels of hierarchy and both continuous and dichotomous data as can HLM. We would like to use MlwiN to analyze dichotomous data based on relatively low prevalences and small level 2 sample sizes.

We need to develop more variables at the neighborhood level that are less SES-related (like the Census variables) and more drug-related. One way to do this might be to take the weighted average of a drug-related variable at level 1 and use it as a level 2 variable. The other way would be to search for drug-related measurements from other existing record systems.

Acknowledgments: The authors gratefully acknowledge the contributions of Mary Ann Hill and Bob Johnson (NORC). Mary Ann contributed her expertise in graphical interpretation and data analysis, and Bob provided needed help in the initial phases of file setup, model specification, and theoretical interpretation of the software results.

References

1. Bryk, Anthony S., Stephen W. Raudenbush (1992) *Hierarchical Linear Models: Applications and Data Analysis Methods* Sage Publications

2. Bryk, Anthony S., Stephen W. Raudenbush, and Richard T. Congdon (1996) *HLM: Hierarchical Linear and Nonlinear Modeling with the HLM/2L and HLM/3L Programs,* Chicago, IL: SSI

3. Goldstein, Harvey (1995) *Multilevel Statistical Models.* London, Edward Arnold: New York, Wiley.

4. Goldstein, Harvey, Rasbash, Jon, Plewis, Ian, Draper, David, Browne, William, Yang, Min, Woodhouse, Geoff, and Healy, Michael. (1998) *A User's Guide to MlwiN*

5. Pfeffermann, D., Skinner, C.J., Holmes, D., Goldstein, H., et al (1997) "Weighting for unequal selection probabilities in multilevel models" Journal of the Royal Statistical Society, B., 60: 23-40.

6. SAS Institute. (1992) *SAS Technical Report P-229, SAS/STAT Software: Changes and Enhancements* Cary, NC: SAS Institute, Inc.