



# H·CUP

HEALTHCARE COST AND UTILIZATION PROJECT

## HCUP Methods Series



Agency for Healthcare  
Research and Quality



U.S. Department of Health and Human Services  
Agency for Healthcare Research and Quality

**Contact Information:**  
**Healthcare Cost and Utilization Project (HCUP)**  
**Agency for Healthcare Research and Quality**  
**540 Gaither Road**  
**Rockville, MD 20850**  
**<http://www.hcup-us.ahrq.gov>**

**For Technical Assistance with HCUP Products:**

**Email: [hcup@ahrq.gov](mailto:hcup@ahrq.gov)**

**or**

**Phone: 1-866-290-HCUP**

Recommended Citation: Houchens R, Chu B, Steiner C. *Hierarchical Modeling using HCUP Data* HCUP Methods Series Report # 2007-01 Online. January 10, 2007. U.S. Agency for Healthcare Research and Quality.  
Available:<http://www.hcup-us.ahrq.gov/reports/methods.jsp>.

## TABLE OF CONTENTS

<b>EXECUTIVE SUMMARY .....</b>	<b>i</b>
<b>INTRODUCTION .....</b>	<b>1</b>
Hierarchical Data Structures .....	3
Strict Hierarchies .....	3
Cross Classifications .....	4
Multiple Memberships .....	4
Other Hierarchical Data Structures .....	5
Hierarchical Regression Models: Key Concepts .....	6
Sources of Variation .....	6
Graphical Typology .....	8
Hierarchical Modeling in Health Services Research .....	1
Why HLM for Health Services Research? .....	1
Examples of HLM in the Health Services Literature .....	1
<b>Statistical methods .....</b>	<b>3</b>
HLM for Continuous Outcomes .....	3
General Two-Level Setup .....	3
Random Intercepts (Empty Model or Variance Components Model) .....	6
Random Intercepts, Fixed (Constant) Slopes .....	8
Random Intercepts, Random Slopes .....	9
Intercept- and Slope-As-Outcomes .....	10
HLM for Binary Outcomes (logistic regression) .....	12
HLM for Count Outcomes .....	13
Poisson Regression .....	13
Negative Binomial Regression .....	14
HLM for Longitudinal Data .....	14
HLM for Cross Classifications .....	15
HLM for Multiple Membership .....	15
Other Methods .....	16
Generalized Estimating Equations (GEE) .....	16
Bayesian Methods .....	16
Using Sampling Weights in HLM .....	16
HLM Compared to Conventional Regression: An Example Using HCUP Data .....	17

Data.....	17
Regressions for Log(Charges) .....	21
Logistic Regressions for In-hospital Mortality.....	23
Summary.....	25
<b>Statistical software .....</b>	<b>26</b>
Programming Code for Continuous Outcomes .....	27
SAS Programs .....	28
Random Intercepts, Empty Model.....	28
Random Intercepts, Fixed Slopes.....	29
Random Intercepts, Random Slopes .....	30
Intercept- and Slopes-As-Outcomes.....	32
Stata Programs .....	34
Random Intercepts, Empty Model.....	34
Random Intercepts, Fixed Slopes.....	34
Random Intercepts, Random Slopes .....	35
Intercept- and Slopes-As-Outcomes.....	36
Programming Code for Binary Outcomes .....	37
SAS Programs .....	38
Random Intercepts, Empty Model.....	38
Random Intercepts, Fixed Slopes.....	40
Random Intercepts, Random Slopes .....	41
Intercept- and Slopes-As-Outcomes.....	42
Stata Programs .....	43
Random Intercepts, Empty Model.....	43
Random Intercepts, Fixed Slopes.....	44
Random Intercepts, Random Slopes .....	45
Intercept- and Slopes-As-Outcomes.....	46
Programming Code for Count Outcomes.....	47
SAS Programs .....	48
Random Intercepts, Empty Model.....	48
Random Intercepts, Fixed Slopes.....	50
Random Intercepts, Random Slopes .....	51
Intercept- and Slopes-As-Outcomes.....	52
Stata Programs .....	53

Random Intercepts, Empty Model.....	53
Random Intercepts, Fixed Slopes.....	54
Random Intercepts, Random Slopes.....	55
Intercept- and Slopes-As-Outcomes.....	56
<b>Recommendations.....</b>	<b>58</b>
<b>REFERENCES.....</b>	<b>59</b>
<b>BIBLIOGRAPHY.....</b>	<b>62</b>

### INDEX OF TABLES

Table 1: OLS vs. HLM Estimates for Log(Total Charges).....	21
Table 2: Conventional vs. HLM Logistic Regression Estimates for In-Hospital Mortality.....	24

### INDEX OF FIGURES

Figure 1: Strict Hierarchy (patients treated by only one physician, and physicians have admitting privileges to only one hospital).....	3
Figure 2: Strict Hierarchy (ED visits nested within patients nested within hospitals).....	4
Figure 3: Cross Classifications (patients treated in one hospital by one physician but physicians have admitting privileges at multiple hospitals).....	4
Figure 4: Multiple Membership (patients 2, 7, and 9 treated by multiple physicians).....	5
Figure 5: Single-Level Information Ignoring Clustering.....	6
Figure 6: Two-Level Information Accounting for Clustering.....	7
Figure 7: Varying Relationships: A Graphical Typography.....	9
Figure 8: Distribution of Logarithm of Total Charges for CHF Patients, 2004 NIS.....	18
Figure 9: Distribution of Age for CHF Patients, 2004 NIS.....	18

## EXECUTIVE SUMMARY

This document continues the HCUP Methods Series of reports, which features information on a broad array of methods as they relate to the HCUP databases and software tools. This series is designed to help HCUP data users work efficiently and effectively with HCUP data. This report is consistent with that goal.

Hierarchical linear modeling (HLM) is a regression technique designed to deal with clustered or grouped data in which analytic units are naturally nested or grouped within other units of interest. For example, a physician's patients form a group nested within that physician. In analyzing outcomes from a sample of patients treated by a number of physicians, interest centers on the effects of both patient and physician characteristics. Since each group of patients is treated by a single physician, it is expected that those patients' outcomes will be correlated, which violates one of the assumptions of standard regression methods. If this correlation is ignored, wrong inferences can result with respect to the effects of both patient and physician factors. HLM accounts for this "within-group" correlation to produce better inferences when the proper model is specified.

HLM has been used increasingly in health services research over the past 20 years. Early adopters tended to be academicians with an interest in the statistical theory or in the technical challenges associated with the development of modeling software. Education was one of the first areas for relatively widespread application. In education, the "hierarchy" consists of students nested within classrooms nested within schools nested within school districts, and so on. Each level of the hierarchy, from students at the first level to school districts at the highest level, can conceivably contribute to variation in student performance. Similar hierarchies exist in health services research. As more researchers appropriately apply HLM, it moves further into the consciousness of the entire health research community.

Nevertheless, many researchers see HLM as a technique that lies too far outside their competency to even contemplate its application. It has only recently made its way into college curriculum; consequently, many analysts have been forced to learn HLM on the job, which can be a daunting task without some guidance. This report was developed to make that task easier. It is intended as a basic introduction to HLM, particularly for HCUP data analysts. It presents key concepts, introduces standard model notation, covers statistical packages for HLM, and gives SAS and Stata program code for a range of standard HLM models.

To be sure, the help of a trained statistician is recommended for the analysis of complex data structures possibly involving non-standard patterns of variances and covariances. However, the basic models should be easily within the grasp of most analysts with a solid background in standard regression techniques. For researchers who want to delve more deeply into HLM, numerous readily-available articles, books, and Websites are devoted to this topic. Many of these resources are listed in the references found throughout this report, especially in the bibliography appearing at the end of this document.

## INTRODUCTION

The Healthcare Cost and Utilization Project (HCUP, pronounced "H-Cup") is a family of health care databases and related software tools and products developed through a Federal-State-Industry partnership and sponsored by the Agency for Healthcare Research and Quality (AHRQ). HCUP databases bring together the data collection efforts of state data organizations, hospital associations, private data organizations, and the Federal government to create a national information resource of patient-level health care data ([HCUP Partners](#)). HCUP includes the largest collection of longitudinal hospital care data in the United States, with all-payer, encounter-level information beginning in 1988. These databases enable research on a broad range of health policy issues, including cost and quality of health services, medical practice patterns, access to health care programs, and outcomes of treatments at the national, state, and local market levels.

This report addresses the practical and statistical reasons and implications for utilizing hierarchical linear modeling (HLM) when analyzing HCUP state and national derivative databases. This report is intended to: 1) define and explain hierarchical models, 2) explore the advantages and disadvantages relative to other statistical approaches, and 3) guide analysts in the application of hierarchical models to HCUP data and the interpretation of results.

It is important to recognize that this report is intended as an *introduction* to HLM. It explains basic concepts, standard notation, and elementary analyses. However, it does not cover model-building strategies and model diagnostics to assess the fit and specification of hierarchical models, for which there are several good references contained in the bibliography.

In this report, the term *hierarchical modeling* refers to regression methods that are pertinent to hierarchical data structures, in which lower-level units are nested or grouped within higher-level units, such as patients nested within hospitals. In the literature, variations of these models are also labeled *mixed models*, *random effects models*, *random coefficient models*, *covariance components models*, or *variance components models*. The terms *hierarchical model* and *multilevel model* are often used interchangeably.<sup>1</sup>

Hierarchical data structures occur naturally in health services research (Goldstein et al., 2002). For example, a study might have a group of patients nested within physicians, physicians nested within hospitals, and hospitals nested within states. This sets up a four-level hierarchy of patients at level one, physicians at level two, hospitals at level three, and states at level four. To the extent that physicians uniquely affect patient therapy, health outcomes might be correlated among patients treated by a single physician. Likewise, outcomes might be correlated among physicians within hospitals and among hospitals within states. In contrast to standard regression techniques, hierarchical methods account for these "within-cluster" correlations at each level of the hierarchy and properly adjust estimates to account for them.

Hierarchical models are especially relevant when (Normand et al., 1997; Goldstein, 2003; Subramanian, 2004; Shahian et al., 2005):

- the observations being analyzed are correlated,
- the causal processes are thought to operate at more than one level; and/or

---

<sup>1</sup>However, Molenberghs and Verbeke (2004) argue that the satisfactory treatment of a *fully hierarchical model* is only possible within a Bayesian context.

- the research interest is especially focussed on describing the variability and heterogeneity in the population, rather than on the average values, or
- there is interest in estimating level-specific effects (e.g., specific hospital effects).

HCUP data often contain nested structures that should not be ignored. For example, patients vary, but so do hospitals. Differences among hospitals might be attributed to the characteristics of the patients they serve, the characteristics of the hospitals themselves, or both. Thus, both patient and hospital factors can affect the study outcome of interest, suggesting that both patient and hospital covariates should enter the analysis. Intuitively, patients treated within one hospital will be more similar than patients treated within another hospital, inducing some correlation among the patients' outcomes within hospitals.

Historically, analysts sometimes have chosen either 1) to aggregate the data to a higher level and conduct analyses on the higher-level units, or more commonly, 2) to disaggregate the data to the lowest level units and conduct analyses on the lower-level units.

The primary drawbacks for an aggregate analysis, such as summarizing patient-level data to the hospital level, are that: 1) it ignores within-hospital variation, 2) the aggregated variables are hospital-level variables, not patient-level variables, and 3) a correlation between hospital-level variables cannot be asserted at the patient level.<sup>2</sup>

The principal drawbacks for a disaggregate analysis, such as inserting hospital-level variables into a traditional patient-level regression, are that: 1) it ignores the data hierarchy, causing the between-hospital variance to be absorbed into the error term, and 2) it inflates the amount of hospital-level information that is available (the patient sample is much larger than the hospital sample). Ordinary regression methods require that observations be independent or uncorrelated. Failure to account for correlated units, the cause of the correlation, and the nature of the correlation, can lead to underestimates of standard errors producing erroneous statistical significance, as illustrated by Austin et al. (2003) and by the example analysis presented later in this report.

There are several statistical methods that deal with clustered data and render valid estimates of standard errors. For example, statistical methods have been devised to handle cluster sampling and other sample survey designs involving group sampling protocols. The method of generalized estimating equations (GEE) accounts for clustering and treats the covariances as "nuisance" parameters not to be estimated. What sets hierarchical modeling apart from these other methods is its ability to model variability at each level of the hierarchy.

Using the example of patients within hospitals, the distinguishing features of a hierarchical model are:

- A separate regression is fit to the patients at each hospital.
- For each regression, the outcome is regressed on patient-level variables.
- The regressions are related across hospitals.
- There are both within-hospital and between-hospital sources of variation.
- Each of the regression coefficients on patient-level variables can be either fixed (constant) or variable (random) across hospitals (Krumholz et al., 2006).

---

<sup>2</sup>Interpreting aggregated data at the individual level is referred to as *the ecological fallacy*.



- For patient-level variables that are allowed to vary across hospitals, the variation can be modeled as a function of hospital-level variables (e.g., Gatsonis et al., 1995). For example, the effect of age can vary between teaching and non-teaching hospitals.
- The residual variance at each level can be modeled as a function of explanatory variables (Snijders and Bosker, 1999).
- Hospital-specific effects can be estimated (Normand et al., 1997).

These aspects provide a rich framework for modeling clustered data and exploring complex relationships among the various patient and hospital characteristics.

### Hierarchical Data Structures

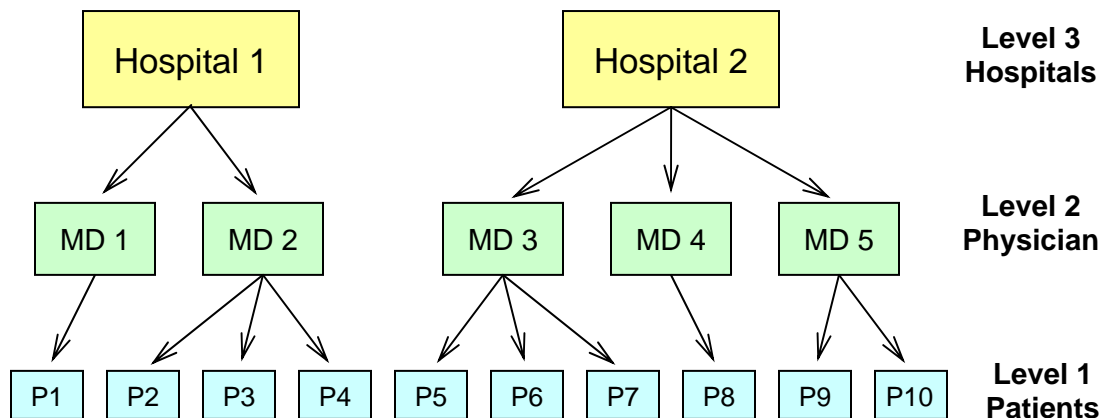
The Healthcare Cost and Utilization Project (HCUP) produces several databases useful to health services researchers that contain hierarchical data structures.

Before a hierarchical model can be formulated properly, it is necessary to recognize the type of hierarchy that the data represent. Not all multiple-level data configurations consist of units that are completely nested within successively larger units. Three main types of data structures are discussed: 1) strict hierarchies, 2) cross-classifications, and 3) multiple memberships. In practice, complicated relationships among entities can be difficult to untangle.

#### Strict Hierarchies

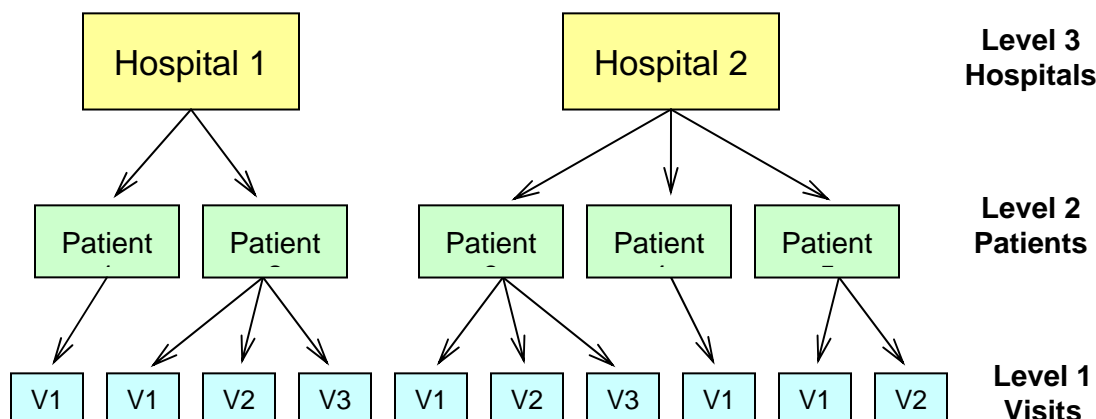
This is the simplest structure. In a three-level structure, level-1 units are strictly nested within level-2 units, which are strictly nested within level-3 units. For example, each patient might have only one physician and each physician might practice in only one hospital. In Figure 1, this structure is illustrated for 10 patients, five physicians, and two hospitals.

**Figure 1: Strict Hierarchy (patients treated by only one physician, and physicians have admitting privileges to only one hospital)**



Another example of a strict hierarchy involves patients with multiple visits. For example, some patients might have multiple emergency department (ED) visits to the same hospital. In this case, the visits (level 1) are nested within patients (level 2), which are nested within hospitals (level 3), as illustrated in Figure 2. This same sort of hierarchy would apply to a hospital discharge database containing multiple hospital stays for some patients.

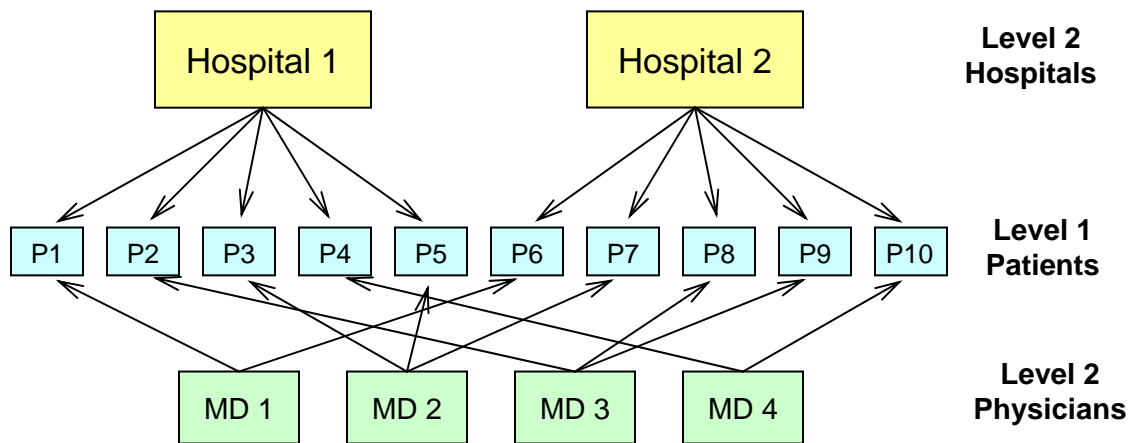
**Figure 2: Strict Hierarchy (ED visits nested within patients nested within hospitals)**



Cross Classifications

Cross classifications exist when higher-level units are not nested within one another. For example, each patient might have only one physician, and be treated at only one hospital, but each physician might use multiple hospitals. Therefore, physicians are not nested within hospitals. Patients are nested within a cross-classification of physicians and hospitals, as illustrated in Figure 3. Patients are grouped in two ways: by physicians and by hospitals. However, physicians and hospitals are considered to be at the same level because physicians are not nested within hospitals and hospitals are not nested within physicians.

**Figure 3: Cross Classifications (patients treated in one hospital by one physician but physicians have admitting privileges at multiple hospitals)**

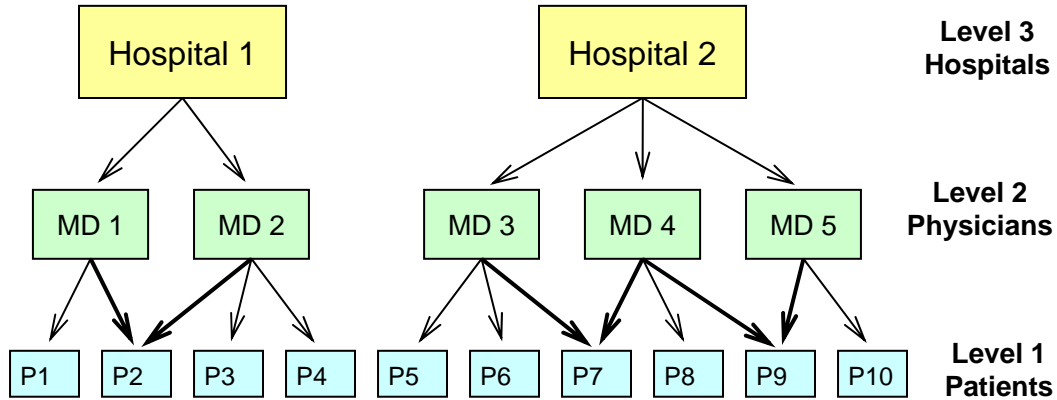


Multiple Memberships

Multiple memberships exist when level-1 units belong to more than one higher-level unit. For example, some patients are treated by more than one physician, as illustrated in Figure 4. In this example, each physician uses only one hospital. However, patients 2, 7, and 9 are each

treated by two different physicians (heavy lines). That is, some patients are members of more than one physician “group.”

**Figure 4: Multiple Membership (patients 2, 7, and 9 treated by multiple physicians)**



### Other Hierarchical Data Structures

More complicated structures can occur in HCUP data. For example, over the course of several hospitalizations, a patient could be treated by more than one physician and be admitted to more than one hospital, setting up a combination of multiple memberships and cross classifications. Diagrams such as those shown in Figures 1 through 4 become unwieldy for more complex data structures.

For simplicity of exposition, this report will be concerned mainly with strict hierarchies. For readers interested in more complicated hierarchical data structures, Browne et al. (2001) discuss Multiple Membership Multiple Classification (MMMC) models and propose a simple scheme for diagramming complex hierarchical data structures. Other useful references for more complicated models will be given throughout this report. A structured bibliography, provided by Gatsonis, Normand, and Zaslavsky (private communication), is presented near the end of this report.

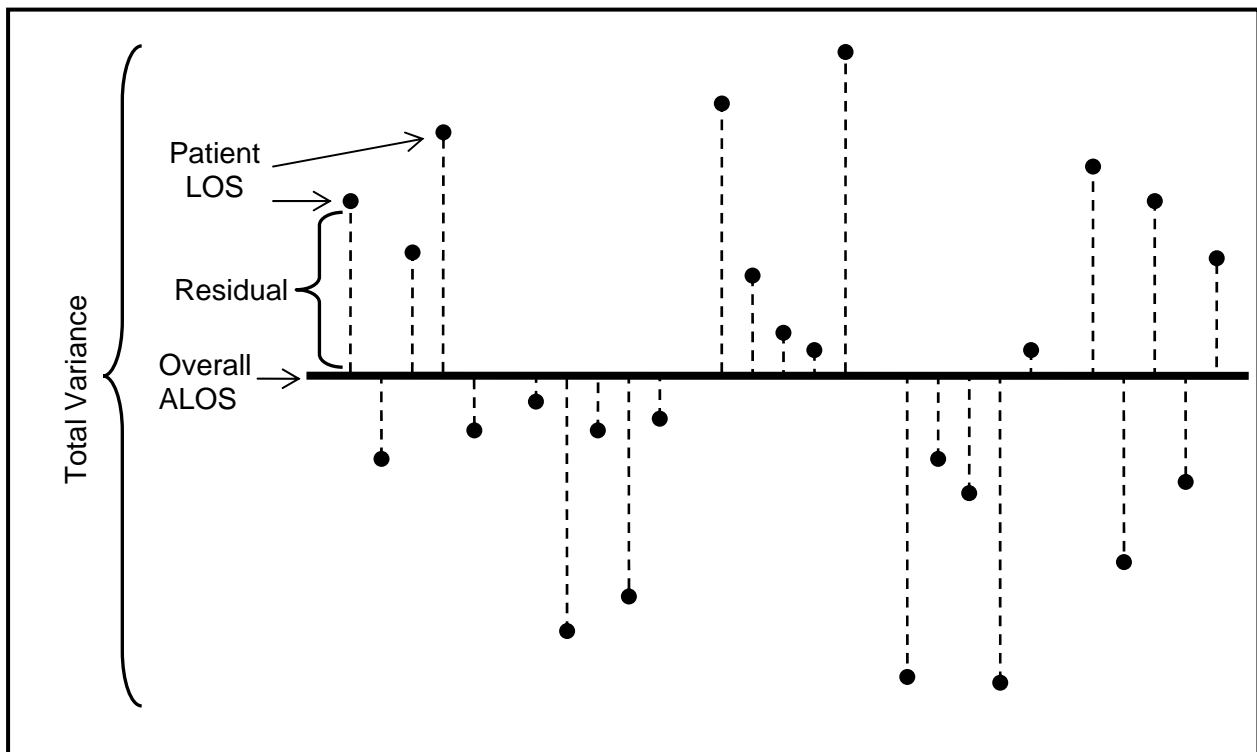
## Hierarchical Regression Models: Key Concepts

Here, some fundamental concepts are introduced through a few simple illustrations.

### Sources of Variation

Consider a simple two-level hierarchy of patients within hospitals for patient lengths of stay (LOS) in the state of California. If one ignores the clustering of patients within hospitals, the patient-level variation in LOS can be illustrated as in Figure 5. The heavy horizontal line represents the overall average length of stay (ALOS) across all patients in California. The length of each vertical dashed line represents the residual, which is the difference between each patient's observed LOS and the overall ALOS. The range of residuals around the overall ALOS represents the total variance in the data.

**Figure 5: Single-Level Information Ignoring Clustering**



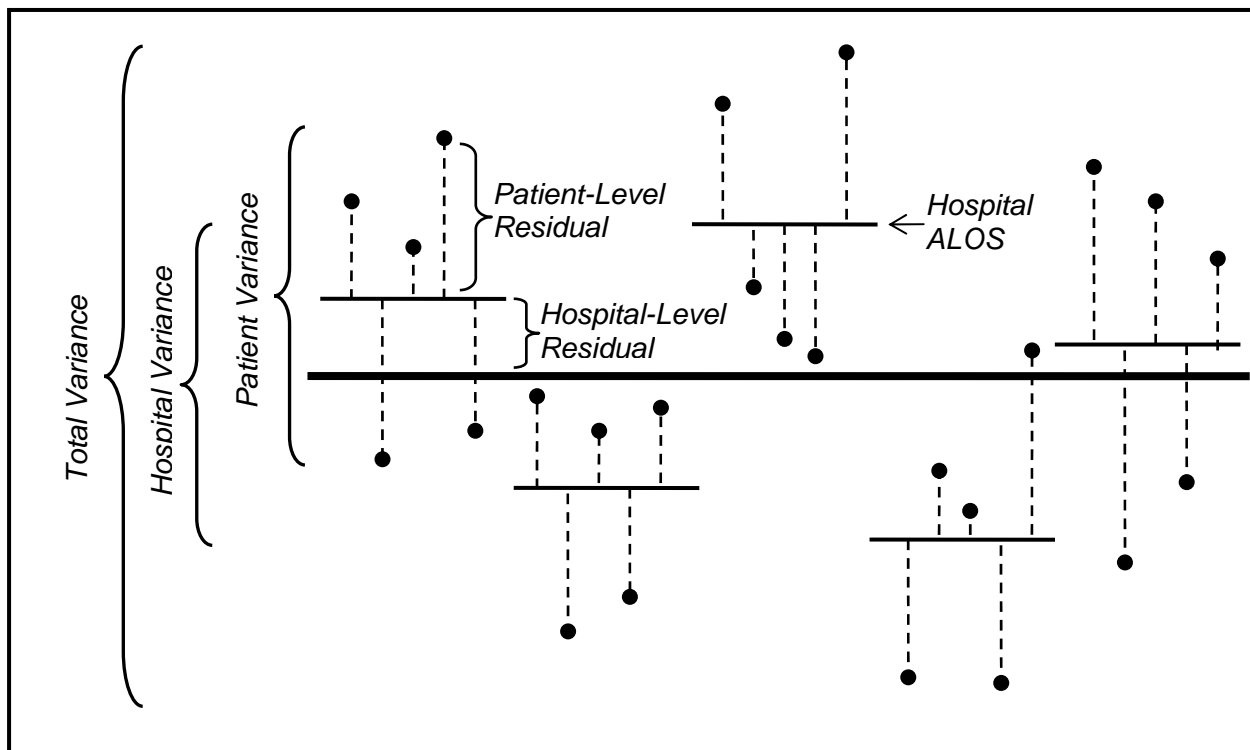
Informally, one can write an “intercept-only” regression model corresponding to Figure 5:

$$L_{ij} = M + E_{ij}$$

where  $L_{ij}$  is the observed LOS for patient  $i$  in hospital  $j$ ,  $M$  is the overall ALOS for California patients, and  $E_{ij}$  is the random error (residual) for patient  $i$  in hospital  $j$ . For simplicity, this regression contains only a constant intercept term and a random error term; it does not have any explanatory variables such as patient age and gender.

Figure 6 shows the same data, accounting for both hospital variance and patient variance. The short horizontal lines represent ALOS for each of five hospitals. Each hospital-level residual is equal to the difference between the hospital's ALOS and the overall ALOS (represented by the thick horizontal line). Each patient's residual is now equal to the difference between the patient's observed LOS and the hospital's ALOS. Thus, the residual in Figure 5 (above) is the sum of the patient-level residual and the hospital-level residual shown in Figure 6. The hospital variance is represented by the spread of the hospital-level residuals around the overall mean. The patient variance is represented by the spread of the patient-level residuals around the hospital means.

**Figure 6: Two-Level Information Accounting for Clustering**



Corresponding to Figure 6, one can write an informal hierarchical model, called a “null,” “empty,” or “variance components” model:

$$L_{ij} = M + H_j + E_{ij}$$

where  $L_{ij}$  is the observed LOS for the patient  $i$  in hospital  $j$ ,  $M$  is the overall ALOS for an average California hospital,  $H_j$  is the residual<sup>3</sup> for hospital  $j$ , and  $E_{ij}$  is the residual for the patient  $i$  in hospital  $j$ . This model has two random errors,  $H_j$  and  $E_{ij}$ . The main purpose of the empty model is to partition the total LOS variance into a variance associated with  $H_j$  that occurs between hospitals, and a variance associated with  $E_{ij}$  that occurs between patients within hospitals:

$$V_{\text{Total}} = V_H + V_E.$$

Estimates for these components of variance can be used to estimate the intraclass correlation coefficient (ICC), which is the proportion of total variance that is between hospitals:

<sup>3</sup>Technically, the hospital residuals are based on Bayes “shrinkage” estimates of the hospital means. This concept is described more fully later in this report.

$$ICC = \frac{V_H}{V_H + V_E}$$

One interpretation for the ICC in the patient-hospital hierarchy is as the average correlation between the outcomes of randomly-paired patients, where patients are paired only within hospitals, not across hospitals. The ICC lies between 0 and 1. If all patients within a hospital have the same LOS, then  $V_P = 0$  and  $ICC = 1$ . At the other extreme, if all hospitals have the same ALOS, then  $V_H = 0$  and the  $ICC = 0$ . Intuitively, the higher the ICC, the more important it is to recognize the clustering of patients within hospitals in estimating variances. HCUP researchers should pay close attention to this because conventional methods will tend to over-estimate the significance of higher-level variables, such as hospital teaching status, when the ICC is greater than zero by an amount that is substantively important.

These fundamental principles will be expanded and made more mathematically precise in later sections.

### Graphical Typology

It is instructive to view some typical hierarchical models from a graphical perspective. The typology in Figure 7 is based on Subramanian, Jones, and Duncan (2003). These plots represent regressions of patient length of stay (response, vertical axis) on patient age (predictor, horizontal axis) for data from a two-level structure: patients at level 1 within hospitals at level 2.

Figure 7(a) represents a “fixed effects” model. The relationship is the same straight line for all patients from all hospitals. This single straight line could have been estimated using ordinary least squares (OLS) regression with a “fixed” intercept and a “fixed” age coefficient. This relationship assumes that the same regression holds for patients from every hospital.

Figure 7(b) represents a *random intercept* or *variance components* model, which is commonly-used in hierarchical modeling. There are separate parallel lines for each of 10 hospitals, with the same slope for all hospitals, but with different intercepts. These hospital-specific lines vary around the underlying average relationship, shown by the thicker line in the middle. These lines represent a hierarchical model with a random intercept and a “fixed” or constant slope. The intercepts for each hospital vary randomly around the overall average intercept. Hierarchical models can be used to estimate the hospital-specific relationships in addition to the overall average relationship between LOS and age.

Figures 7(c) - 7(f) represent various random intercept and random slope models. In each plot, the thicker line represents the overall (population) average relationship between patient age and patient LOS. However, the hospital-specific lines are not parallel because the hospital-specific age coefficients are conceptually random draws from the “population” of hospital-specific age coefficients. In some of the diagrams, the intercepts and slopes are correlated, and hierarchical models can be used to estimate the correlation between these parameters along with the intercepts and slopes.

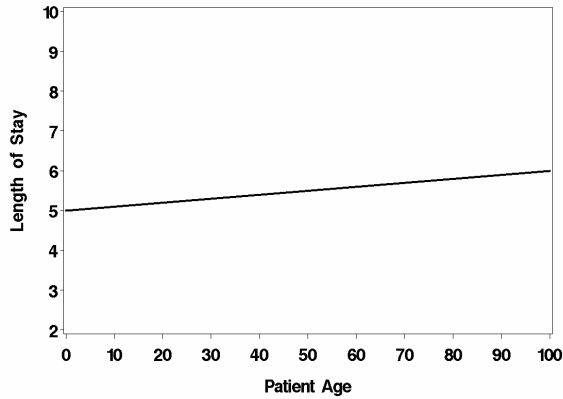
In Figure 7(c), the intercepts and slopes are positively correlated. Hospitals with higher intercepts have steeper slopes. The lines are farther apart for older people than they are for younger people, indicating that the variation in hospital ALOS is greater for older people than it is for younger people. Thus, one would conclude that hospitals have more of an (absolute) effect on LOS for older patients than for younger patients. Figure 7(d) exemplifies intercepts and

slopes that are negatively correlated. Figure 7(e) illustrates intercepts and slopes that are uncorrelated.

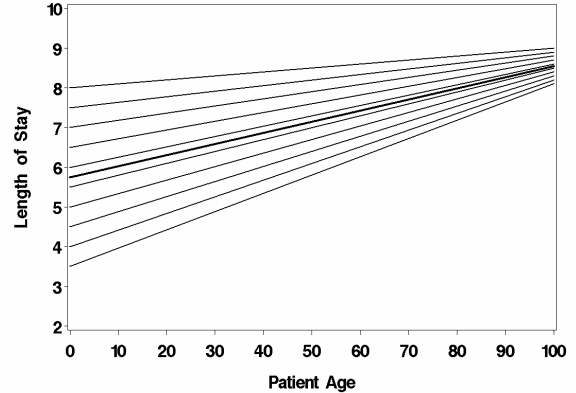
Figure 7(f) shows an example where there is no overall relationship between LOS and age (thick line has zero slope), but individual hospitals do have relationships.

Figure 7: Varying Relationships: A Graphical Typography.

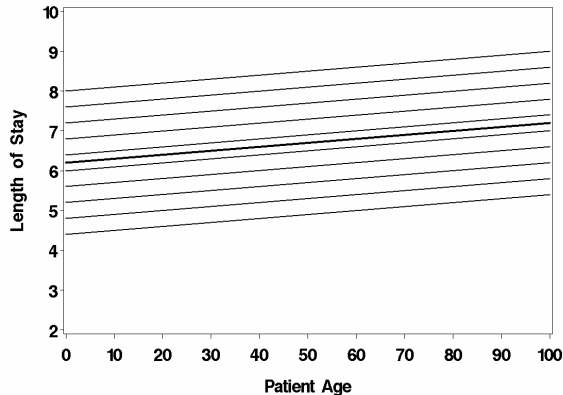
a) constant intercept and constant slope



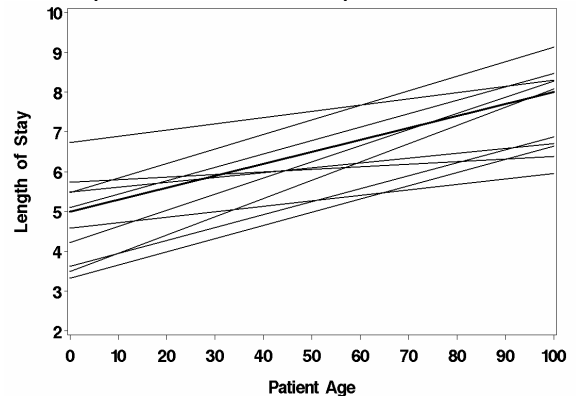
d) negative correlation between random intercepts and random slopes



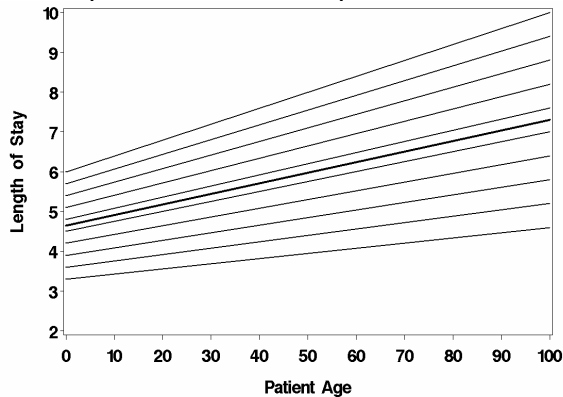
b) random intercepts, constant slope



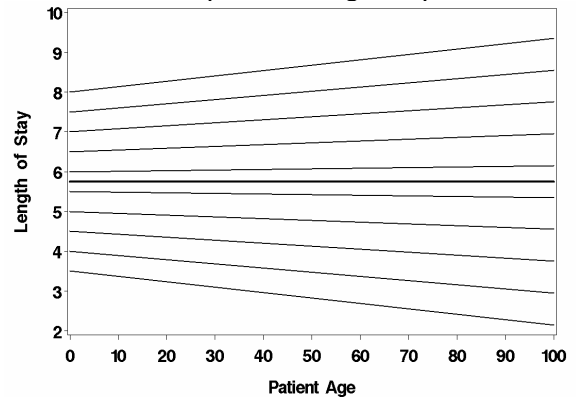
e) zero correlation between random intercepts and random slopes



c) positive correlation between random intercepts and random slopes



f) positive correlation between random intercepts and random slopes, average slope is zero



## Hierarchical Modeling in Health Services Research

### *Why HLM for Health Services Research?*

The term *hierarchical model* represents a regression model that explicitly recognizes the nested structure of data. The nature of the clustering and its effects are of substantive interest, and therefore modeled explicitly. The term *marginal models* or *population-average models* represent regression models that also recognize the nested structure of data, but which treat the clustering as an uninteresting nuisance that is incorporated only to adjust the standard errors associated with the regression estimates. They estimate the population average effects, like those represented by the heavy lines in Figure 7.

Although the use of HLM is increasing, to date most studies using clustered data in the health services literature have employed OLS models (for continuous outcomes) and marginal models. OLS models ignore the clustering and hence tend to under-estimate standard errors for the regression coefficients. Marginal models, like GEE, are robust (Liang and Zeger, 1986) and they address the OLS problem with standard errors. Both of these methods produce estimates of “fixed” (average) effects of a predictor, such as age, that are nearly identical to those produced by HLM, when the error terms are normally distributed.

So why bother with HLM? Health services research (HSR) often attempts to distinguish between compositional effects and contextual effects. For example, differences in hospital mortality rates might be explained by each hospital’s mix of patients (composition), by each hospital’s characteristics (context), or by a combination. For instance, inpatient mortality rates might differ among hospitals because some hospitals serve more severely ill patients than others, or because some hospitals have available more advanced technology than others. If so, the benefits of advanced technology might accrue only to specific types of patients. If hospital characteristics are found to have a significant impact, then policy might be established to direct certain kinds of patients toward appropriate types of hospitals for inpatient care. Thus, it is critical to consider both patient and hospital sources of variation, and appropriately assess the statistical significance of both patient and hospital effects.

The key to HLM lies in the treatment of the clustering and the heterogeneity of the outcome. OLS and marginal models take a “means only” perspective, ignoring potential information about the variability in the response that one seeks to explain through the regression model. With HLM, one can ask (and try to answer) a number of new questions, such as the interpretation of the variances and covariances of the hospital effects, and their predictors. Suddenly, one can ask questions about both the averages and the covariance structure. For example, when adding a patient-level predictor such as disease severity, how much hospital-to-hospital variability remains? Does the relationship between mortality and severity vary among hospitals, and if so, what does it mean? Hierarchical models enable analysts to investigate and explain the sources of both within-hospital and between-hospital variation.

### *Examples of HLM in the Health Services Literature*

Hierarchical linear models have been used in HSR since the late 1980s. Numerous authors proposed studying “relative effectiveness” or “provider profiling” by means of HLM (Dubois et al., 1987; Jencks et al., 1988; Thomas et al., 1994; Normand et al., 1995; Epstein, 1995; Schneider and Epstein, 1996; Morris and Christiansen, 1996; Goldstein and Spiegelhalter, 1996; Rice and Leyland, 1996; Leyland and Boddy, 1998; Marshall and Spiegelhalter, 2001). Subsequently, the



use of HLM has grown over the past 15 years. For example, a search of PUBMED for HLM-related terms<sup>4</sup> returned titles for more than 650 published articles, over half of which were published within the past 5 years. A further search of *ongoing* health services research projects returned more than 75 studies that proposed the use of HLM.

Several general articles on the use of HLM in HSR have appeared. Rice and Leyland (1996) published an introductory account of HLM, drawing on then-recent applications in HSR. Duncan, Jones, and Moon (1998) also considered the use of HLM in health research. They developed four graphical typologies to outline the questions that hierarchical models can address and illustrated their potential by drawing on published examples from different research areas. The article by Sullivan, Dukes, and Losina (1999) introduced HLM generally and applied a two-level model to data from the Type II Diabetes Patient Outcomes Research Team (PORT) study. More recently, Goldstein, Browne, and Rasbash (2002) published a tutorial on HLM for medical data. They covered a wide range of models and introduced the MLwiN modeling software. Finally, Leyland and Goldstein (2001) presented a series of articles on HLM in health statistics.

To date, there have been only a few published examples of HLM applied specifically to HCUP data:

- Dowell et al. (2004) applied HLM to the 1994-1997 HCUP Nationwide Inpatient Sample (NIS) to study disparities in treatment and outcomes for patients hospitalized with Type II diabetes.
- Carey (2002) fitted hierarchical models to the 1996 HCUP NIS to demonstrate that declines in hospital ALOS resulted in hospital cost savings.
- Bramble and colleagues (2004) also used the 1996 NIS to fit a three-level hierarchical linear model, with patients (level one) clustered within hospitals (level two) clustered within markets (level 3), to determine the effect of hospital teaching status and hospital experience on resource use (charges and length of stay).

However, there have been several published examples of HLM applied to administrative databases like those produced by the HCUP project (e.g., Localio et al., 1997; Amaravadi et al., 2000; Spiegelhalter et al., 2002; Scott et al., 2004; Estabrooks et al., 2005).

More generally, there exist many specific examples of HLM in the HSR literature and other areas of research, some of which can be found in the list of references and bibliography at the end of this report.

---

<sup>4</sup>PUBMED (<http://www.ncbi.nlm.nih.gov/entrez/query.fcgi>) and HSRproj (<http://www.nlm.nih.gov/hsrproj>) using search terms “hierarchical model,” “hierarchical linear model,” “multilevel model,” or “multi-level model.” Some of the articles incorporating these terms actually might not have applied HLM methods. However, a perusal of the abstracts suggests that the vast majority did fit HLM regressions.

## STATISTICAL METHODS

In this section, standard mathematical notation and statistical assumptions are introduced, so that analysts can understand applications in the literature and apply the statistical software used to fit these models. These models are compared to other regression methods.

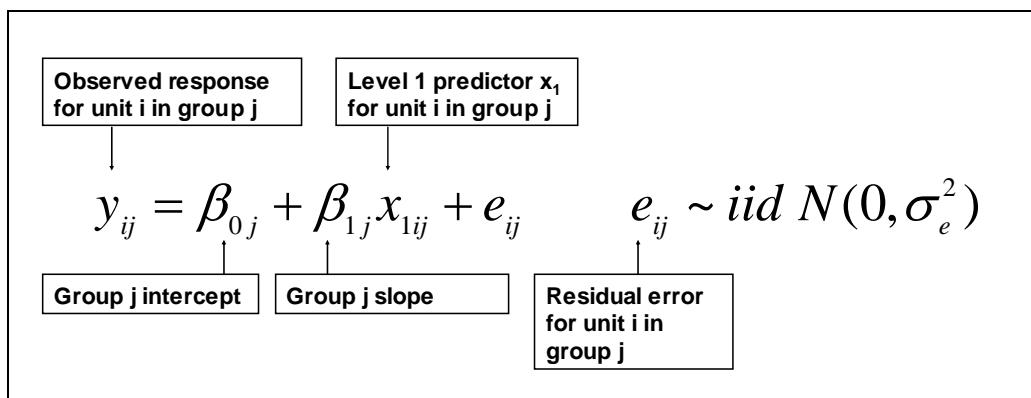
In HLM, each level corresponds to a distinct population from which the sample of units at that level are drawn. As an example, the NIS is a sample of about 1,000 hospitals from the population of about 5,000 hospitals nationwide. Although the NIS contains 100 percent of that year's discharges from each sampled hospital, each hospital's patients can be viewed as a sample from a larger population of potential patients. Their outcomes can be viewed as a random sample of outcomes from a population of "similar" patients. Many illnesses strike at random, causing people to be hospitalized at random times. If a patient obtained care on a Wednesday, the treatment and outcome could well have been different if care had been rendered on a Saturday, or even if it had occurred an hour earlier or later than it did, because of the way medical staff are assigned to cases, especially emergencies. Therefore, the collection of patient outcomes during a given year for a specific facility can easily be considered a random sample, even for that specific "population" of patients.

### HLM for Continuous Outcomes

The hierarchical linear model is an extension of the general linear model in which the probability model for the errors, or residuals, has a structure reflecting the hierarchical structure of the data. In this section, models for hierarchies with two levels are constructed and explained. However, the extension to more levels is straightforward. Throughout this section, level-1 units are nested within level-2 units, such as patients nested within hospitals. The notation used in this section of the report is commonly used in texts, journal articles, and specialized computer software for HLM.

#### General Two-Level Setup

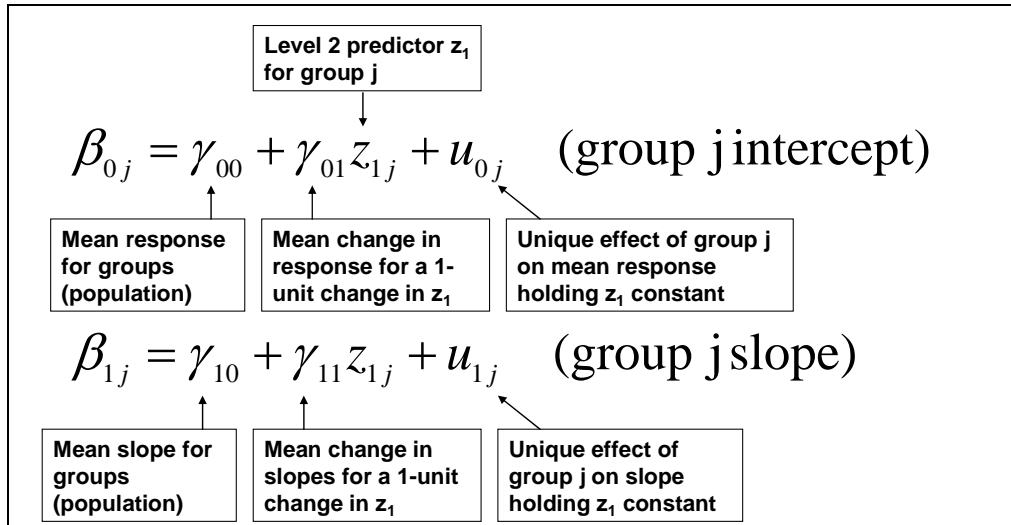
The model for a two-level hierarchy can be built up as follows. First, specify the level-1 regression of a response  $y_{ij}$  (e.g., hospital charges for patient  $i$  in hospital  $j$ ) on a single level-1 predictor  $x_{1ij}$  (e.g., age):



Level-1 units are indexed by  $i$  and level-2 units are indexed by  $j$ . In the equations above, and throughout this report, the term "group" is synonymous with the term "level-2 unit." This is a linear regression model, in which the intercepts and the slopes can depend on the level-2 units (groups or clusters) and therefore are indicated with the subscript  $j$ . The level-1 residuals,  $e_{ij}$ , are

assumed to be independent and identically distributed (iid) as normal, with a mean of zero and a variance of  $\sigma^2$ .

Next, the intercepts and slopes in the level-1 equation can be modeled as a function of level-2 predictors through the following level-2 regressions, which can involve a level-2 predictor,  $z_1$  (e.g., hospital teaching status):



In the level-2 equations, the group  $j$  intercept is equal to the sum of:

1. the “fixed” population mean,  $\gamma_{00}$ , for the level-2 intercepts,
2. a “fixed” multiple of a level-2 variable,  $z_1$ , and
3. a random error  $u_{0j}$  for level-2 unit  $j$ , also called the level-2 residual.

Similarly, the group  $j$  slope is the sum of three elements analogous to those for the group  $j$  intercept. The fact that the level-2 units are regarded as a random sample from a population is reflected by the presence of the random errors  $u_{0j}$  and  $u_{1j}$ .

In the literature, whenever at least one group variable,  $z$ , is included in both equations, the equation for the intercept is called an “intercepts-as-outcomes” model, and the equation for the slope is called a “slopes-as-outcomes” model (Snijders and Bosker, 1999; Raudenbush and Bryk, 2002).

The standard statistical assumptions for the error terms in the level-2 equations are as follows:

Variance of random intercepts

Covariance between intercepts and slopes

$$\begin{pmatrix} u_{0j} \\ u_{1j} \end{pmatrix} \sim iid N \left[ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \tau_{00} & \tau_{01} \\ \tau_{10} & \tau_{11} \end{pmatrix} \right] \quad \forall j$$

Variance of random slopes

$$\text{correlation} = \rho(u_{0j}, u_{1j}) = \frac{\tau_{01}}{\sqrt{\tau_{00} \cdot \tau_{11}}}$$

$u_{0j}$  and  $u_{1j}$  are uncorrelated with  $e_{ij}$

That is, the error terms for the intercepts and slopes are assumed to be distributed as bivariate normal, with possibly positive or negative covariance (Figures 7(c) and 7(d), respectively). This assumption allows the parameters to be estimated by maximum likelihood and allows for statistical tests of the regression coefficients.

Finally, substitute the level-2 equations into the level-1 equation to formulate a single, combined equation:

$$y_{ij} = \beta_{0j} + \beta_{1j}x_{1ij} + e_{ij}$$

$$= (\gamma_{00} + \gamma_{01}z_{1j} + u_{0j}) + (\gamma_{10} + \gamma_{11}z_{1j} + u_{1j})x_{1ij} + e_{ij}$$

$$= \underbrace{[\gamma_{00} + \gamma_{01}z_{1j} + \gamma_{10}x_{1ij} + \gamma_{11}z_{1j}x_{1ij}]}_{\text{Fixed part}} + \underbrace{[u_{0j} + u_{1j}x_{1ij} + e_{ij}]}_{\text{Random part}}$$

The first line shows the level-1 equation. The second line substitutes the level-2 equations for the level-1 regression coefficients,  $\beta_{0j}$  and  $\beta_{1j}$ . The third line rearranges the terms in the second line into a “fixed” part and a “random” part. The fixed part contains the “fixed” (constant or deterministic) coefficients and the random (or stochastic) part contains the random error terms.

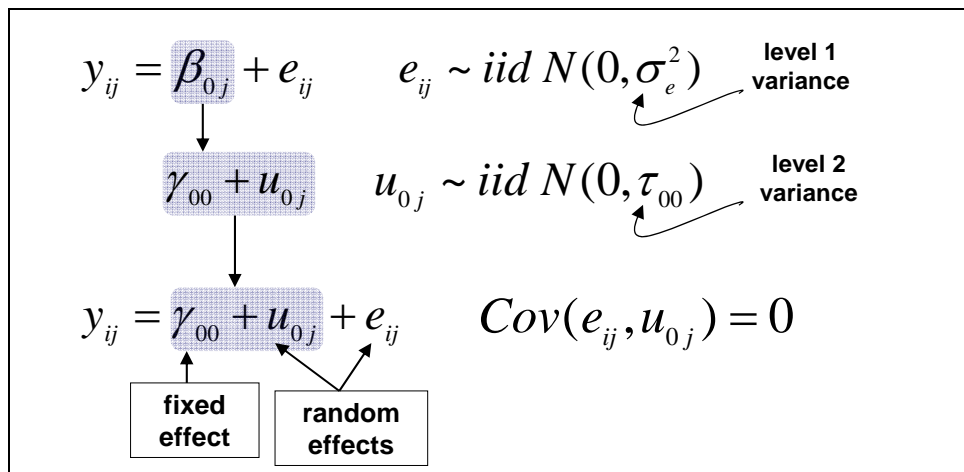
The fixed part of this combined model looks like an ordinary regression, with one level-2 variable ( $z$ ), one level-1 variable ( $x$ ), and an interaction between the two variables ( $z \cdot x$ ). The random part is what distinguishes it as a hierarchical model. It has the usual random error ( $e$ ) corresponding to level-1 residuals, but it also has two other error terms ( $u_0$  and  $u_1$ ) associated with the level-2 residuals. This combined model is commonly called a “mixed model” because it includes a mixture of fixed and random effects.

It is important to understand how to construct both the level-1 and the level-2 regressions, as well as how to construct the combined equation, because some computer programs for HLM require that regressions be specified at each level individually, while others require that the combined equation be specified.

The discussion now turns to some special cases of the two-level model.

Random Intercepts (Empty Model or Variance Components Model)

The simplest non-trivial specification for the random part of a two-level model is a model in which only the intercept varies between Level-2 units and no predictor variables are entered, either for level-1 units or for level-2 units. This model is variously called the empty, random intercept, or variance component model:



An important statistic from the empty model is the intraclass correlation coefficient (ICC):

$$ICC = \frac{\tau_{00}}{\tau_{00} + \sigma_e^2}$$

which is the proportion of total variance that is accounted for at the group level. If there is no between-group variance, then the ICC is zero and a hierarchical model is unnecessary, assuming that the model is correct. The ICC can also be interpreted as the correlation between level-1 units randomly paired within groups (Snijders and Bosker, 1999).

As an example, this empty model is fit to the logarithm of total charges (log(charges)) for the subset of diabetes patients in the 2003 NIS using SAS PROC MIXED. The program code is contained in the chapter on statistical software later in this report. The results are as follows:

Fixed Effect	Coefficient	S.E	t-value	Probability
Intercept (log \$)	$\gamma_{00} = 9.0569$	0.0205	441.52	.000
Random Effect	Variance Component	S.E	z-value	Probability
Level 1 residual $e_{ij}$	$\sigma^2 = 0.6838$	0.0031	221.13	.000
Level 2 residual $u_{0j}$	$\tau_{00} = 0.3735$	0.0187	19.97	.000

The overall hospital average of log(charges) is 9.06. The level-1 (patient) error variance is estimated as 0.6838 and the level-2 (hospital) variance is estimated to be 0.3735, which is statistically significant (z-value = 19.97, prob. < .001). The estimated intraclass correlation coefficient is  $0.35 = 0.3735 / (0.3735 + 0.6838)$ , indicating that over one-third of the total variance in log(charges) is between hospitals. Of course, the ICC might be reduced by adding explanatory variables to explain more of the patient-level variation.

From this model, the average log(charge) for hospital  $j$  can be estimated by an empirical Bayes “shrinkage estimator,” which is a weighted average of the hospital’s observed mean,  $\bar{y}_j$ , and the estimated overall hospital mean,  $\hat{\gamma}_{00}$ :

$$\hat{\beta}_{0j} = \hat{\lambda}_j * \bar{y}_j + (1 - \hat{\lambda}_j) * \hat{\gamma}_{00}$$

$$\text{weight} = \hat{\lambda}_j = \left( \frac{\hat{\tau}_{00}}{\hat{\tau}_{00} + \hat{\sigma}_e^2 / n_j} \right) = \text{reliability}$$

Here,  $n_j$  is the sample size (number of discharges) for hospital  $j$ . The weight, which lies between zero and one, is also an estimate of the reliability of the hospital’s observed mean (Snijders and Bosker, 1999). As the reliability approaches zero, the hospital’s estimate “shrinks” toward the overall mean. For this reason, it is an example of a “shrinkage estimator.” Intuitively, the reliability is near zero if the variance of the hospital’s mean (second term in the denominator) is large or if the between-hospital variance is large (indicating evidence of wide variation among hospitals).

Although the empirical Bayes estimate is biased, it is (on average) closer than the observed mean to the hospital’s true mean,<sup>5</sup> and it is often a preferred basis for ranking hospitals because it tends to draw small hospitals with unreliable estimates closer to the overall mean. Keep in mind that the null model assumes that each true hospital average is drawn from a population of hospital averages described by the normal distribution with mean  $\gamma_{00}$  and variance  $\tau_{00}$ . The shrinkage estimate takes advantage of this assumption to sharpen each hospital’s estimate. This is sometimes called “borrowing strength” from the whole (of all hospital means).

As an example, suppose a hospital’s average observed log(charge) was 8.67 based on a sample of 10 discharges. Then based on the above results the reliability would be:

$$.85 = \left( \frac{.3735}{.3735 + .6838/10} \right)$$

and the empirical Bayes estimate for the hospital would be:

$$8.73 = .85 * 8.67 + (1 - .85) * 9.06.$$

The hospital’s observed mean of 8.67 is pulled 15% of the way toward the overall hospital mean of 9.06, for an empirical Bayes estimate of 8.73. If the hospital’s mean was based on  $n = 10,000$

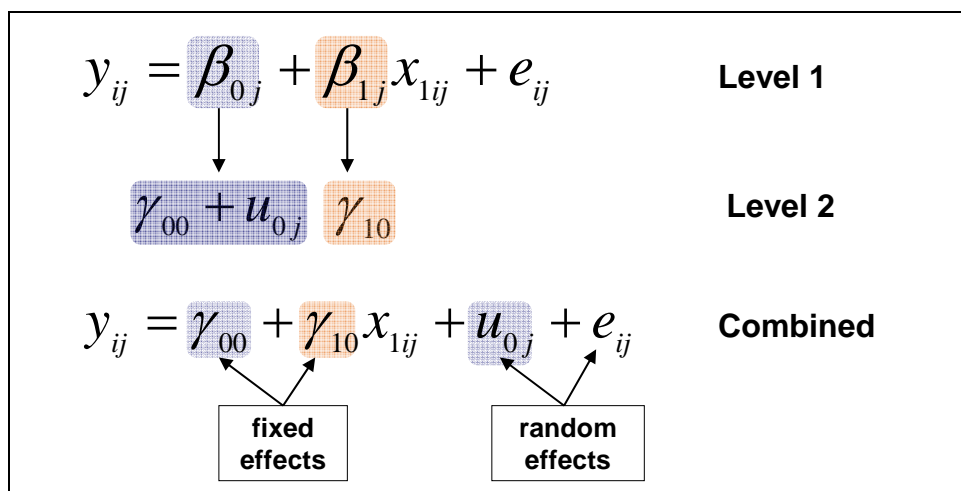
---

<sup>5</sup>Under standard assumptions, the empirical Bayes estimate has a lower mean squared error (MSE).

discharges, then the reliability would be nearly equal to 1, and the empirical Bayes estimate would be virtually the same as the hospital's observed mean of 8.67.

### Random Intercepts, Fixed (Constant) Slopes

This model, graphically exemplified by Figure 7(b), adds a “fixed” explanatory variable,  $x_1$ , to the empty model.



In the combined model, the regression coefficient  $\beta_{1j}$  is simply renamed  $\gamma_{10}$  to preserve the general notation for the combined models. The coefficient just as easily could have been renamed  $\beta_1$ . In this particular model, the subscript  $j$  is unnecessary because the slope is constant across level-2 units. The ICC can still be calculated for this model, and it is called the “residual” ICC controlling for the effect of  $x_1$ . It represents the group variance that remains after accounting for the effect of  $x_1$ . The addition of more level-1 explanatory variables is straightforward.

Continuing the example for NIS diabetes patients in the 2003 NIS, age is added as a patient-level variable, with the following results:

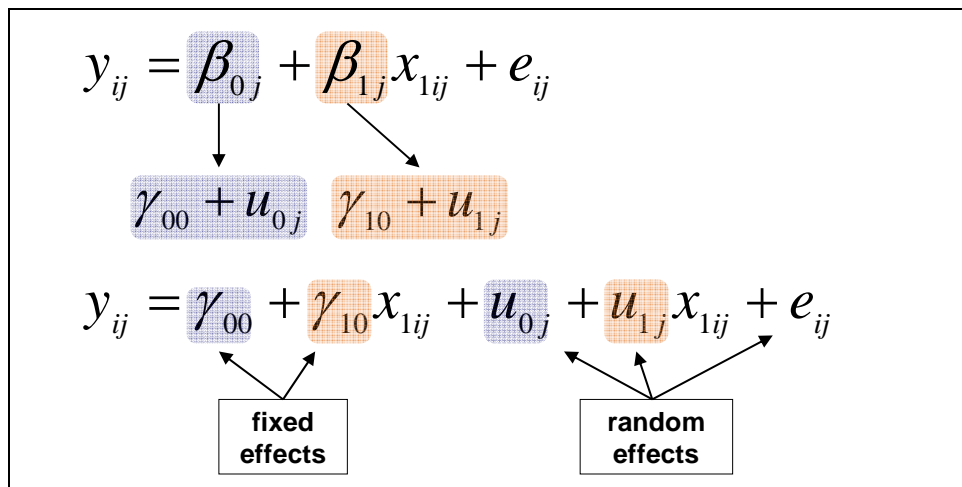
Fixed Effect	Coefficient	S.E	t-value	Probability
Intercept (log \$)	$\gamma_{00} = 8.9481$	0.0207	433.43	< .0001
Age (decades over the mean)	$\gamma_{01} = 0.0713$	0.0014	52.99	< .0001
Random Effect	Variance Component	S.E	z-value	Probability
Level 1 residual, $e_{ij}$	$\sigma^2 = 0.6647$	0.0030	221.05	< .0001
Level 2 residual, $u_{0j}$	$\tau_{00} = 0.3754$	0.0188	19.98	< .0001

The age coefficient, which is a “fixed” effect, is statistically significant. Age was constructed to be centered at the overall mean age, divided by 10, so age is measured in decades. The average of log(charges) is estimated to increase by about 7 percent for every decade increase

in age. Notice that the estimates of variance components changed very little compared with the empty model. The patient-level variance decreased slightly, from 0.68 to .66, with the addition of age to the model. Although the hospital-level variance did not decline, it could have declined if considerable hospital-to-hospital variation existed in the patient age distribution. There are more patient-level variables, such as severity of diabetes, that could be added to explain more variation. However, this example is only meant to illustrate the methodology rather than provide a comprehensive analysis.

### Random Intercepts, Random Slopes

This model allows both the intercept and the slope to vary across level-2 units, as shown graphically in Figures 7(c) through 7(f).



Similar to the intercept, the slope is modeled by a fixed effect,  $\gamma_{10}$ , representing the overall average slope across level-2 units, and a random effect,  $u_{1j}$ , representing the group-specific slope residual for each level-2 unit. Notice that the level-2 slope residual,  $u_{1j}$ , is multiplied by  $x_1$ , resulting in a total error that varies according to the value of  $x_1$ . This is a strong assumption that should be confirmed in practice.

For the 2003 NIS diabetes subset, the following results were obtained by specifying age as a random effect:

Fixed Effect	Coefficient	S.E	t-value	Probability
Intercept,	$\gamma_{00} = 8.9617$	0.0207	433.84	< .0001
Age (decades over the mean)	$\gamma_{01} = 0.0636$	0.0021	30.37	< .0001
Random Effect	Variance Component	S.E	z-value	Probability
Level 1 residual $e_{ij}$	$\sigma^2 = 0.6589$	0.0030	220.42	< .0001
Level 2 residual $u_{0j}$	$\tau_{00} = 0.3729$	0.0191	19.51	< .0001

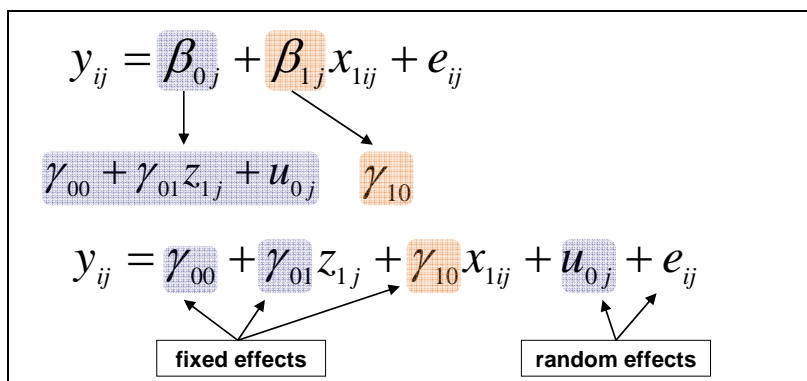


Level 2 age effect (decades over the mean), $u_{1j}$	$\tau_{11} = 0.0014$	0.00015	9.58	< .0001
Cov( $u_0, u_1$ )	$\tau_{01} = -0.0004$	0.0014	-0.30	.7666

The estimated coefficient for age, .0636, now represents the *average* increase in log(charges) across hospitals per decade of patient age. Notice that this coefficient (.0636) is different from the fixed coefficient for age in the previous model (.0713). The component of variance for age is significant, indicating that the age coefficient does vary from hospital to hospital. The other variance components are very slightly smaller as a result. Although the variance of the age coefficient seems small, the standard deviation (the square root of the variance) is about .04, which is fairly large relative to the mean coefficient of .06. Based on the assumption that the coefficients are normally distributed, one would expect about 95 percent of the hospital-specific age coefficients to fall in the range -.02 to .10. Finally, the covariance between the random intercepts and slopes is -.0004, which is not significantly different from zero ( $p$ -value = .77). Therefore, one would infer that there is no correlation between the hospital-specific intercepts and slopes, based on this model.

### Intercept- and Slope-As-Outcomes

The intercept as outcomes model allows the intercept to vary as a function of one or more level-2 variables. For example, the intercept for teaching hospitals might differ from the intercept for non-teaching hospitals. In the example below, the variable  $z_1$  would be teaching status (1 for teaching hospitals and 0 for non-teaching hospitals), and the average outcome for teaching hospitals would differ by an amount  $\gamma_{01}$  compared with non-teaching hospitals. The hospital-level residual,  $u_{0j}$ , is conditional on teaching status. In general,  $z_1$  could be any discrete or continuous level-2 variable.



This model merely adds a fixed hospital effect, so it is not very different from the random intercept, fixed slope model discussed in the previous section. Nevertheless, the addition of a group variable can reduce the hospital-level residual variance. Again,  $\beta_1$  could have been substituted for  $\beta_{1j}$ . However,  $\gamma_{10}$ , is substituted to preserve the general notation in the combined model.

If both the intercept and the slope are a function of a group variable, then the intercept- and slopes-as-outcomes model results (Raudenbush and Bryk, 2002):

$$\begin{aligned}
y_{ij} &= \beta_{0j} + \beta_{1j}x_{1ij} + e_{ij} \\
&= (\gamma_{00} + \gamma_{01}z_{1j} + u_{0j}) + (\gamma_{10} + \gamma_{11}z_{1j} + u_{1j})x_{1ij} + e_{ij} \\
&= [\gamma_{00} + \gamma_{01}z_{1j} + \gamma_{10}x_{1ij} + \gamma_{11}z_{1j}x_{1ij}] + [u_{0j} + u_{1j}x_{1ij} + e_{ij}] \\
&\quad \underbrace{\hspace{10em}}_{\text{Fixed part}} \quad \underbrace{\hspace{10em}}_{\text{Random part}}
\end{aligned}$$

This model was fit to the diabetes patients in the 2003 NIS, using age as the patient-level variable and teaching status as the hospital-level variable. The results are as follows:

Fixed Effect	Coefficient	S.E	t-value	Probability
Intercept (log \$)	$\gamma_{00} = 8.8802$	0.0219	405.37	< .0001
Age (decades over the mean)	$\gamma_{10} = 0.0590$	0.0025	23.94	< .0001
Teaching hospital	$\gamma_{01} = 0.4684$	0.0506	9.26	< .0001
Age * Teaching hospital	$\gamma_{11} = 0.0169$	0.0046	3.71	< .0001
Random Effect	Variance Component	S.E	z-value	Probability
Level 1 residual $e_{ij}$	$\sigma^2 = 0.6589$	0.0029	220.43	< .0001
Level 2 residual $u_{0j}$	$\tau_{00} = 0.3374$	0.0171	19.74	< .0001
Level 2 age effect (decades over the mean), $u_{1j}$	$\tau_{11} = 0.0014$	0.0001	9.41	< .0001
Cov( $u_0, u_1$ )	$\tau_{01} = 0.6589$	0.0029	220.43	< .0001

The variance components for both the intercept and age are statistically significant, indicating significant hospital-to-hospital variation in those coefficients. The estimated log(charges) for teaching hospitals are higher by 0.47 compared with non-teaching hospitals. Also, the estimated age coefficient for teaching hospitals is 0.076, calculated as the sum of 0.017 and 0.059. Notice that, in the regression results shown in the previous section, the overall mean age coefficient is .063, which falls between the non-teaching age coefficient of .059 and the teaching age coefficient of .076 in the present regression. Notice, too, that the addition of the teaching variable slightly reduced the hospital-level residual variance for the intercepts, as compared with that estimated for the random intercept and random slope model in the previous section. It is possible that the addition of more hospital-level variables would further reduce the hospital-level variance.

## HLM for Binary Outcomes (logistic regression)

This section describes only the most general version of the two-level HLM with one level-1 predictor and one level-2 predictor. Special cases of this model, such as the null model, the random intercept model, the random slope model, and so on, are completely analogous to those cases for the continuous HLM.

Binary outcomes—those taking on only two values—are commonplace in health services research. Examples include in-hospital mortality, the occurrence of a medical error, admission through the ER (versus treat-and-release), surgical treatment (versus non-surgical treatment), and whether a patient is readmitted to the hospital within 30 days of a prior discharge. These examples also demonstrate the potential importance of analyzing the sources of variation for these kinds of outcomes. If analysts can learn which patient and institutional factors distinguish between successes and failures, then those factors might lead to policy instruments that benefit patients in the future.

Binary outcomes are often analyzed using logistic regression, in which the predicted log-odds of an event is a linear model. The hierarchical logistic model asserts a logistic regression as the level-1 model:

$$\ln\left(\frac{\pi_{ij}}{1 - \pi_{ij}}\right) = \beta_{0j} + \beta_{1j}x_{1ij}$$

$$y_{ij} \sim \text{Bernoulli}(\pi_{ij})$$

In this case, the “link function” is called the logit. The Bernoulli distribution is a special case of the binomial distribution that takes on only the values 0 and 1. The variance of the Bernoulli distribution,  $\pi_{ij}(1 - \pi_{ij})$ , is a function of the mean,  $\pi_{ij}$ . Consequently, the specification of a level-1 “residual” error term is unnecessary.

The general setup for the two-level hierarchical logistic model parallels that for the continuous model. Importantly, the statistical models and assumptions for the second-level regressions are precisely the same as those described above for the continuous model. The second-level random effects (group-level residuals) are assumed to be normally distributed, and the second-level residuals are assumed to be independent of the first-level residuals. However, in the first-level models, the residuals are assumed to be distributed as Bernoulli. Consequently, the combined model is:

$$\ln\left(\frac{\pi_{ij}}{1 - \pi_{ij}}\right) = \gamma_{00} + \gamma_{01}z_{1j} + \gamma_{10}x_{1ij} + u_{0j} + u_{1j}x_{1j}$$

$$y_{ij} \sim \text{Bernoulli}(\pi_{ij})$$

$$\begin{pmatrix} u_{0j} \\ u_{1j} \end{pmatrix} \sim \text{iid } N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \tau_{00} & \tau_{01} \\ \tau_{10} & \tau_{11} \end{pmatrix}\right] \quad \forall j$$

Maximum likelihood is the method used to estimate the coefficients for this model. It should be noted that this model can be re-written in terms of the expected odds-ratio:

$$\frac{\pi_{ij}}{1 - \pi_{ij}} = e^{(\gamma_{00} + \gamma_{01}z_{1j} + \gamma_{10}x_{1ij} + u_{0j} + u_{1j}x_{1j})} = e^{\gamma_{00}} \cdot e^{\gamma_{01}z_{1j}} \cdot e^{\gamma_{10}x_{1ij}} \cdot e^{u_{0j}} \cdot e^{u_{1j}x_{1j}}$$

which illustrates the multiplicative form of the model.

## HLM for Count Outcomes

Again, this section describes only the most general version of the two-level HLM with one level-1 predictor and one level-2 predictor. Special cases of this model, such as the null model, the random intercept model, the random slope model, and so on, are completely analogous to those cases for the continuous HLM.

Hierarchical models for count outcomes are often based either on the Poisson distribution or the negative binomial distribution, leading to Poisson regression or negative binomial regression models, respectively. The range of outcomes for both distributions is the set of natural numbers 0, 1, 2, 3, 4,.... The Poisson distribution is commonly used for modeling rates of low-frequency events, such as mortality counts for low-mortality diseases. A fundamental assumption for the Poisson model is that the mean is equal to the variance. This assumption is often untenable. Consequently, the negative binomial distribution is the preferred model for count outcomes for which the variance exceeds the mean. In fact, the Poisson distribution is a special case of the negative binomial distribution, enabling analysts to decide between the two models based on the value of an estimated parameter.

Again, the general setup for the two-level hierarchical count model parallels that for the continuous model. The statistical models and assumptions for the second-level regressions are the same as those for the continuous model. However, in the first-level models, the residuals are assumed to be distributed as Poisson or negative binomial.

### Poisson Regression

For Poisson regression, the dependent variable is the logarithm of the population mean count,  $\theta_{ij}$ , given that level-1 unit  $i$  is contained in level-2 unit  $j$ , and given the value of any explanatory variables. In this context, the logarithm is called the “link function” because it links the mean count to the level-1 linear equation:

$$\ln(\theta_{ij}) = \beta_{0j} + \beta_{1j}x_{1ij}$$
$$y_{ij} \sim \text{Poisson}(\theta_{ij})$$

The assumption that the observed counts,  $y_{ij}$ , are distributed as Poisson with mean  $\theta_{ij}$ , emphasizes the difference between this level-1 equation and the level-1 equation for the continuous (normally-distributed) outcome. This level-1 equation is written in terms of the population mean (expected value) of the response, which requires no error term. Whereas, the continuous version is written in terms of the observed response, which includes a normally-distributed residual error term to explain the difference between the observed and expected response.

The assumption that the observed counts  $y_{ij}$  are distributed as Poisson with mean  $\theta_{ij}$  is used to estimate the regression coefficients and other parameters by the method of maximum likelihood. For the Poisson distribution, the mean and the variance are equal. Therefore, the level-1 error variance is not constant for each observation, as it is in the continuous case. Instead, it fluctuates with the value of the mean.

The level-2 equations and the statistical assumptions for the level-2 equations are precisely the same as those given above for the continuous case. Therefore, the combined model becomes:

$$\begin{aligned} \ln(\theta_{ij}) &= \gamma_{00} + \gamma_{01}z_{1j} + \gamma_{10}x_{1ij} + u_{0j} + u_{1j}x_{1j} \\ y_{ij} &\sim \text{Poisson}(\theta_{ij}) \\ \begin{pmatrix} u_{0j} \\ u_{1j} \end{pmatrix} &\sim \text{iid } N \left[ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \tau_{00} & \tau_{01} \\ \tau_{10} & \tau_{11} \end{pmatrix} \right] \quad \forall j \end{aligned}$$

The “fixed” effects are  $\gamma_{00}$  (average intercept),  $\gamma_{01}$  (coefficient on the group variable  $z_1$ ), and  $\gamma_{10}$  (average coefficient on the level-1 variable  $x_1$ ). The random effects are  $u_{0j}$  and  $u_{1j}$ , corresponding to the random intercepts and random slopes, respectively.

Because the dependent variable is the logarithm of the expected count, each regression coefficient can be exponentiated to estimate the multiplicative effect on the mean of the corresponding predictor variable. For example, an age coefficient of .05 would translate into a multiplicative effect of approximately 1.05 ( $= e^{.05}$ ), meaning that the average count increases by about 5 percent for each year of age.

### Negative Binomial Regression

Most HLM computer programs optionally allow the user to specify the negative binomial distribution for the response. The main difference between the Poisson model and the negative binomial model is the assumption concerning the distribution of counts. The second-level models and assumptions are the same. However, the counts  $y_{jk}$  are assumed to be distributed as a negative binomial with mean  $\theta_{ij}$  and dispersion parameter  $k$ :

$$\begin{aligned} \ln(\theta_{ij}) &= \beta_{0j} + \beta_{1j}x_{1ij} \\ y_{ij} &\sim \text{Negative Binomial}(\theta_{ij}, k) \end{aligned}$$

Again, the link function is the logarithm, which implies multiplicative effects. The negative binomial distribution is often used in situations where the counts have a larger variance than that implied by the Poisson distribution.<sup>6</sup>

The level-2 models are the same as those specified for the continuous outcome. Therefore, the combined model is:

$$\begin{aligned} \ln(\theta_{ij}) &= \gamma_{00} + \gamma_{01}z_{1j} + \gamma_{10}x_{1ij} + u_{0j} + u_{1j}x_{1j} \\ y_{ij} &\sim \text{Negative Binomial}(\theta_{ij}, k) \\ \begin{pmatrix} u_{0j} \\ u_{1j} \end{pmatrix} &\sim \text{iid } N \left[ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \tau_{00} & \tau_{01} \\ \tau_{10} & \tau_{11} \end{pmatrix} \right] \quad \forall j \end{aligned}$$

The coefficients are estimated by the method of maximum likelihood.

### **HLM for Longitudinal Data**

<sup>6</sup>In particular, if the counts are distributed as Poisson conditional on means that are distributed as gamma, then the resulting counts are unconditionally distributed as negative binomial. Fully Bayesian HLM sometimes uses this derivation directly in specifying the hierarchical model for counts.

A special case of a hierarchical model is one in which the level-1 units are repeated measures taken on each level-2 unit. For example, suppose the data consisted of visits to the emergency department for asthma treatment, and researchers were interested in the probability that each visit resulted in a hospital admission (versus treat and release). Some patients would have multiple asthma ED visits, and these would comprise repeated measures for those patients. Further, suppose that time is thought to be an important factor. These data could be modeled using a hierarchical model with asthma ED visits at level 1 and patients at level 2.

Importantly, an HLM could be fit using all of the asthma ED visits for all of the patients, even though patients have different numbers of ED visits (occasions) that occur at irregular times. There are few opportunities for these types of models using HCUP data. Therefore, readers are referred to the excellent text on the subject by Singer and Willett (2003) and to Singer's Website (<http://gseweb.harvard.edu/%7Efaculty/singer/>) for more information.

### **HLM for Cross Classifications**

As illustrated earlier in Figure 3, the group structure might not be strictly hierarchical, and cross-classifications are an example of a more complicated structure. Figure 3 shows an example of patients nested within hospitals and physicians, but physicians are not nested within hospitals. Both physicians and hospitals can have an effect on patient outcomes, so it remains important to estimate their effects. It also remains important to account for the potential correlation among patients treated by the same physician and treated in the same hospital.

Cross-classifications, also called crossed random coefficients, can be fit in the HLM framework with a properly-specified model. The simplest specification for the patient/physician/hospital structure is as follows:

$$y_{i(jk)} = \alpha + \mu_j + \mu_k + e_{i(jk)}$$

where the outcome  $y_{i(jk)}$  for patient  $i$  treated by physician/hospital combination  $jk$  is modeled by an overall mean  $\alpha$  with random departures (residuals)  $\mu_j$  and  $\mu_k$  for physician  $j$  and hospital  $k$ , respectively, and a patient-level random departure  $e_{i(jk)}$ . Other terms involving patient, physician, and hospital factors can be added to this basic model.

Readers interested in these models should consult the references provided at the end of this report. In particular, Hox (2002), Snijders and Bosker (1999), and Rasbash et al. (2005) all have informative chapters on this topic.

### **HLM for Multiple Membership**

Multiple membership, as illustrated earlier in Figure 4, is also an example of a more complicated grouping structure. In that example, some patients are treated by more than one physician. Patients are clustered or grouped by physician. Therefore, some patients belong to multiple groups and they are said to have multiple memberships. More generally, multiple memberships exist whenever lower-level units belong to two or more higher-level units.

An interesting aspect of these models is that for each level-1 unit, the analyst must specify weights for the level-2 units. The choice of weights might or might not be straightforward. As an example, for a patient with multiple physicians, the weight for each physician could be the proportion of the patient's visits to that physician. Alternatively, the weight could be the

proportion of the patient's total treatment costs associated with that physician, or the proportion of the year that physician served as the patient's primary care provider.

Readers interested in pursuing these models can check the references and Websites given later in this report. Rasbash et al. (2005) provide a nice introductory chapter in the MLwiN User's Guide.

## **Other Methods**

### Generalized Estimating Equations (GEE)

The Generalized Estimating Equation (GEE) approach is one method that accounts for within-cluster correlations by modelling the correlations rather than by modelling random effects, as hierarchical models do. Different correlation structures can be specified for GEE models. HLM models the outcome conditional on the random effects (or random coefficients), while GEE methods model the marginal expectation of the outcome "averaged" over the population.

Consequently, the interpretation of regression coefficients are slightly different between HLM and GEE models. In HLM, the regression coefficient is an estimate of how the outcome changes as a function of predictors conditional on the random effects. In GEE, the coefficient is an estimate of how the outcome changes as a function of predictors "averaged" over the group random effects. For continuous outcomes, the predictor coefficients are mathematically equivalent, but for count and binary variables, the GEE estimates will usually be smaller than the corresponding HLM estimates (in absolute value).

### Bayesian Methods

Fully Bayesian hierarchical models specify probability distributions, called *prior* distributions, for the population parameters, such as the population intercept,  $\gamma_{00}$ . The prior distribution is combined with the observed data to produce a *posterior* distribution, which forms the basis for inferences. By contrast, the "frequentist" or "classical" approach explained in this report considers the population parameters to be fixed, unknown constants.

Bayesian methods are usually more demanding, both analytically and computationally. However, correctly specified Bayesian models provide more precision and better estimates of variances in exchange for a small amount of bias (Hox, 2002). Using a Bayesian model, it is not possible to estimate negative variances, which is possible with standard hierarchical models. Further, when the number of higher-level units is small, estimates and statistical inferences from frequentist hierarchical models may be unreliable because the underlying methods rely on large-sample theory. In those cases, especially when the group sample sizes are highly variable, Bayesian methods provide an especially attractive alternative. A brief, readable introduction to Bayesian hierarchical models is contained in Raudenbush and Bryk (2002, Chapter 13). Browne and Draper (2005) compare Bayesian and likelihood-based methods. These models require special software, such as WinBUGS, which is briefly described later in this report.

## **Using Sampling Weights in HLM**

For complex sample survey data, methods for incorporating sampling weights in hierarchical models is a subject of active research in the literature. Although HCUP data are not derived

from survey data, the NIS and the KID are the results of sampling procedures. Therefore, these methods could be applied to the NIS and the KID.

That said, the authors of this report do not recommend these methods for HCUP data. One reason is philosophical. It seems incongruous to incorporate finite population weights into models that specifically consider these same populations to be conceptually infinite. Another reason stems from practicality. It complicates an already-complicated estimation procedure, possibly for little or no gain. For example, when sampling weights were incorporated into an analysis of NIS data, the results with and without the weights were virtually indistinguishable from one another.

For analysts who are concerned about the potential effects of sample stratification, one approach is to include the stratifiers as predictors in the model. For example, when using the NIS, hospital-level variables like teaching status and bed size can be incorporated into the model, and their effects can be estimated.

A good reference for sampling weights in HLM is the article by Rabe-Hesketh and Skrondal (2005b). Also, more information and other references can be obtained at the following Websites: [http://www.cpc.unc.edu/restools/data\\_analysis/ml\\_sampling\\_weights](http://www.cpc.unc.edu/restools/data_analysis/ml_sampling_weights) <http://www.statmodel.com/download/asparouhovgmms.pdf>.

### **HLM Compared to Conventional Regression: An Example Using HCUP Data**

This section is loosely based on an analysis by Joshi, D'Souza, and Madhavan (2004) of hospital charges, mortality, and length of stay for congestive heart failure (CHF) patients, which used a 10 percent sample of the 1997 Nationwide Inpatient Sample (NIS). The authors used conventional regression models in their analyses, which included both patient-level and hospital-level predictors. This example does not assess the fit, check the assumptions, or question the specifications of the models. For this example, similar (but not identical) models are fit for total charges and in-hospital mortality using the complete 2004 NIS. These models are fit once using a conventional regression model (disaggregated analysis) and once using a random-intercept model with discharges at level 1 and hospitals at level 2. It is intended only to demonstrate how estimates and subsequent inferences can differ between the two approaches.

#### Data

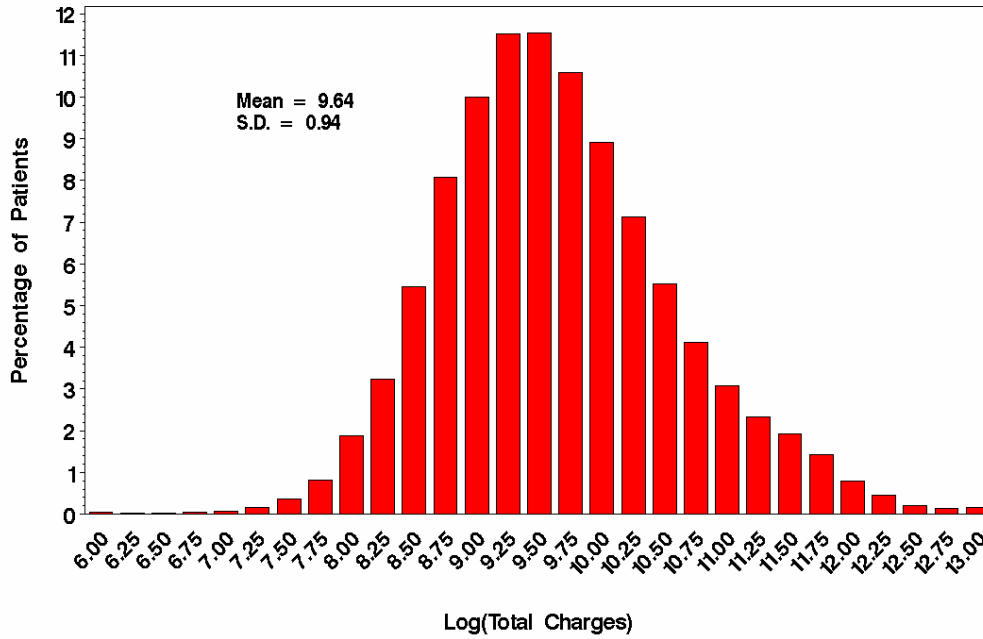
For this analysis, all CHF discharges (principal diagnosis of 428.xx) were selected from the 2004 NIS. This resulted in a total of 222,200 discharges from 978 hospitals. Hospital-level variables were merged onto this file from the 2004 NIS hospital file that accompanies the 2004 NIS. Patient severity measures were also merged onto this file from the 2004 Severity file that accompanies the 2004 NIS.

The logarithm of total charges is the dependent variable for the charge regressions. The distribution of  $\log(\text{charges})$  is shown in Figure 8, below. The distribution is nearly symmetrical, but very slightly skewed to the right, with a mean of 9.64 and a standard deviation of 0.94.

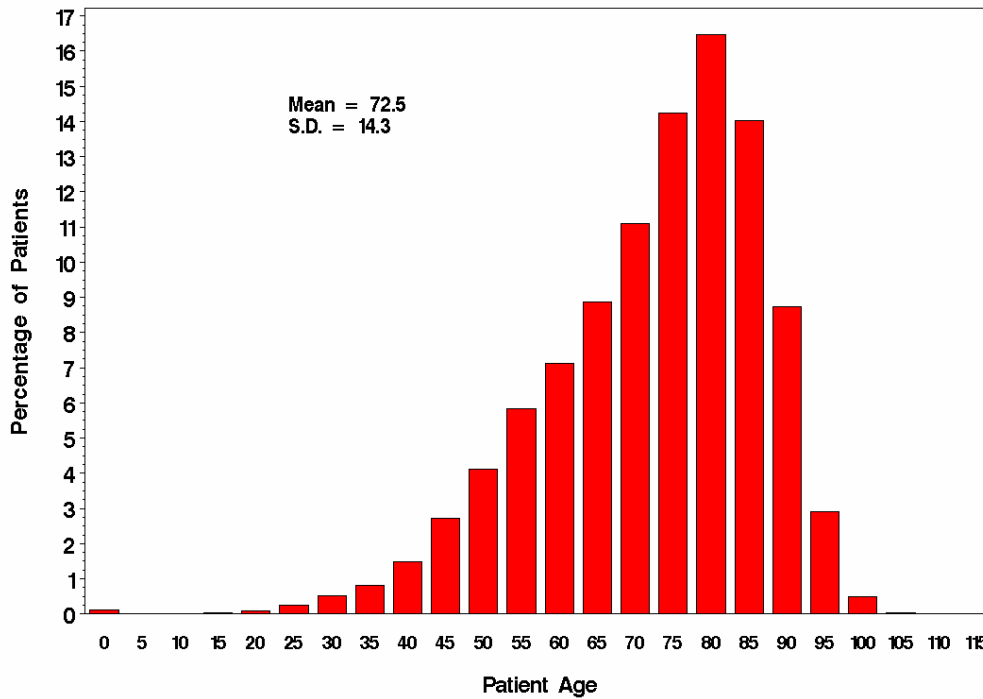
One of the patient-level predictors is age. The age distribution is shown in Figure 9, which is highly skewed, as one would expect for a sample of CHF patients. The average age is 72.5 with a standard deviation of 14.3.



**Figure 8:** Distribution of Logarithm of Total Charges for CHF Patients, 2004 NIS



**Figure 9:** Distribution of Age for CHF Patients, 2004 NIS



Other predictor variables are as follows:

Log(Resource Demand Scale) – This is the logarithm of the Disease Staging Resource Demand Scale, which will be used in the model for total charges. This is an inpatient severity measure that predicts hospital charges as a function of patient characteristics including age, sex, diagnosis related group (DRG), the patient’s principal disease, the severity of the principal disease, the patient’s comorbidities, the severity of comorbidities, the admission source, and the urgency of admission. This scale predicts the charge for an admission as a percentage of the average charge for all admissions to the hospital. A scale value of 100 corresponds to the cost of an average admission. The average scale value for CHF patients is 129.31 with a standard deviation of 135.60. The average value log(scale) value for CHF patients is 4.66 with a standard deviation of 0.54.

Log(Mortality Scale) – This is the logarithm of the Disease Staging Mortality Scale, which will be used in the model for in-hospital mortality. This inpatient severity measure predicts hospital mortality as a function of patient characteristics including age, sex, diagnosis related group (DRG) for surgical cases (not used for non-surgical cases), the patient’s principal disease, the severity of the principal disease, the patient’s comorbidities, the severity of comorbidities, the admission source, and the urgency of admission. This scale predicts the mortality risk for an admission as a percentage of the average mortality risk for all admissions. A scale value of 100 corresponds to the mortality rate of an average admission. The average mortality scale value for CHF patients is 287.21 with a standard deviation of 444.61. The average value of the log(mortality scale) value for CHF patients is 4.87 with a standard deviation of 1.30.

Age – 70 (decades) – This is calculated as (age – 70) / 10, so that age is centered at 70 and measured in decades.

Female – This is an indicator for gender. Female = 1 for females and = 0 for males.

Median ZIP Income – This is a four-category variable designating the median income for the patient’s ZIP Code. The categories are 1) under \$36,000, 2) \$36,000 - \$44,999, 3) \$45,000 – \$59,999, and 4) \$60,000 and up. The income group frequencies are as follows:

ZIP Income Quartile	Frequency	Percent	Cumulative Frequency	Cumulative Percent
<b>Missing</b>	4,482	2.02	4,482	2.02
<b>Q1: &lt; \$ 36,000</b>	74,565	33.56	79,047	35.57
<b>Q2: \$36,000 - 44,999</b>	58,740	26.44	137,787	62.01
<b>Q3: \$45,000 - 59,999</b>	44,931	20.22	182,718	82.23
<b>Q4: \$60,000 +</b>	39,482	17.77	222,200	100.00

Expected Primary Payer – This is a categorical variable with frequencies as follows. Medicare is the omitted category in the regressions.

Primary Payer	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Missing/Invalid	185	0.08	185	0.08
1: Medicare	168,359	75.77	168,544	75.85
2: Medicaid	15,899	7.16	184,443	83.01
3: Private insurance	27,400	12.33	211,843	95.34
4: Self-pay	6,289	2.83	218,132	98.17
5: No charge	761	0.34	218,893	98.51
6: Other	3,307	1.49	222,200	100.00

Hospital Region – This is a four-category variable with frequencies as follows. West is the omitted category in the regressions.

Region	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Northeast	44,230	19.91	44,230	19.91
Midwest	51,623	23.23	95,853	43.14
South	96,196	43.29	192,049	86.43
West	30,151	13.57	222,200	100.00

Hospital Bed Size and Location/Teaching Status – Hospitals in the NIS are classified as small, medium, or large bed sizes separately for rural, urban teaching, and urban non-teaching hospitals. This combination is treated as a nine-category variable. Large teaching hospital is the omitted category in the regressions.

Bed Size	Location/Teaching	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1: Small	1: Rural	7,027	3.16	7,027	3.16
1: Small	2: Urban non-teaching	14,843	6.68	21,870	9.84
1: Small	3: Urban teaching	11,122	5.01	32,992	14.85
2: Medium	1: Rural	8,627	3.88	41,619	18.73
2: Medium	2: Urban non-teaching	28,790	12.96	70,409	31.69
2: Medium	3: Urban teaching	19,789	8.91	90,198	40.59
3: Large	1: Rural	22,445	10.10	112,643	50.69
3: Large	2: Urban non-teaching	61,257	27.57	173,900	78.26
3: Large	3: Urban teaching	48,300	21.74	222,200	100.00

## Regressions for Log(Charges)

Both a conventional regression, labeled “OLS,” and a hierarchical regression, labeled “HLM,” are fit to the CHF sample. For the HLM, discharges are at level 1, hospitals are at level 2, and a random-intercept model is fit using SAS PROC MIXED to produce restricted maximum likelihood estimates (the default in SAS). For the OLS model, the R-square value was 40 percent. For an empty HLM model (not shown), the random intercepts variance was 0.4012, while the residual variance was 0.5982, resulting in an intra-class correlation of 0.40, indicating that 40 percent of the total variance is between hospitals. For the HLM model with covariates, the intercept variance fell to 0.1849 and the residual variance fell to 0.3589, indicating that the predictor variables explained a large fraction of the variance at both the patient and hospital levels.

The results for the “fixed” effects are shown in Table 1. The shaded rows correspond to effects that are statistically significant at the 5 percent level using conventional regression, but not significant at the 5 percent level using HLM. The estimated coefficients and their standard errors are in close agreement between the two estimation methods for severity, patient age, and female gender. However, based on the OLS results, there are several instances where one would infer that the average log(charges) differ significantly between types of patients or types of hospitals. For example, under OLS, large urban teaching hospitals (the omitted category) differ significantly from other types of hospitals. However, based on the HLM results, large urban teaching hospitals are not significantly different from hospitals in four of the five other categories: small urban teaching, medium urban teaching, medium urban non-teaching, and large urban non-teaching. Even when the effects are significant under both models, the sign of the coefficient is sometimes different. For example, the OLS coefficient has a different sign than the HLM coefficient for patients in ZIP Codes with a median income of \$36,000 - \$45,999 (versus \$60,000 and higher).

Parameter	Model	Fixed Effects	Standard Error	t Value	Pr >  t
Intercept	OLS	5.6042	0.016019	349.85	<.0001
	HLM	5.4654	0.065191	83.84	<.0001
Log(Resource Demand Scale)	OLS	0.9577	0.002992	320.14	<.0001
	HLM	0.9595	0.002555	375.51	<.0001
Age – 70 (decades)	OLS	-0.0255	0.001368	-18.66	<.0001
	HLM	-0.0271	0.001159	-23.42	<.0001
Female	OLS	0.0101	0.003211	3.16	0.0016
	HLM	0.0122	0.002656	4.58	<.0001
Median ZIP Income < \$ 36,000 (vs. \$60,000+)	OLS	-0.0054	0.005009	-1.07	0.2831
	HLM	0.0007	0.005229	0.13	0.8981
Median ZIP Income \$36,000 - 44,999 (vs. \$60,000+)	OLS	-0.0775	0.004991	-15.54	<.0001
	HLM	0.0121	0.004999	2.42	0.0155

Table 1: OLS vs. HLM Estimates for Log(Total Charges)

Parameter	Model	Fixed Effects	Standard Error	t Value	Pr >  t
Median ZIP Income \$45,000 - 59,999 (vs. \$60,000+)	OLS	-0.0544	0.005160	-10.54	<.0001
	HLM	0.0009	0.004758	0.19	0.8531
Medicaid (vs. Medicare)	OLS	0.0321	0.006897	4.65	<.0001
	HLM	-0.0054	0.005794	-0.94	0.3471
Private insurance (vs. Medicare)	OLS	-0.0200	0.005266	-3.79	0.0001
	HLM	-0.0292	0.004450	-6.56	<.0001
Self-pay (vs. Medicare)	OLS	0.0480	0.010014	4.79	<.0001
	HLM	-0.0070	0.008618	-0.81	0.4171
No charge (vs. Medicare)	OLS	0.0720	0.027656	2.60	0.0092
	HLM	0.0082	0.024085	0.34	0.7321
Other payer (vs. Medicare)	OLS	0.0330	0.013183	2.50	0.0123
	HLM	-0.0922	0.011693	-7.88	<.0001
Northeast (vs. West)	OLS	-0.0630	0.005787	-10.89	<.0001
	HLM	-0.1120	0.051248	-2.19	0.0288
Midwest (vs. West)	OLS	-0.3691	0.005634	-65.51	<.0001
	HLM	-0.3346	0.043531	-7.69	<.0001
South (vs. West)	OLS	-0.3148	0.005173	-60.86	<.0001
	HLM	-0.1751	0.040590	-4.31	<.0001
Small rural hospital (vs. large urban teaching)	OLS	-0.7138	0.009667	-73.84	<.0001
	HLM	-0.7747	0.065131	-11.89	<.0001
Small urban non-teaching hospital (vs. large urban teaching)	OLS	-0.2256	0.007093	-31.81	<.0001
	HLM	-0.2829	0.066161	-4.28	<.0001
Small urban teaching hospital (vs. large urban teaching)	OLS	-0.1655	0.007795	-21.23	<.0001
	HLM	-0.1068	0.085301	-1.25	0.2106
Medium rural hospital (vs. large urban teaching)	OLS	-0.5349	0.008926	-59.92	<.0001
	HLM	-0.6149	0.071147	-8.64	<.0001
Medium urban non-teaching hospital (vs. large urban teaching)	OLS	-0.1231	0.005641	-21.83	<.0001
	HLM	-0.0737	0.067978	-1.08	0.2783
Medium urban teaching hospital (vs. large urban teaching)	OLS	-0.1839	0.006350	-28.97	<.0001
	HLM	-0.0707	0.082670	-0.86	0.3921
Large rural hospital (vs. large urban teaching)	OLS	-0.3691	0.006233	-59.21	<.0001
	HLM	-0.4217	0.067686	-6.23	<.0001

Parameter	Model	Fixed Effects	Standard Error	t Value	Pr >  t
Large urban non-teaching hospital (vs. large urban teaching)	OLS	-0.0406	0.004649	-8.74	<.0001
	HLM	0.0024	0.067378	0.04	0.9710

### Logistic Regressions for In-hospital Mortality

The overall mortality rate for CHF patients was 4.0 percent. Both a conventional logistic regression, labeled “CONV,” and a hierarchical logistic regression, labeled “HLM,” were fit to the CHF sample. For the HLM, discharges are at level 1, hospitals are at level 2, and a random-intercept model was fit using SAS PROC GLIMMIX to produce restricted maximum likelihood estimates (the default in SAS).

For the conventional logistic model, the c-statistic has a value of 0.78 and the Hosmer-Lemeshow goodness of fit statistic has a value of 14.73 with a *p*-value of .065, indicating no significant lack of fit at the 5 percent significance level.

By using GLIMMIX to fit an empty model (not shown), the random intercept variance was estimated as 0.1083 with a standard error of 0.0116, indicating significant hospital-to-hospital variation for the intercept. The intercept variance for the model with predictors was estimated as 0.1311 with a standard error of 0.01394. Consequently, the predictors did not reduce the hospital-to-hospital intercept variance for in-hospital mortality.

The results for the “fixed” effects are shown in Table 2. The shaded rows correspond to effects that are statistically significant at the 5 percent level using conventional logistic regression, but not significant at the 5 percent level using HLM. As was the case for log(charges), the estimated coefficients and their standard errors are in close agreement between the two methods for severity, patient age, and female gender. Also, the estimated standard errors are similar for ZIP income and primary payer. However, the estimated standard errors for the conventional logistic model are quite different from the hierarchical model for the hospital regions and the hospital bed size / location / teaching categories. For example, based on the conventional logistic results, one would infer that the average in-hospital mortality differs significantly between large urban teaching hospitals (the omitted category) and both large urban non-teaching hospitals and small urban teaching hospitals. However, based on the HLM results, the average mortality is not significantly different between those hospital types. Conversely, the estimate for medium rural hospitals is significant at the 5 percent level under HLM, but not under conventional logistic regression.

Table 2: Conventional vs. HLM Logistic Regression Estimates for In-Hospital Mortality

Parameter	Model	Fixed Effects	Standard Error	$\chi^2$ Value	Pr > $\chi^2$
Intercept	CONV	-5.4805	0.198899	759.2462	<.0001
	HLM	-5.5483	0.208146	710.5394	<.0001
Log(Mortality Scale)	CONV	-0.1285	0.067740	3.5986	0.0578
	HLM	-0.1153	0.068168	2.8615	0.0907
Log(Mortality Scale) Squared	CONV	0.0889	0.005816	233.5033	<.0001
	HLM	0.0893	0.005859	232.3854	<.0001
Age – 70 (decades)	CONV	0.0918	0.011177	67.4949	<.0001
	HLM	0.0855	0.011316	57.1161	<.0001
Female	CONV	0.0442	0.023232	3.6232	0.0570
	HLM	0.0382	0.023405	2.6648	0.1026
Median ZIP Income < \$ 36,000 (vs. \$60,000+)	CONV	0.0222	0.035548	0.3912	0.5316
	HLM	-0.0445	0.041263	1.1639	0.2807
Median ZIP Income \$36,000 - 44,999 (vs. \$60,000+)	CONV	0.0542	0.034749	2.4291	0.1191
	HLM	-0.0018	0.039512	0.0021	0.9635
Median ZIP Income \$45,000 - 59,999 (vs. \$60,000+)	CONV	-0.0279	0.035802	0.6070	0.4359
	HLM	-0.0623	0.038790	2.5801	0.1082
Medicaid (vs. Medicare)	CONV	0.2137	0.060835	12.3410	0.0004
	HLM	0.2042	0.061833	10.9044	0.0010
Private insurance (vs. Medicare)	CONV	0.2495	0.039742	39.3984	<.0001
	HLM	0.2296	0.040925	31.4751	<.0001
Self-pay (vs. Medicare)	CONV	0.4206	0.084777	24.6100	<.0001
	HLM	0.3888	0.087378	19.7961	<.0001
No charge (vs. Medicare)	CONV	0.3598	0.242096	2.2094	0.1372
	HLM	0.4241	0.250776	2.8602	0.0908
Other payer (vs. Medicare)	CONV	0.7136	0.084127	71.9454	<.0001
	HLM	0.7503	0.088699	71.5631	<.0001
Northeast (vs. West)	CONV	0.1752	0.039752	19.4188	<.0001
	HLM	0.1742	0.064550	7.2810	0.0070
Midwest (vs. West)	CONV	-0.0613	0.040323	2.3144	0.1282
	HLM	0.0010	0.061912	0.0002	0.9874
South (vs. West)	CONV	0.1188	0.036350	10.6765	0.0011
	HLM	0.1183	0.056302	4.4165	0.0356

Parameter	Model	Fixed Effects	Standard Error	$\chi^2$ Value	Pr > $\chi^2$
Small rural hospital (vs. large urban teaching)	CONV	0.4283	0.066855	41.0386	<.0001
	HLM	0.4616	0.088850	26.9867	<.0001
Small urban non-teaching hospital (vs. large urban teaching)	CONV	-0.0318	0.051938	0.3750	0.5403
	HLM	0.0476	0.079297	0.3607	0.5481
Small urban teaching hospital (vs. large urban teaching)	CONV	-0.2116	0.059595	12.6077	0.0004
	HLM	-0.1855	0.097393	3.6260	0.0569
Medium rural hospital (vs. large urban teaching)	CONV	0.1026	0.065105	2.4854	0.1149
	HLM	0.2020	0.091674	4.8569	0.0275
Medium urban non-teaching hospital (vs. large urban teaching)	CONV	-0.0091	0.039717	0.0523	0.8191
	HLM	0.0024	0.071497	0.0011	0.9732
Medium urban teaching hospital (vs. large urban teaching)	CONV	-0.2059	0.046124	19.9265	<.0001
	HLM	-0.1771	0.085671	4.2735	0.0387
Large rural hospital (vs. large urban teaching)	CONV	0.1322	0.044575	8.7975	0.0030
	HLM	0.1649	0.074881	4.8475	0.0277
Large urban non-teaching hospital (vs. large urban teaching)	CONV	-0.1479	0.033348	19.6630	<.0001
	HLM	-0.1308	0.066816	3.8340	0.0502

### Summary

The analyses in this section demonstrate that different inferences can result when conventional methods are applied to hierarchical data in which large group effects (large ICC values) are apparent. As it turns out, the differences between the conventional model and the HLM model were greater for the regression on log(charges) than they were for the regression on mortality. The important message is that substantial differences are possible, and that the analyst cannot foresee what those differences will be. Austin et al. (2003) similarly demonstrate these points in an analysis of patients hospitalized with acute myocardial infarction.

These analyses bypassed the usual model-building steps that include checking for influential observations, checking modeling assumptions, and so forth, which are beyond the scope of this report. Analysts are encouraged to consult the references for further guidance.



## STATISTICAL SOFTWARE

The most common hierarchical models can be fit using either of two general statistical software packages, SAS or Stata, and the program code for several models is provided in this chapter. There are also a number of specialized HLM software packages available. Here is a brief description of three special programs and some references for their use.

HLM was written by Bryk, Raudenbush, and Congdon (1996). The book by Raudenbush and Bryk (2002) provides the theory for the models supported by the software, and it is also a good general reference on hierarchical models. HLM is an interactive program in which the user specifies the variables for the model at each level. It requires separate files for each level in the hierarchy. For example, if the hierarchy consists of patients nested in hospitals, then it expects one file with patient-level data and a separate file with hospital-level data. Once the models are specified for each level, the software will optionally display the combined model. Details can be found at the Website: <http://www.ssicentral.com/hlm/>.

MLwiN was written by researchers at the London Institute of Education. The User Manual includes examples of several HLM models, including cross-classifications and multiple memberships (Rasbash et al., 2005). MLwiN is an interactive program that operates on a single, integrated database in which users specify random effects at each level. Information can be obtained from the Website: <http://www.cmm.bristol.ac.uk/>.

WinBUGS is an interactive program for Bayesian analysis of complex statistical models using Markov Chain Monte Carlo (MCMC) techniques. Researchers who want to fit Bayesian hierarchical models should acquire this program, which is constantly evolving and improving. Information is available at the Website: <http://www.mrc-bsu.cam.ac.uk/bugs>. Also of potential interest to SAS users are some SAS macros to convert SAS data for input to WinBUGS (and vice-versa) available at the following Website: <http://www.mcw.edu/pcor/bugs/>.

Finally, OpenBUGS (the open source version of WinBUGS) is directly accessible through the statistical software R (downloadable at <http://cran.r-project.org/>). This software provides two potential advantages over the standalone WinBUGS package. First, no key code is necessary to unlock R/OpenBUGS. Second, it has all of the power of R directly interfacing with BUGS. Also, for the Linux user, JAGS (Just Another Gibbs Sampler: downloadable at <http://www-fis.iarc.fr/~martyn/software/jags/>) can be directly interfaced with R.

The remainder of this chapter offers program code for SAS and Stata that can be used to fit hierarchical models with those software packages.

In SAS, continuous (normal) outcomes are modeled using PROC MIXED, which is a SAS STAT procedure. Count and binary outcomes are modeled using PROC NL MIXED or PROC GLIMMIX. PROC GLIMMIX is not incorporated as a standard SAS procedure. It must be downloaded from the SAS Institute Website:

<http://www.sas.com/apps/demosdownloads/setupcat.jsp?jsessionid=4540306B9B52B3ADF5E462E35138F29C.tomcat1?cat=SAS%2FSTAT+Software>.

The GLIMMIX User's Manual is also available for download. The book by Littell, et al. (2006) is an excellent reference for fitting hierarchical models using SAS.

In Stata, continuous outcomes are modeled using the XTMIXED command. Count and binary outcomes are modeled using the GLLAMM (Generalized Linear and Latent Mixed Models) command, which was written by Sophia Rabe-Hesketh, and is available for download from the

Website: [www.gllamm.org](http://www.gllamm.org). In addition Rabe-Hesketh and Skrondal have written a useful book that covers hierarchical modeling using XTMIXED and GLLAMM (Rabe-Hesketh and Skrondal, 2005a). Their text contains many examples and details concerning the use of XTMIXED and GLLAMM in Stata.

In addition to the Websites mentioned that are specific to these packages, the UCLA Academic Technology Service Website has program code for textbook examples using several software packages: <http://www.ats.ucla.edu/stat/>.

The discussion now turns to the SAS and Stata program code that can be used to fit the models presented earlier in this report. These programs suppose that an analyst is interested in analyzing a subset of discharges with a diagnosis of diabetes. To create the analysis file, all discharges were selected from the 2003 NIS 10 percent "A" sample (there is also a different 10 percent "B" sample available) with a principal diagnosis of diabetes as classified by the Clinical Classification System (NIS data element CCSDX1 = 50). The analysis file is named "nis\_diabetes" in the program code that follows. The reader should be able to reproduce the output in this section of the report using the programs given below.

Note that the results on the logarithm of charges for diabetes patients in this section, based on a 10 percent NIS sample, differ from the results presented in the main body of the report, based on the full NIS sample. The 10 percent sample is used here to make it easier to replicate the results, for those wishing to do so.

### **Programming Code for Continuous Outcomes**

The continuous outcome for these examples is "logtotchg" which is the logarithm of total charges for each hospital stay. The patient-level predictor variable is "CAge," which is age centered at 40 and measured in decades = (age-40)/10. The hospital-level predictor is "Teach," which is equal to 1 for teaching hospitals and 0 for non-teaching hospitals.

The SAS and the Stata programs are given for each of the following models: 1) the empty model, 2) the random intercepts, fixed slopes model, 3) the random intercepts and random slopes model, and 4) the intercepts-as-outcomes model. These correspond to the theoretical models discussed earlier in this report. The program code and output described in this section of the report are solely for the purpose of establishing some basic sample programs. Consequently, these models are very simple. SAS and Stata offer several estimation options. The default estimation option for SAS is Restricted Maximum Likelihood (REML), which is used in the examples that follow. The estimation method used for Stata is Maximum Likelihood Estimation (MLE), which tends to produce virtually the same estimates as REML does. The reader should consult the documentation for each procedure for alternative specifications and for other estimation options.

## SAS Programs

### *Random Intercepts, Empty Model*

The SAS code for the empty model is as follows:

```
proc mixed data=in.nis_diabetes noitprint noclprint covtest scoring=8 ;
  class hospid ;
  model logtotchg = / solution ;
  random intercept / subject=hospid ;
  title 'Empty model for Log of total charges';
run;
```

The CLASS statement designates the hospital identifier, hospid, as a categorical variable. The NOCLPRINT option prevents the printing of the CLASS level information while the NOITPRINT option suppresses the display of the “Iteration History” table. The COVTEST option invokes hypothesis tests for the variance and covariance components. The option SCORING=8 requests that Fisher scoring be used for the first eight iterations.

The MODEL statement gives the outcome variable on the left side of the equal sign and the predictor variable(s) on the right side. In this case, no predictor variables are specified because the intercept term is included by default. The SOLUTION option in the MODEL statement requests estimates for the fixed effect(s).

The RANDOM statement specifies random effects, which in this model is just the intercept term (intercept is a reserved keyword in the RANDOM statement). The SUBJECT option on the RANDOM statement indicates the hierarchical structure. In this case, discharges are grouped into hospitals. Consequently, the “subject” is the hospital identifier, which is designated as a CLASS variable to indicate that it should be treated as a classification variable with discrete values.

### Results

The key output is as follows:

Covariance Parameter Estimates					
Cov Parm	Subject	Estimate	Standard Error	Z Value	Pr Z
UN(1,1)	HOSPID	0.3079	0.02061	14.94	<.0001
Residual		0.6882	0.01027	67.03	<.0001

Solution for Fixed Effects					
Effect	Estimate	Standard Error	DF	t Value	Pr >  t
Intercept	9.1814	0.02288	833	401.32	<.0001

### Interpretation

By default, SAS PROC MIXED produces restricted maximum likelihood (REML) parameter estimates. The first row of the “Covariance Parameter Estimates,” labeled “UN(1,1),” gives the hospital-to-hospital intercept variance as 0.3079 with an estimated standard error of 0.02061, producing a Z-value of 14.94, corresponding to a *p*-value of less than .0001. This suggests that hospitals differ significantly in their average log(charges) for diabetes patients. The estimate of the level-1 random error, the patient residual variance, is 0.6882, which is also statistically significant. The estimated intraclass correlation coefficient is

$$ICC = 0.31 = 0.3079 / (0.3079 + 0.6882),$$

indicating that approximately one-third of the total variance in log(charges) is between hospitals. This value is higher than that found in many applications, and it indicates a substantial level of between-hospital variation.

The “Solution for Fixed Effects” table displays parameter estimates for the fixed effects. In the empty model there is only one fixed effect, the average intercept, estimated as 9.1814.

### *Random Intercepts, Fixed Slopes*

The SAS code for the random intercepts, fixed slopes model is as follows:

```
proc mixed data=in.nis_diabetes noitprint noclprint covtest scoring=8 ;
  class hospid ;
  model logtotchg = CAge / solution ;
  random intercept / subject=hospid;
  title 'Random intercepts, fixed slope for Log of total charges';
run;
```

The only change compared with the empty model occurs in the Model statement, which now includes the additional fixed effect for CAge = (Age – 40) / 10, which is age, centered at 40 years, and measured in decades.

### Results

The key output is as follows:

Covariance Parameter Estimates					
Cov Parm	Subject	Estimate	Standard Error	Z Value	Pr Z
UN(1,1)	HOSPID	0.3036	0.02025	14.99	<.0001
Residual		0.6661	0.009943	66.99	<.0001

Solution for Fixed Effects					
Effect	Estimate	Standard Error	DF	t Value	Pr >  t
Intercept	9.0676	0.02356	833	384.86	<.0001
CAge	0.07717	0.004366	8913	17.68	<.0001

### Interpretation

The “Covariance Parameter Estimates” table shows that the hospital-to-hospital intercept variance is 0.3036, which is only slightly lower than the value of .03079 that was estimated from the empty model presented in the last section. The estimate of the level-1 residual error is 0.6661, compared with 0.6882 in the empty model. Consequently, the addition of CAge to the model slightly reduced the variance at both the hospital level and the patient level. The estimated intraclass correlation coefficient can be estimated as  $0.31 = 0.3036 / (0.3036 + 0.6661)$ .

The “Solution for Fixed Effects” table contains REML estimates for the fixed effects. The average intercept is estimated as 9.0676, controlling for CAge. The estimated coefficient of CAge is 0.07717, indicating that log(charges) increase by an average of about 0.077 for each decade increase in patient age. An alternative interpretation is that total charges are multiplied by the anti-log of the CAge coefficient,  $1.08 = \exp(0.07717)$ , so that average charges increase by approximately 8 percent for every decade increase in age. The standard errors for these fixed effects are relatively small, resulting in large *t*-statistics and low *p*-values.

### *Random Intercepts, Random Slopes*

The SAS code for the random intercepts, random slopes model is as follows:

```
proc mixed data=in.nis_diabetes noitprint noclprint covtest scoring=8 ;
  class hospid ;
  model logtotchg = CAge / solution ;
  random intercept CAge / type=un subject=hospid G Gcorr ;
  title 'Random intercepts, random slopes model for log of total charges';
run;
```

As indicated in the random statement, this model has two random effects: one for the intercept and one for the CAge slope. The G option in the Random statement requests that the G matrix be printed. In SAS, the G matrix is the variance-covariance matrix for the random effects. The Gcorr option asks for the correlation between the random intercepts and random slopes. The covariance structure option, UN, requests an unstructured covariance matrix, which implies three components for the G matrix: 1) the random intercept variance, 2) the random slope variance, and 3) the covariance between the random intercepts and random slopes.

### Results

The key output is as follows:

Estimated G Matrix				
Row	Effect	HOSPID	Col1	Col2
1	Intercept	04001	0.2835	0.006002
2	CAge	04001	0.006002	0.001011

Estimated G Correlation Matrix				
Row	Effect	HOSPID	Col1	Col2
1	Intercept	04001	1.0000	0.3546
2	CAge	04001	0.3546	1.0000

Covariance Parameter Estimates					
Cov Parm	Subject	Estimate	Standard Error	Z Value	Pr Z
UN(1,1)	HOSPID	0.2835	0.02043	13.87	<.0001
UN(2,1)	HOSPID	0.006002	0.002672	2.25	0.0247
UN(2,2)	HOSPID	0.001011	0.000598	1.69	0.0454
Residual		0.6624	0.01007	65.75	<.0001

Fit Statistics	
-2 Res Log Likelihood	24965.5
AIC (smaller is better)	24973.5
AICC (smaller is better)	24973.5
BIC (smaller is better)	24992.4

Solution for Fixed Effects					
Effect	Estimate	Standard Error	DF	t Value	Pr >  t
Intercept	9.0753	0.02296	833	395.18	<.0001
CAge	0.07183	0.004651	732	15.45	<.0001

### Interpretation

The REML estimates of the variances and covariances are contained in the “Estimated G Matrix” and in the “Covariance Parameter Estimates” tables. The value of UN(1,1), 0.2835, is the estimated intercept variance; the value of UN(2,2), 0.001011, is the estimated slope (of CAge) variance. The value of UN(2,1), 0.006002, is the estimated covariance between hospital intercepts and hospital slopes. The corresponding *t*-values and *p*-values are small, suggesting that hospital intercepts and slopes vary significantly and that the correlation is significant between intercepts and slopes. The G-correlation matrix contains the estimated correlation between intercepts and slopes, 0.35. This implies that hospitals with higher intercepts also tend to have higher slopes on CAge—older patients tend to cost more than younger patients by a larger factor at hospitals with higher intercepts compared to hospitals with lower intercepts.

The “Fit Statistics” table contains a series of statistics that can be used to compare alternative covariance structures. Models that fit better have smaller values on the Akaike’s Information Criteria (AIC), the AIC Corrected (AICC), or the Bayesian Information Criterion (BIC). The reader is referred to Little et al. (2006) for a discussion of these statistics and their use in selecting covariance structures.

The “Solution for Fixed Effects” table displays parameter estimates for the fixed effects. The average intercept is estimated as 9.0753, controlling for CAge. The estimated average slope of CAge is 0.07183, indicating an increase in total charges of about 7.4 percent for every decade increase in patient age. The standard errors for these fixed effects are very small, resulting in large *t*-statistics and low *p*-values.

*Intercept- and Slopes-As-Outcomes*

The SAS code for the intercept- and slopes-as-outcomes model is as follows:

```
proc mixed data=in.nis_diabetes noitprint noclprint covtest scoring=8 ;
  class hospid ;
  model logtotchg = CAge Teach CAge*Teach / solution ;
  random intercept CAge / type=un subject=hospid G Gcorr ;
  Title 'Intercepts and slopes as outcomes model for Total Charges';
run;
```

This model adds a hospital-level predictor, Teach, equal to 1 for teaching hospitals and equal to 0 for non-teaching hospitals. The term, CAge\*Teach, adds the interaction between CAge and Teach, implying that the random slopes of CAge could differ between teaching hospitals and non-teaching hospitals.

Results

The key output is as follows:

Estimated G Matrix				
Row	Effect	HOSPID	Col1	Col2
1	Intercept	04001	0.2648	0.004886
2	CAge	04001	0.004886	0.000958

Estimated G Correlation Matrix				
Row	Effect	HOSPID	Col1	Col2
1	Intercept	04001	1.0000	0.3068
2	CAge	04001	0.3068	1.0000

Covariance Parameter Estimates					
Cov Parm	Subject	Estimate	Standard Error	Z Value	Pr Z
UN(1,1)	HOSPID	0.2648	0.01944	13.62	<.0001
UN(2,1)	HOSPID	0.004886	0.002599	1.88	0.0602
UN(2,2)	HOSPID	0.000958	0.000595	1.61	0.0536
Residual		0.6617	0.01006	65.76	<.0001

Fit Statistics	
-2 Res Log Likelihood	24921.1
AIC (smaller is better)	24929.1
AICC (smaller is better)	24929.1
BIC (smaller is better)	24948.0

Solution for Fixed Effects					
Effect	Estimate	Standard Error	DF	t Value	Pr >  t
Intercept	8.9969	0.02609	832	344.81	<.0001
CAge	0.06708	0.006049	731	11.09	<.0001
Teach	0.3659	0.05247	8181	6.97	<.0001
CAge*Teach	0.01890	0.009451	8181	2.00	0.0456

### Interpretation

The random intercept variance is estimated to be 0.2648 and it is significant. However, the estimates of the other random effects—the random effect of CAge and the covariance between intercepts and slopes—are not statistically significant at the 5 percent level. Notice also that the random intercept variance, 0.2648, is smaller in this model than it was in the previous model (0.3036), as a result of adding the hospital-level predictor, Teach, and its interaction with CAge.

The fixed effects are all statistically significant. The estimated effect of Teach implies that log(charges) average 0.3659 higher for teaching hospitals than for non-teaching hospitals, for 40-year-old patients (when CAge = 0). The interaction between CAge and TEACH is significant at the 5 percent level. It implies that the average slope for CAge is 0.0189 higher for teaching hospitals compared with non-teaching hospitals, for 40-year-old patients.

As an aside, the estimates in the main body of the text for this model, based on the full 2003 NIS, are close to those shown here, except that the full sample standard errors are smaller. The full sample estimates of the random effects are all statistically significant.



## Stata Programs

### *Random Intercepts, Empty Model*

The Stata command for the empty model is:

```
xtmixed logtotchg || hospid: , variance mle
```

The Stata command is XT MIXED and logtotchg is the dependent variable. Random effects are specified after the double-bar ||. In this case, the term “hospid: “ after the || indicates random intercepts (included by default), which vary by hospid. The variance option calculates variances and the mle option requests maximum likelihood estimation.

### Results

The key output is as follows:

logtotchg	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
+						
_cons	9.18145	.0228634	401.58	0.000	9.136639	9.226261
-----						
Random-effects Parameters			Estimate	Std. Err.	[95% Conf. Interval]	
-----						
hospid: Identity						
	var(_cons)		.3073635	.0208819	.2690436	.3511413
-----						
	var(Residual)		.6882129	.01028	.6683566	.7086591
-----						

### Interpretation

The \_cons coefficient is the estimate of the fixed effect, the average intercept. The “Random-effects Parameters” table contains the estimated variances of the random effects and their standard errors. The hospital-level random intercept variance is 0.3074 and the patient-level residual variance is 0.6882. With this information, the ICC =  $0.3074 / (0.3074 + 0.6882) = 0.31$ .

### *Random Intercepts, Fixed Slopes*

The Stata command for the random intercept, fixed slope model is:

```
xtmixed logtotchg cage || hospid: , variance mle
```

As with the empty model, the Stata command is XT MIXED and logtotchg is the dependent variable, while cage is added as a discharge-level predictor. The specification “hospid: “ after the || indicates random intercepts, which vary by hospid.

## Results

The key output is as follows:

logtotchg	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
cage	.0771748	.0043658	17.68	0.000	.0686179	.0857316
_cons	9.06769	.023547	385.09	0.000	9.021538	9.113841

Random-effects Parameters	Estimate	Std. Err.	[95% Conf. Interval]	
hospid: Identity				
var(_cons)	.3031478	.0205646	.2654066	.3462559
var(Residual)				
	.6660383	.0099562	.6468076	.6858408

## Interpretation

The estimated coefficient of cage is 0.0772 and the average hospital intercept is 9.0677. The variance of the random intercepts is estimated as 0.3031, and the patient-level residual variance is estimated to be 0.6660. All estimates are statistically significant.

### *Random Intercepts, Random Slopes*

The Stata commands for the random intercepts, random slope model are:

```
xtmixed logtotchg cage || hospid: cage , cov(unstr) mle
```

For this model, cage is added as a random effect, listed after the double-bar ||. The random intercept is included by default. An unstructured covariance matrix is specified by the option cov(unstr) and maximum likelihood estimates are requested.

## Results

The key output is as follows:

logtotchg	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
cage	.0718641	.0046444	15.47	0.000	.0627613	.0809669
_cons	9.075364	.0229475	395.48	0.000	9.030388	9.12034

Random-effects Parameters	Estimate	Std. Err.	[95% Conf. Interval]	
hospid: Unstructured				

sd(cage)	.031461	.0091529	.0177883	.055643
sd(_cons)	.5319018	.0197335	.4945976	.5720196
corr(cage,_cons)	.3594371	.1973968	-.0679466	.6753583
-----+				
sd(Residual)	.8138795	.0061971	.8018236	.8261167
-----				
LR test vs. linear regression:	chi2(3) =	2187.80	Prob > chi2 =	0.0000

Note: LR test is conservative and provided only for reference.

### Interpretation

The average cage coefficient is estimated to be 0.0719 and the average intercept is estimated to be 0.0754. The standard deviation of the random slopes, sd(cage), is estimated as 0.0315. The standard deviation of the random intercepts, sd(\_cons), is estimated as 0.5319. The correlation between the random intercepts and random slopes, corr(cage, \_cons), is estimated as 0.3594. The 95 percent confidence interval for the correlation is (-0.0679, 0.6754), which includes zero. Consequently, the correlation between intercepts and slopes is not significantly different from zero at the 5 percent significance level.

### Intercept- and Slopes-As-Outcomes

The Stata commands are:

```
xtmixed logtotchg cage teach teachXcage || hospid: cage, variance cov(un) mle
```

The Stata command is XT MIXED, logtotchg is the dependent variable, cage is a discharge-level predictor, and teach is a hospital-level predictor. The newly added term, teachXcage, is the interaction between cage and teach, which must be constructed as a new variable in Stata. Random effects are specified after the double-bar ||. The intercept and CAge vary with the value of hospid. Maximum likelihood estimates are requested along with variance estimates corresponding to an unstructured covariance matrix.

### Results

The key output is as follows:

logtotchg	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
cage	.0671278	.0060379	11.12	0.000	.0552937	.078962
teach	.3656849	.0523762	6.98	0.000	.2630293	.4683404
teachXcage	.0189177	.0094195	2.01	0.045	.0004558	.0373796
_cons	8.996938	.0260543	345.32	0.000	8.945873	9.048003

Random-effects Parameters	Estimate	Std. Err.	[95% Conf. Interval]	
-----+				
hospid: Unstructured				
var(cage)	.000914	.0005684	.0002701	.0030927

var(_cons)	.2636992	.0197212	.2277457	.3053285
cov(cage,_cons)	.0049209	.0026008	-.0001766	.0100184
-----+				
var(Residual)	.6617481	.010063	.6423159	.6817681
-----				
LR test vs. linear regression:	chi2(3) =	2093.34	Prob > chi2 =	0.0000

Note: LR test is conservative and provided only for reference.

### Interpretation

The average hospital cage slope indicates an increase in log(charges) equal to 0.0671 for every decade increase in age. Teaching hospitals average 0.3657 higher than non-teaching hospitals, when age = 40. The age slope is 0.0189 higher for teaching hospitals compared with non-teaching hospitals, when age = 40. The average hospital intercept is estimated to be 8.997, when age = 40.

### **Programming Code for Binary Outcomes**

SAS can fit binary or binomial outcomes using either PROC NL MIXED or PROC GLIMMIX. Although NL MIXED has a lot of flexibility, it requires special programming to fit hierarchical models. On the other hand, GLIMMIX was especially developed as a SAS procedure with syntax similar to that for PROC MIXED, described in the previous section of this report. Stata can fit binary or binomial outcomes using the GLLAMM command. Both PROC GLIMMIX and GLLAMM must be downloaded as described earlier.

The probability of in-hospital mortality for patients with diabetes is modeled using the same subset of data that was used to model the logarithm of charges in the previous section. The patient-level predictor variable is "CAge," which is age, centered at 40, and measured in decades = (age-40)/10. The hospital-level predictor is "Teach," which is equal to 1 for teaching hospitals and 0 for non-teaching hospitals.

Both the SAS and the Stata program code are given for each of the following models: 1) the empty model, 2) the random intercepts, fixed slopes model, 3) the random intercepts and random slopes model, and 4) the intercepts-as-outcomes model. These correspond to the theoretical models discussed earlier in this report. The program code and output described in this section of the report are solely for the purpose of establishing some basic sample programs. SAS and Stata offer several estimation options. The default estimation option for SAS is Restricted Maximum Likelihood (REML), which is used in the examples that follow. The estimation method used for Stata is Maximum Likelihood Estimation (MLE), which usually produces estimates similar to REML. The reader should consult the documentation for each procedure for alternative specifications and for other estimation options.

## SAS Programs

### *Random Intercepts, Empty Model*

The SAS code for the empty model is as follows:

```
proc glimmix data=in.nis_diabetes noitprint noclprint ;
  class hospid ;
  model died (descending) = / dist=binomial solution ;
  random intercept / sub=hospid ;
  title 'One-way ANOVA with random effects for Mortality';
run;
```

The PROC GLIMMIX code for modeling binary outcomes is similar to that used by the PROC MIXED procedure for modeling continuous outcomes. The GLIMMIX procedure can fit continuous, binary, or count outcomes, but it is primarily for non-continuous outcomes. The objective here is to model the probability of in-hospital mortality. Therefore, a logistic model is fit.

The NOCLPRINT option prevents the printing of the CLASS level information, while the NOITPRINT option suppresses the display of the “Iteration History” table. The CLASS statement designates the hospital identifier, hospid, as a categorical variable.

The MODEL statement gives the outcome variable, died, on the left side of the equal sign and the predictor variable(s) on the right side. The “descending” option is specified to model the probability that died = 1, instead of the probability that died = 0. No predictor variables are specified because the intercept term is included by default. The SOLUTION option specifies that the estimates for fixed effects should be printed. DIST=BINOMIAL specifies the binomial for the patient-level mortality distribution. The default “link” function, linking the dependent variable to the linear predictor, is the logit or log-odds.

The RANDOM statement identifies random effects, which in this model is one term, the intercept. The SUB option on the random statement indicates the hierarchical structure. In this case, discharges are grouped into hospitals (e.g., subject = hospid).

### Results

The key output is as follows:

<b>Model Information</b>	
<b>Data Set</b>	IN.NIS_DIABETES
<b>Response Variable</b>	DIED
<b>Response Distribution</b>	Binomial
<b>Link Function</b>	Logit
<b>Variance Function</b>	Default
<b>Variance Matrix Blocked By</b>	HOSPID

Model Information	
Estimation Technique	Residual PL
Degrees of Freedom Method	Containment

Dimensions	
G-side Cov. Parameters	1
Columns in X	1
Columns in Z per Subject	1
Subjects (Blocks in V)	842
Max Obs per Subject	141

Fit Statistics	
-2 Res Log Pseudo-Likelihood	69661.38
Generalized Chi-Square	7208.58
Gener. Chi-Square / DF	0.73

Covariance Parameter Estimates			
Cov Parm	Subject	Estimate	Standard Error
Intercept	HOSPID	0.3582	0.2045

Solutions for Fixed Effects					
Effect	Estimate	Standard Error	DF	t Value	Pr >  t
Intercept	-4.4720	0.09943	840	-44.98	<.0001

### Interpretation

The "Model Information" table provides basic model information, including the method of estimation. The "Dimensions" table shows how many columns exist in the fixed effects matrix (X) and in the random effects matrix (Z) parts of the model. In this case, there is only one fixed effect, the average intercept, and one random effect, the intercept. In PROC GLIMMIX, the covariance matrix for the fixed effects is labeled "R" and the covariance matrix for the random effects is labeled "G." There is one G-side covariance parameter.

In the "Fit Statistics" table, PROC GLIMMIX produces log-likelihood and chi-square statistics, which can be compared to other models that include more terms in addition to the intercept. The "Covariance Parameter Estimates" table contains the variance of the random intercept, which is 0.3582, with a standard error of 0.2045, indicating that the hospital-to-hospital intercept variance is not significantly different from 0. The "Solutions for Fixed Effects" table contains the estimate of the overall intercept, -0.4720, which is the log-odds of mortality for the average hospital.

### Random Intercepts, Fixed Slopes

The SAS code for the random intercepts, fixed slopes model is as follows:

```
proc glimmix data=in.nis_diabetes noitprint noclprint ;
  class hospid ;
  model died (descending) = CAge / dist=binomial solution ;
  random intercept / sub=hospid ;
  title 'One-way ANCOVA with random effects for Mortality';
run;
```

The only syntax change from the empty model, described above, is in the model statement, which now includes CAge as a “fixed” effect.

### Results

The key output is as follows:

Covariance Parameter Estimates			
Cov Parm	Subject	Estimate	Standard Error
Intercept	HOSPID	0.2883	0.2012

Solutions for Fixed Effects					
Effect	Estimate	Standard Error	DF	t Value	Pr >  t
Intercept	-5.6528	0.2140	840	-26.42	<.0001
CAge	0.5150	0.06240	9006	8.25	<.0001

Type III Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
CAge	1	9006	68.10	<.0001

### Interpretation

The REML estimate of the hospital intercept variance, 0.2883, is contained in the “Covariance Parameter Estimates” table. The relatively large standard error for this estimate, .2012, suggests that hospitals do not differ statistically in their average diabetes mortality, after accounting for age.

The “Solution for Fixed Effects” table displays parameter estimates for the fixed effects. The estimate of -5.6528 for the intercept is the log-odds of mortality for the average hospital, when age is equal to 40 (CAge is equal to 0). The estimated coefficient of CAge is 0.5150, which

indicates that the log-odds of mortality increases by 0.515 for every decade increase in age. The standard errors for these fixed effects are very small, resulting in large *t*-statistics and low *p*-values.

*Random Intercepts, Random Slopes*

The SAS code for the random intercepts, random slopes model is as follows:

```
proc glimmix data=in.nis_diabetes noitprint noclprint ;
  class hospid ;
  model died (descending) = CAge / dist=binomial solution ;
  random intercept CAge / type=un sub=hospid ;
  parms / noiter ;
  title 'Random intercepts, random slopes model for Mortality';
```

Compared with the random intercept constant slope model, the random statement is changed to request random slopes for CAge. This model has two random effects, one for the intercept and one for the slope of CAge. The covariance structure option, type = un, requests an unstructured covariance matrix with three components: the intercept variance, the CAge slope variance, and the covariance between the random slopes and random intercepts. GLIMMIX offers a host of possible specifications for the covariance structure. The user should consult the GLIMMIX User's Guide for an explanation of other structures.

Results

The key output is as follows:

Covariance Parameter Estimates			
Cov Parm	Subject	Estimate	Standard Error
UN(1,1)	HOSPID	0.002395	0.9290
UN(2,1)	HOSPID	0.000482	0.2304
UN(2,2)	HOSPID	0.003430	0.06338

Solutions for Fixed Effects					
Effect	Estimate	Standard Error	DF	t Value	Pr >  t
Intercept	-5.6367	0.2117	840	-26.63	<.0001
CAge	0.5157	0.06224	739	8.29	<.0001



Type III Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
CAGE	1	739	68.66	<.0001

### Interpretation

The variance-covariance estimates are shown in the table “Covariance Parameter Estimates.” The row labeled “UN(1,1)” is for the random intercept variance, the row labeled “UN(2,2)” is for the random slope variance, and the row labeled “UN(2,1)” is for the covariance between the two. All of these estimates are small and statistically insignificant.

The estimates of the fixed effects are -5.6367 for the average intercept and 0.5157 for the coefficient of CAGE, which are similar to the coefficients estimated for the previous model.

### Intercept- and Slopes-As-Outcomes

The SAS code for the intercept- and slopes-as-outcomes model is as follows:

```
proc glimmix data=in.nis_diabetes noitprint noclprint ;
  class hospid ;
  model died (descending) = CAGE Teach CAGE*Teach / dist=binomial solution;
  random intercept CAGE / type=un sub=hospid ;
  parms / noiter ;
  Title 'Intercepts and slopes as outcomes model for Mortality';
```

Compared with the random intercepts and random slopes model, two terms are added to the model statement: Teach and CAGE\*Teach. Teach is a hospital-level variable equal to 1 for teaching hospitals and equal to 0 for non-teaching hospitals. CAGE\*Teach is an interaction term between CAGE and TEACH. The random slopes of CAGE are assumed to depend on hospital teaching status.

### Results

The key output is as follows:

Covariance Parameter Estimates			
Cov Parm	Subject	Estimate	Standard Error
UN(1,1)	HOSPID	0.002497	0.8779
UN(2,1)	HOSPID	0.000540	0.2238
UN(2,2)	HOSPID	0.003318	0.06359

Solutions for Fixed Effects					
Effect	Estimate	Standard Error	DF	t Value	Pr >  t
Intercept	-5.8830	0.3040	839	-19.35	<.0001
CAge	0.5988	0.08494	738	7.05	<.0001
Teach	0.5307	0.4177	8267	1.27	0.2040
CAge*Teach	-0.1981	0.1252	8267	-1.58	0.1136

### Interpretation

The variance and covariance parameters are all very small and statistically insignificant, implying that the intercepts and slopes do not vary with hospital teaching status.

For the fixed effects, the coefficient of CAge is significantly different from 0 whereas the main effect of Teach is not, indicating that the intercept for the average teaching hospital is not different from the average intercept for non-teaching hospitals. The interaction between CAge and Teach is not significant, indicating that the average age effect does not differ significantly by teaching status.

### Stata Programs

#### *Random Intercepts, Empty Model*

The Stata command is:

```
gllamm died, i(hospid) link(logit) family(binom) adapt
```

The Stata command is GLLAMM and died is the dependent variable. The binomial distribution is specified along with the logit link. The option i() lists the hierarchical nested clusters from smallest to largest. For example, if the patients were nested within physicians and physicians were nested in hospitals, then the option would be i(physician hospital). In the present application, hospid is listed as the group or cluster identifier. The adapt option requests that adaptive quadrature be used in maximizing the likelihood. Adaptive quadrature is generally considered superior to other methods, especially for large cluster sizes and large intraclass correlations. However, it is more demanding computationally (Rabe-Hesketh et al., 2002).

### Results

The key output is as follows:

```
gllamm model
```

```
log likelihood = -619.71704
```

```
-----+-----
      died |      Coef. Std. Err.      z    P>|z|     [95% Conf. Interval]
-----+-----
      _cons | -4.658136   .1591222   -29.27   0.000   -4.97001   -4.346263
-----+-----
```

Variiances and covariances of random effects

\*\*\*level 2 (hospid)

var(1): .41705228 (.25297922)

Interpretation

The estimates using GLLAMM are different from those using GLIMMIX primarily due to the adaptive quadrature estimation used by GLLAMM (as an option). For example, the GLLAMM estimate of the overall intercept is -4.658, as compared with -4.472 using GLIMMIX. The GLLAMM estimate for the hospital-to-hospital intercept variance is 0.417, with a standard deviation of 0.253.

*Random Intercepts, Fixed Slopes*

The Stata commands are:

```
gllamm died cage, i(hospid) link(logit) family(binom) adapt
```

The only change from the empty model for the GLLAMM statement, described earlier, is that CAge is added as a discharge-level predictor.

Results

The key output is as follows:

gllamm model

log likelihood = -575.94527

died	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
cage	.5178015	.0627319	8.25	0.000	.3948492	.6407537
_cons	-5.809921	.2486373	-23.37	0.000	-6.297242	-5.322601

Variiances and covariances of random effects

\*\*\*level 2 (hospid)

var(1): .33347854 (.24481664)

### Interpretation

The estimate of the random intercept using GLLAMM with adaptive quadrature differs from the GLIMMIX estimate. The GLLAMM variance estimate of the random intercepts is 0.3335 while its estimation using GLIMMIX is 0.2883, although it is not statistically significant in either case. In addition, the estimates of the fixed effects are 0.5178 for the Cage slope and -5.8099 for the average hospital intercept.

### *Random Intercepts, Random Slopes*

The Stata commands are:

```
gen cons = 1
eq cons: cons
eq cage: cage
gllamm died cage, i(hospid) link(logit) family(binom) nrf(2) eqs(cage cons)
adapt
```

The first statement generates a vector ones and places it in the variable “cons.” The second statement and third statement create “equations” for the random effects. The nrf option specifies the number of random effects, which is 2. The other options were explained earlier for models previously discussed.

### Results

The key output is as follows:

**gllamm model**

**log likelihood = -575.42525**

```
-----+-----
      died |      Coef. Std. Err.      z    P>|z|     [95% Conf. Interval]
-----+-----
      cage |   .4731404   .1135109     4.17   0.000     .2506632     .6956176
     _cons |  -5.678198   .4056365    -14.00   0.000    -6.473231    -4.883165
-----+-----
```

**Variances and covariances of random effects**

**\*\*\*level 2 (hospid)**

```
var(1): .02956028 (.04297933)
cov(2,1): -.0234992 (.15374653) cor(2,1): -.36208346
var(2): .14248858 (.73567786)
-----
```

### Interpretation

The estimates of the fixed effects are 0.47 for the average age slope and -5.68 for the average intercept. The slope and intercept variances are labeled “var(1)” and “var(2),” respectively, because that is the order in which they are listed in the eqs() option. The covariance estimate is labeled “cov(2,1).” The estimated slope variance is 0.0296 with a standard error of 0.0430. The estimated intercept variance is 0.1425 with a standard error of 0.7357. The estimated covariance is 0.0235 with a standard error of 0.1537. The estimated correlation, labeled “cor(2,1),” is -0.36 between the random slopes and random intercepts.

### Intercept- and Slopes-As-Outcomes

The Stata commands are:

```
gen cons = 1
gen teachXcage = teach * cage
eq cage = cage
eq cons = cons
gllamm died cage teach teachXcage, i(hospid) link(logit) family(binom)
nrf(2) eqs(cage cons) adapt
```

The only difference between this program and the previous program is the specification of the interaction term teachXcage as a “fixed” effect.

### Results

The key output is as follows:

**gllamm model**

**log likelihood = -574.47715**

died	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
cage	.5568866	.1256532	4.43	0.000	.3106109	.8031624
teach	.5251542	.4335477	1.21	0.226	-.3245836	1.374892
teachXcage	-.1792738	.1307934	-1.37	0.170	-.4356241	.0770766
_cons	-5.93533	.4528876	-13.11	0.000	-6.822973	-5.047686

### **Variances and covariances of random effects**

**\*\*\*level 2 (hospid)**

```
var(1): .01573126 (.04290265)
cov(2,1): .0009305 (.14929964) cor(2,1): .02014846
var(2): .13557539 (.69979177)
```

### Interpretation

The average slope on cage is .5569 and the average intercept is -5.9353, controlling for the value of age. Neither the teaching effect nor the crossed age/teaching effect is statistically significant. The hospital-to-hospital variance of the CAge slopes is given by var(1), while the variance of the intercepts (conditional on age) is given by var(2)—var(1) and var(2) correspond to variables in the same order as they are listed in the eqs() option. The covariance between the random intercepts and random slopes is given by var(2,1) and the correlation between them is given by cor(2,1). However, none of the variances are statistically different from zero, as they have relatively large estimated standard errors (in parentheses).

### **Programming Code for Count Outcomes**

SAS can fit count outcomes using either PROC NLMIXED or PROC GLIMMIX. Although NLMIXED has a lot of flexibility, it requires special programming to fit hierarchical models. On the other hand, GLIMMIX was especially developed as a SAS procedure with syntax similar to that for PROC MIXED, used for continuous outcomes. Stata can fit count outcomes using the GLLAMM command. Both PROC GLIMMIX and GLLAMM must be downloaded, as described earlier.

The patient length of stay (LOS) for patients with diabetes is modeled using the same subset of data that was used to model the logarithm of charges and in-hospital mortality in the previous sections. The dependent variable, measured in days, is redefined by subtracting 1 to create a variable with integer values starting at zero, just as the Poisson distribution does. The patient-level predictor variable is “CAge,” which is age, centered at 40, and measured in decades = (age-40)/10. The hospital-level predictor is “Teach,” which is equal to 1 for teaching hospitals and 0 for non-teaching hospitals.

Both the SAS and the Stata program code are given for each of the following models: 1) the empty model, 2) the random intercepts, fixed slopes model, 3) the random intercepts and random slopes model, and 4) the intercepts-as-outcomes model. These correspond to the theoretical models discussed earlier in this report. The program code and output described in this section of the report are solely for the purpose of establishing some basic sample programs. SAS and Stata offer several estimation options. The default estimation option for SAS is Restricted Maximum Likelihood (REML), which is used in the examples that follow. The estimation method used for Stata is Maximum Likelihood Estimation (MLE), which usually produces estimates similar to REML. The reader should consult the documentation for each procedure for alternative specifications and for other estimation options.

## SAS Programs

### *Random Intercepts, Empty Model*

The SAS code for the empty model is as follows:

```
proc glimmix data=in.nis_diabetes noitprint noclprint ;
  class hospid ;
  model los_1 = / dist=poisson solution ;
  random intercept / sub=hospid ;
  parms / noiter ; * This statement not generally necessary ;
  Title 'One-way ANOVA with random effects for LOS-1';
run;
```

The SAS code for count outcomes is similar to the PROC GLIMMIX code that was described for binary outcomes. The CLASS statement designates the hospital identifier, hospid, as a categorical variable. The NOCLPRINT option prevents the printing of the CLASS level information while the NOITPRINT option suppresses the display of the “Iteration History” table.

The dependent variable in the MODEL statement is los\_1 (LOS minus 1). The DIST = POISSON option specifies the Poisson distribution. The default link function for the Poisson distribution is the log link. No predictor variables are entered on the right side of the equal sign because the intercept term is included by default. The SOLUTION option requests the printing of estimates for fixed effects.

The RANDOM statement instructs the procedure to fit random intercepts that vary at the hospital level (sub=hospid).

For this particular data set, the procedure failed to converge without the PARMS statement. This can happen for any number of reasons. GLIMMIX provides many estimation options that can be employed to solve convergence problems. In this case, the PARMS statement, which is usually used to set initial values for the covariance or scale parameters, is used with the “noiter” option to solve for minimum variance quadratic unbiased estimates. Other options could have been employed in an attempt to solve the convergence problem. This is just one example.

### Results

The key output is as follows:

<b>Model Information</b>	
<b>Data Set</b>	IN.NIS_DIABETES
<b>Response Variable</b>	Los_1
<b>Response Distribution</b>	Poisson
<b>Link Function</b>	Log
<b>Variance Function</b>	Default
<b>Variance Matrix Blocked By</b>	HOSPID
<b>Estimation Technique</b>	Residual PL
<b>Degrees of Freedom Method</b>	Containment

Dimensions	
G-side Cov. Parameters	1
Columns in X	1
Columns in Z per Subject	1
Subjects (Blocks in V)	842
Max Obs per Subject	141

Fit Statistics	
-2 Res Log Pseudo-Likelihood	66834.19
Generalized Chi-Square	60829.02
Gener. Chi-Square / DF	6.14

Covariance Parameter Estimates			
Cov Parm	Subject	Estimate	Standard Error
UN(1,1)	HOSPID	0.2183	0.01175

Solutions for Fixed Effects					
Effect	Estimate	Standard Error	DF	t Value	Pr >  t
Intercept	1.2606	0.01831	840	68.84	<.0001

### Interpretation

The “Model Information” table provides information about how the GLIMMIX procedure proceeds with the model and the methods of estimation. The “Dimensions” table shows how many columns are in the fixed effects matrix (X) and in the random effects matrix (Z) parts of the model. In this case, there is only one fixed and one random effect, the intercept. The number of hospitals (subjects) is 842, and the maximum number of patients in a hospital (max obs per subject) is 141.

The “Fit Statistics” table presents statistics that can be used to compare the fit with other model specifications.

The “Covariance Parameter Estimates” table contains the estimate of the random intercept variance, 0.2183, and its standard error, 0.01175. The “Solutions for Fixed Effects” table contains the estimate of the overall average hospital intercept, 1.2606, which represents the hospital average for log(LOS-1), because the link function is the logarithm. This null model is useful for estimating the magnitude and significance of the between-hospital variance, and not much else.



## Random Intercepts, Fixed Slopes

The SAS code for the random intercepts, fixed slopes model is as follows:

```
proc glimmix data=in.nis_diabetes noitprint noclprint ;
  class hospid ;
  model los_1 = CAge / dist=poisson link=log solution ;
  random intercept / sub=hospid ;
  parms / noiter ; * This statement not generally necessary ;
  Title 'One-way ANCOVA with random effects for LOS-1';
run;
```

Notice that nothing has changed except for the Model statement, which now includes the additional fixed effect for CAge.

### Results

The key output is as follows:

Fit Statistics	
-2 Res Log Pseudo-Likelihood	63856.68
Generalized Chi-Square	57610.09
Gener. Chi-Square / DF	5.83

Covariance Parameter Estimates			
Cov Parm	Subject	Estimate	Standard Error
UN(1,1)	HOSPID	0.2110	0.01181

Solutions for Fixed Effects					
Effect	Estimate	Standard Error	DF	t Value	Pr >  t
Intercept	1.0554	0.01874	840	56.32	<.0001
CAge	0.1231	0.002675	9048	46.02	<.0001

### Interpretation

Notice that the estimate of the ratio of the Generalized Chi-Square statistic has diminished from 6.14 to 5.83, indicating that the current model, which includes age, is a better fit than the null model. The intercept variance is slightly smaller in this model, 0.2110, compared with 0.2183 in the null model. The estimated coefficient of CAge is 0.1231. Recall that this Poisson regression models the average of  $\log(\text{LOS}-1)$ . Consequently, the exponentiated coefficient,  $1.13 = \exp(0.1231)$ , indicates that the average patient LOS increases about 13 percent for every decade increase in patient age.

## Random Intercepts, Random Slopes

The SAS code for the random intercepts, random slopes model is as follows:

```
proc glimmix data=in.nis_diabetes noitprint noclprint ;
  class hospid ;
  model los_1 = CAge / dist=poisson solution ;
  random intercept CAge / type=un sub=hospid ;
  parms / noiter ; * This statement not generally necessary ;
  Title 'Random intercepts, random slopes model for LOS-1';
run;
```

Compared with the model in the previous section, the RANDOM statement includes two random effects, one for the INTERCEPT and one for the CAge slopes, both of which will vary by hospid (sub=hospid). An unstructured covariance option is selected (type=un) for the variance of the random intercepts, random slopes, and their covariance. The remaining parameters are explained in the previous two sections.

### Results

The key output is as follows:

Fit Statistics	
<b>-2 Res Log Pseudo-Likelihood</b>	59477.84
<b>Generalized Chi-Square</b>	52045.03
<b>Gener. Chi-Square / DF</b>	5.26

Covariance Parameter Estimates			
Cov Parm	Subject	Estimate	Standard Error
UN(1,1)	HOSPID	0.2743	0.01955
UN(2,1)	HOSPID	-0.04368	0.005683
UN(2,2)	HOSPID	0.02562	0.002439

Solutions for Fixed Effects					
Effect	Estimate	Standard Error	DF	t Value	Pr >  t
<b>Intercept</b>	1.0664	0.02211	840	48.23	<.0001
<b>CAge</b>	0.1173	0.007223	740	16.23	<.0001

### Interpretation

In the table "Covariance Parameter Estimates," the value of UN(1,1), 0.2743, is the estimated intercept variance; the value of UN(2,2), 0.02562, is the estimated slope (of CAge) variance; and the value of UN(2,1), -0.04368, is the estimated covariance between hospital intercepts and

hospital slopes. The corresponding standard errors are small, suggesting that hospital intercepts and slopes vary significantly and that the correlation between them is negative (see figure 7(d)). The estimated intercept/slope correlation is  $-0.52 = -0.04368 / \sqrt{(0.2743)(0.02562)}$ . Therefore, hospitals with higher intercept values tend to have smaller slopes.

The average coefficient for CAge is estimated as 0.1173. Therefore, the average LOS increases by an estimated 12 percent for every decade increase in age at the average hospital (the multiple of CAge is  $1.1245 = \exp(0.1173)$ ).

### *Intercept- and Slopes-As-Outcomes*

The SAS code for the intercept- and slopes-as-outcomes model is as follows:

```
proc glimmix data=in.nis_diabetes noitprint noclprint ;
  class hospid ;
  model los_1 = CAge Teach CAge*Teach / dist=Poisson solution ;
  random intercept CAge / type=un sub=hospid ;
  parms / noiter ; * This statement not generally necessary ;
  Title 'Intercepts and slopes as outcomes model for LOS-1';
run;
```

The only change from the random-intercepts, random-slopes model described in the previous section is the addition of the interaction term, CAge\*Teach.

### Results

The key output is as follows:

Fit Statistics	
<b>-2 Res Log Pseudo-Likelihood</b>	59407.34
<b>Generalized Chi-Square</b>	51960.03
<b>Gener. Chi-Square / DF</b>	5.26

Covariance Parameter Estimates			
Cov Parm	Subject	Estimate	Standard Error
<b>UN(1,1)</b>	HOSPID	0.2767	0.01993
<b>UN(2,1)</b>	HOSPID	-0.04336	0.005640
<b>UN(2,2)</b>	HOSPID	0.02528	0.002378

Solutions for Fixed Effects					
Effect	Estimate	Standard Error	DF	t Value	Pr >  t
Intercept	1.0125	0.02562	839	39.52	<.0001
CAge	0.1166	0.008400	739	13.88	<.0001
Teach	0.2280	0.05154	8308	4.42	<.0001
CAge*Teach	0.005448	0.01630	8308	0.33	0.7382

### Interpretation

The estimate for the random intercept variance is 0.2767, the random slope variance estimate is 0.02528, and the estimate for the covariance between the intercepts and slopes is -0.04336. These estimates are very close to the estimates from the previous model.

The estimates of “fixed” effects, patient age and hospital teaching status, are both statistically significant. On average, the LOS is about 25 percent higher at teaching hospitals than it is at non-teaching hospitals (the multiple is  $1.256 = \exp(0.2280)$ ). Of course, this difference might be explained by differences in case-mix severity between teaching and non-teaching hospitals, which could be tested by adding patient severity measures to the regression. The interaction between age and teaching status is not statistically significant. Consequently, the effect of age on patient LOS does not differ appreciably between teaching and non-teaching hospitals.

### Stata Programs

#### *Random Intercepts, Empty Model*

The Stata command is:

```
gllamm los_1, i(hospid) link(log) family(poisson) adapt
```

The Stata command GLLAMM is used to obtain maximum likelihood estimates for the empty model. The dependent variable is los\_1 and no predictor variables are entered. The options family(poisson) and link(log) are specified to model the outcome. The option i(hospid) indicates that patients are grouped by hospital. The “adapt” option requests that adaptive quadrature be used for numerical integration in maximizing the likelihood function. This method often results in better estimates than those produced by other methods (Rabe-Hesketh et al., 2002).

### Results

The key output is as follows:

```
gllamm model
```

```
log likelihood = -36907.51
```

```
-----+-----
      los_1 |      Coef. Std. Err.      z    P>|z|     [95% Conf. Interval]
-----+-----
      _cons |     1.21957   .0213992    56.99   0.000     1.177629     1.261512
```

-----  
**Variiances and covariances of random effects**  
 -----

\*\*\*level 2 (hospid)

var(1): .29732598 (.01917026)  
 -----

Interpretation

The estimate of the random intercept for the average hospital is 1.2196, which is the average of the log(LOS-1) for the average hospital. The hospital-to-hospital intercept variance is estimated to be 0.2973, with a standard error of 0.0192.

*Random Intercepts, Fixed Slopes*

The Stata command is:

```
gllamm los_1 cage, i(hospid) link(log) family(poisson) adapt
```

Here, cage is added as a “fixed” predictor to the null model described in the previous section.

Results

The key output is as follows:

**gllamm model**

**log likelihood = -35785.412**

-----

los_1	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
cage	.1231055	.0026815	45.91	0.000	.1178498	.1283612
_cons	1.017756	.0213553	47.66	0.000	.9759006	1.059612

-----

-----  
**Variiances and covariances of random effects**  
 -----

\*\*\*level 2 (hospid)

var(1): .27581157 (.01802368)  
 -----

### Interpretation

The estimate of the average intercept is 1.0178, conditional on patient age. The estimated age coefficient is 0.1231, indicating that LOS increases by about 13 percent for every decade increase in age (the multiple is  $1.1310 = \exp(0.1231)$ ). The estimated variance of the random intercepts is 0.2758, with a standard error of 0.0180.

### *Random Intercepts, Random Slopes*

The Stata commands are:

```
gen cons = 1
eq cage = cage
eq cons = cons
gllamm los_1 cage, i(hospid) link(log) family(poisson) nrf(2) eqs(cage
cons) adapt
```

For random slope models, gllamm requires all random effects to be specified using the eqs() option. Its arguments are “equations” defined by “eq” commands. Consequently, the first command in the Stata program, “gen” creates (the intercept variable) cons, which is assigned a value of 1 for every case. The second and third (eq) commands create “equations” for age and the intercept. Compared with the previous model, with cage as a fixed effect, the gllamm command includes the option eqs(cage cons). It is important to pay attention to the order of the arguments in the eqs() option because the variances for random effects are labeled in the same order.

### Results

The key output is as follows:

**gllamm model**

**log likelihood = -35104.891**

los_1	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
cage	.1193281	.0071251	16.75	0.000	.1053632	.133293
_cons	1.015129	.0244502	41.52	0.000	.9672074	1.06305

### **Variances and covariances of random effects**

**\*\*\*level 2 (hospid)**

```
var(1): .02109545 (.00179995)
cov(2,1): -.03559773 (.00521775) cor(2,1): -.42943507
var(2): .32573232 (.02440816)
```

### Interpretation

The estimates of the fixed effects are 0.1193 for the average age coefficient, and 1.0151 for the average intercept. Exponentiating the average age coefficient gives a multiple of 1.127 for the effect of each decade of age on ALOS at the average hospital. That is, ALOS increases an average of 12.7 percent for every additional decade of patient age at the average hospital.

The estimate of the variance for the random age coefficients is 0.0211 and the estimate for the random intercept variance is 0.3257. Remember that these variances, var(1) and var(2), are numbered in the same order as they are listed in the eqs() option. The estimated covariance between the random intercepts and random slopes is -0.0356, yielding an estimated correlation of -0.43. This implies that hospitals with larger intercepts tend to have smaller slopes (as in figure 7(d)).

### *Intercept- and Slopes-As-Outcomes*

The Stata commands are:

```
gen cons = 1
gen teachXcage = teach * cage
eq cage = cage
eq cons = cons
gllamm los_1 cage teach teachXcage, i(hospid) link(log) family(poisson)
nrf(2) eqs(cage cons) adapt
```

This model contains commands and options identical to the model in the previous section, with the addition that a new variable, teachXcage, is created as the product of cage and teach, which is entered as an additional predictor. Teach is a hospital-level variable equal to 1 for teaching hospitals and equal to 0 for non-teaching hospitals.

### Results

The key output is as follows:

**gllamm model**

**log likelihood = -35089.15**

los_1	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
cage	.1188993	.0083335	14.27	0.000	.1025661	.1352326
teach	.2636844	.0548162	4.81	0.000	.1562466	.3711222
teachXcage	.0039677	.0153416	0.26	0.796	-.0261013	.0340367
_cons	.9579603	.0277979	34.46	0.000	.9034774	1.012443

Variiances and covariances of random effects

---

\*\*\*level 2 (hospid)

var(1): .02119057 (.00180921)  
cov(2,1): -.03581881 (.00515187) cor(2,1): -.43708377  
  
var(2): .31692054 (.02372454)

---

Interpretation

The fixed effects are all significant except for the interaction between patient age and hospital teaching status. The estimated coefficient for Teach is 0.2637, indicating that the ALOS for teaching hospitals is about 30 percent higher for teaching hospitals than that it is for non-teaching hospitals (note:  $\exp(0.2637) = 1.3017$ ). Of course, this difference might be explained by different case-mix severity between teaching and non-teaching hospitals, which could be tested by adding a measure of patient severity as a predictor variable. The average coefficient for age is 0.1189, implying that ALOS increases (at the average hospital) by about 12.6 percent for every decade increase in patient age.



## RECOMMENDATIONS

Hierarchical models should be one of many statistical techniques in every HCUP analyst's toolkit. It is a useful regression methodology for nested data: discharges within hospitals, visits within patients, patients within physicians, and so on. When properly specified, these models can lead to better inferences and yield more information than is available from traditional, standard regression models.

Hierarchical data structures almost always induce within-group correlation that should not be ignored. Statistical packages like Stata, SUDAAN, and SAS provide procedures for survey data that explicitly account for sampling designs involving probability samples with stratification and clustering, such as the designs employed for some HCUP databases. Also, the Generalized Estimating Equations method properly accounts for clustering and grouping for the purpose of estimating regression coefficients and their standard errors. However, analysts who want more flexibility in their models or who want to analyze the variances and covariances at each level of their data structure should employ HLM.

The analyst should begin with the most basic models, starting with the (empty) components of variance model and estimating the intraclass correlation coefficient (ICC) to determine the relative magnitude of the within-group and between-group variance. Then, assuming that the ICC is positive, predictors should be added to the random-intercepts model, based on substantive theory, while monitoring the reduction in variance at each level of the hierarchy. Complexity should be added only if the model is found to be deficient. Random slopes might be suggested either by the data or by theory. However, the analyst should be wary of making the model overly-complicated because it makes estimation and interpretation more difficult.

There are numerous other resources available to aid further study and to help researchers fit these models to their data. The introductory texts listed in the bibliography are a good place to start. In addition, the recent book by Twisk (2006), which is available in paperback, provides a very readable introduction to HLM written especially for non-statistician researchers. There are several good Websites containing downloadable data, programs, papers, presentations, and links to other HLM Websites. Two especially comprehensive sites are the Centre for Multilevel Modelling Website: <http://www.cmm.bristol.ac.uk/> and the UCLA Academic Services Website: <http://www.ats.ucla.edu/stat/>.

## REFERENCES

- Amaravadi JB, Dimick PJ, Pronovost, PA. ICU nurse-to-patient ratio is associated with complications and resource use after esophagectomy. *Intensive Care Medicine*, 26(12): 1857-1862, 2000.
- Austin PC, Tu JV, Alter DA. Comparing hierarchical modeling with traditional logistics regression analysis among patients hospitalized with acute myocardial infarction: Should we be analyzing cardiovascular outcomes data differently? *American Heart Journal*, 145(1): 27-35, 2003.
- Bramble JD, Sakowski H, Rich EC, Esterbrooks, D. Resource use in treating alcohol- and drug-related diagnoses. Does teaching status and experience matter? *Journal of General Internal Medicine*, 19: 36-42, 2004.
- Browne, WJ, Draper, D. A comparison of Bayesian and likelihood-based methods for fitting multilevel models. *Bayesian Analysis*, 1: 473-514, 2005.
- Browne, WJ, Goldstein, H, Rasbash, J. Multiple membership multiple classification (MMMC) models. *Statistical Modelling*, 1: 103-124, 2001.
- Bryk, AS, Raudenbush, SW, Congdon, R. *Hierarchical Linear and nonlinear modeling with the HLM/2L and HLM/3L Programs*. Chicago: Scientific Software International, 1996.
- Carey, K. Hospital length of stay and cost: A multilevel modeling analysis. *Health Services & Outcomes Research Methodology*, 3: 41-56, 2002.
- Dowell MA, Rozell B, Roth D, Delugach H, Chaloux P, Dowell J. Economic and clinical disparities in hospitalized patients with type 2 diabetes. *Journal of Nursing Scholarship*, 36: 1-66, 2004.
- Dubois RW, Brook RH, Rogers, WH. Adjusted hospital death rates: Potential screen for quality of medical care. *American Journal of Public Health*, 77: 1162-1167, 1987.
- Duncan C, Jones K, Moon G. Context, composition and heterogeneity: Using multilevel models in health services research. *Social Science and Medicine*, 46(1): 97-117, 1998.
- Epstein, A. Performance reports on quality-prototypes, problems and prospects. *New England Journal of Medicine*, 333: 57-61, 1995.
- Estabrooks CA, Midodzi WK, Cummings GG, Ricker KL, Giovannetti P. The Impact of Hospital Nursing Characteristics on 30-Day Mortality. *Nursing Research*, 54(2): 74-84, 2005.
- Gatsonis CA, Epstein AM, Newhouse JP, Normand SL, McNeil BJ. Variations in the utilization of coronary angiography for elderly patients with an acute myocardial infarction: An analysis using hierarchical logistic regression. *Medical Care*, 33: 625-642, 1995.
- Goldstein HI. *Multilevel Statistical Models* (Third Edition). London: Edward Arnold, 2003.
- Goldstein HI, Browne W, & Rasbash J. Tutorial in biostatistics: multilevel modelling of medical data. *Statistics in Medicine*, 21: 3291-3315, 2002.
- Goldstein H, Spiegelhalter D. League tables and their limitations: Statistical issues in comparisons of institutional performance. *Journal of the Royal Statistical Society, Series A*, 159: 385-444, 1996.
- Hox, J. *Multilevel Analysis: Techniques and Applications*. Mahwah, NJ: Erlbaum, 2002.
- Jencks SF, Daley J, Draper D, Thomas N, Lenhart G, Walker J. Interpreting hospital mortality data, *Journal of the American Medical Association*, 260: 3611-3616, 1988.

- Joshi AV, D'Souza AO, Madhavan SS. Differences in hospital length-of-stay, charges, and mortality in congestive heart failure patients. *Congestive Heart Failure*, 10: 76-84, 2004.
- Krumholz, HM, Wang, Y, Mattera, JA, Wang, Y, Han, LF, Ingber, MJ, Roman, S, Normand, SL. An administrative claims model suitable for profiling hospital performance based on 30-day mortality rates among patients with an acute myocardial infarction. *Circulation*, 113: 1683-1692.
- Leyland AH, Boddy FA. League tables and acute myocardial infarction, *Lancet*, 351: 555-558, 1998.
- Leyland, AH, Goldstein, H. (eds.) *Multilevel Modelling of Health Statistics*, London: Wiley, 2001.
- Liang K, Zeger SL. Longitudinal data and analysis using generalized linear models. *Biometrika*, 73: 45-51, 1986.
- Littell RC, Milliken GA, Stroup WW, Wolfinger RD, Schabenberger, O. *SAS for Mixed Models* (Second Edition), Carey, NC: SAS Institute, 2006.
- Localio RA, Hamony BH, Fisher AC, TenHave TR. The public release of hospital and physician mortality data in Pennsylvania: A case study. *Medical Care*, 35(3): 272-286, 1997.
- Marshall EC, Spiegelhalter DJ. "Institutional Performance." In Goldstein H, Leyland AH (eds.). *Multilevel Modelling of Health Statistics*, NY: Wiley, 2001.
- Molenberghs G, Verbeke, G. Meaningful statistical model formulations for repeated measures. *Statistica Sinica*, 14: 989-1020, 2004.
- Morris CN, Christiansen CL. "Hierarchical Models for Ranking and for Identifying Extremes, with Application." In Bernardo JO, Berger Dawid AP (eds.), *Bayesian Statistics 5*, Oxford University Press, 1996: 277-297.
- Normand SL, Glickman ME, Ryan TJ. "Modelling Mortality Rates for Elderly Heart Attack Patients: Profiling Hospitals." In Gatsonis C, et al. (eds.). *The Cooperative Cardiovascular Project. Case Studies in Bayesian Statistics*. NY: Springer-Verlag, 1995: 435-456.
- Normand SL, Glickman ME, Gatsonis CA. Statistical models for profiling providers of medical care: Issues and applications. *Journal of the American Statistical Association*, 92: 803-814, 1997.
- Rabe-Hasketh SA, Skrondal A. *Multilevel and Longitudinal Models Using Stata*. College Station, TX: Stata Press, 2005a.
- Rabe-Hesketh SA, Skrondal A. Multilevel modeling of complex survey data. *Journal of the Royal Statistical Society, Series A*. In press, 2005b.
- Rabe-Hesketh SA, Skrondal A, Pickles A. Reliable estimation of generalized linear models using adaptive quadrature. *Stata Journal*, 2(1): 1-21, 2002.
- Rasbash J, Steele F, Browne W, Prosser B. *A User's Guide to MLwiN*. Centre for Multilevel Modelling, University of Bristol, 2005.
- Raudenbush SW, Bryk AS. *Hierarchical Linear Models: Applications and Data Analysis Methods* (Second Edition). Thousand Oaks, CA: Sage, 2002.
- Rice N, Leyland A. Multilevel models: Applications to health data. *Journal of Health Services Research and Policy*, 1(3): 154-164, 1996.
- Schneider EC, Epstein AM. Influence of cardiac surgery performance reports on referral practices and access to care. *New England Journal of Medicine*, 335: 251-256, 1996.

Scott I, Youlden D, Coory M. Are diagnosis specific outcome indicators based on administrative data useful in assessing quality of hospital care? *Quality and Safety in Healthcare*, 13: 32-39, 2004.

Shahian DM, Torchiana DF, Shemin RJ, Rawn JD, Normand ST. Massachusetts Cardiac Surgery Report Card: Implications of Statistical Methodology. *Annals of Thoracic Surgery*, 80: 2106-2113, 2005.

Singer JD, & Willett JB. *Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence*. New York: Oxford Press, 2003.

Snijders TAB, & Bosker RJ. *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling*. London: Sage, 1999.

Spiegelhalter DJ, Aylin P, Best NG, Evans SJW and Murray GD. Commissioned analysis of surgical performance using routine data: lessons from the Bristol inquiry. *Journal of the Royal Statistical Society: Series A*, 165(2): 191-221, 2002.

Subramanian SV. "Multilevel Methods, Theory, and Analysis," In Anderson, NB (Ed.) *Encyclopedia of Health & Behavior*, Thousand Oaks, CA: Sage, 2004: 602-609.

Subramanian SV, Jones K, Duncan C. "Multilevel Methods For Public Health Research," In Kawachi I, Berkman L. (Eds.) *Neighborhoods and Health*, New York: Oxford University Press, 2003: 65-111.

Sullivan LM, Dukes KA, Losina E. Tutorial in biostatistics: An introduction to hierarchical linear modeling. *Statistics in Medicine*, 18: 855-888, 1999.

Thomas N, Longford NT, Rolph JE. Empirical Bayes methods for estimating hospital specific mortality rates. *Statistics in Medicine*, 13: 889-903, 1994.

Twisk, JWR. *Applied Multilevel Analysis: A Practical Guide for Medical Researchers*. Cambridge University Press: 2006.

## BIBLIOGRAPHY

The authors of this report would like to thank Constantine Gatsonis, Ph.D. (Brown University), Sharon-Lise Normand, Ph.D. (Harvard), and Alan Zaslavsky, Ph.D. (Harvard), for providing this bibliography, which was distributed at their presentation on HLM at the 2006 Annual AcademyHealth Meeting.

### General Introductions (in order of increasing difficulty):

Singer JD and Willett JB, *Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence*, Oxford University Press, 2003.

Snijders T and Bosker R, *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling*, Sage Publishers, 1999.

Bryk AS and Raudenbush SW, *Hierarchical Linear Models (2<sup>nd</sup> edition)*, Sage Publications, 2002.

Hox J, *Multilevel Analysis, Techniques and Applications*, Lawrence Erlbaum Associates, 2002. (Sample chapters available for download at <http://www.fss.uu.nl/ms/jh/mlbook/leabook.htm>)

Fitzmaurice G, Laird N, and Ware J, *Applied Longitudinal Analysis*, John Wiley and Sons, 2004.

Leyland AH and Goldstein H, *Multilevel Modeling of Health Statistics*, Wiley, 2001.

Spiegelhalter DJ, Abrams KR, Myles JP, *Bayesian Approaches to Clinical Trials and Health-Care Evaluation*, Wiley, 2004.

Goldstein H, *Multilevel Statistical Models*, Edward Arnold, 1995. (Available for free download, <http://www.arnoldpublishers.com/support/goldstein.htm>.)

McCulloch CE and Searle SR, *Generalized, Linear and Mixed Models*, Wiley, 2000.

Longford NT, *Random Coefficient Models*, Clarendon (Oxford University Press), 1993.

### Selected applications in health services research:

McNeil BJ, Pederson S, Gatsonis C. Current issues in profiling quality of care. *Inquiry* 1992; **29**:298.

Gatsonis CA, Normand SL, Liu C, Morris C. Geographic variation of procedure utilization: A Hierarchical Model Approach. *Medical Care* 1993; **31**:YS54-YS59.

Gatsonis CA, Epstein AM, Newhouse JP, Normand SL, McNeil BJ. Variations in the utilization of coronary angiography for elderly patients with an acute myocardial infarction: An analysis using hierarchical logistic regression. *Medical Care* 1995; **33**:625-642.

Goldstein H and Spiegelhalter D. League tables and their limitations: Statistical issues in comparisons of institutional performance. *Journal of the Royal Statistical Society, Series A*, 1996; **159**:385-444.

Normand SL, Glickman ME, Gatsonis CA. Statistical models for profiling providers of medical care: Issues and applications. *Journal of the American Statistical Association* 1997; **92**:803-814

Gatsonis, C.A. Hierarchical Models in Health Services Research. *Encyclopedia of Biostatistics*, New York: Wiley, 1998.

Daniels MJ and Gatsonis CG. Hierarchical Generalized Linear Models in the analysis of variations in health care utilization. *Journal of the American Statistical Association* 1999; **94**:29-42.

Landrum MB, McNeil BJ, Silva L, Normand SL. Understanding variability in physician ratings of the appropriateness of coronary angiography after Acute Myocardial Infarction. *Journal of Clinical Epidemiology* 1999; **52**:309-319.

Landrum MB, Normand SLT. Applying Bayesian ideas to the development of medical guidelines. *Statistics in Medicine* 1999;**18**:117-137.

Landrum, M.B, Bronskill, S.E. and Normand, S-L. Analytic methods for constructing cross-sectional profiles of health care providers. *Health Services and Outcomes Research Methodology*, 2000, 1: 23-47.

Burgess, J., Lourdes, West, M. Profiling Substance Abuse Provider Trends in Health Care Delivery Systems. *Health Services and Outcomes Research Methodology*, 2000; 1: 253-276.

Spiegelhalter D, Aylin P, Best N, Evans SJW, and Murray, GD. Commissioned analysis of surgical performance by using routine data: lessons from the Bristol inquiry. *Journal of the Royal Statistical Society, Series A* 2002; 165: 1-31.

Normand S-LT, Zou KH. Sample size considerations in observational health care quality studies. *Statistics in Medicine* 2001;**21**:331-345.

Bronskill SE, Normand S-LT, Landrum MB, Rosenheck RA. Longitudinal profiles of health care providers. *Statistics in Medicine* 2002;**21**:1067-1088.

Landrum, M.B., Normand, S-L., Rosenheck R.A. Selection of related multivariate means: Monitoring psychiatric care in the Department of Veterans Affairs. *Journal of the American Statistical Association*, 2003;98(461):7-16.

Zaslavsky AM, Zaborski LS, Cleary PD. Plan, geographical, and temporal variation of consumer assessments of ambulatory health care. *Health Services Research* 2004;**39**(5):1467-1486.

Shahian DM, Torchiana DF, Shemin RJ, Rawn JD, Normand S-LT. The Massachusetts cardiac surgery report card: implications of statistical methodology. *Annals of Thoracic Surgery*, 2005;80:2106-13.

Daniels MJ and Normand S-LT. Longitudinal profiling of health care units based on continuous and discrete patient outcomes. *Biostatistics* 2006;7(1):1-15.

### **Multilevel Models Project home page (sponsors of MLwiN software):**

<http://multilevel.ioe.ac.uk/> with links to other software for multilevel modeling

Multilevel modeling newsgroup/discussion list: <http://www.jiscmail.ac.uk/lists/multilevel.html>

### **Widely-used software for hierarchical modeling (not complete!):**

BUGS (General-purpose Bayesian modeling software, well adapted to development of fully Bayesian nonstandard models.) <http://www.mrc-bsu.cam.ac.uk/bugs/>

HLM (Standalone package for hierarchical models.) <http://www.ssicentral.com/hlm/hlm.htm>

MLwiN (Another standalone package, with Windows and command line interface. Both maximum likelihood and Bayesian approaches to inference.) <http://multilevel.ioe.ac.uk/features/index.html>

SAS: PROC MIXED for maximum likelihood estimation of linear mixed models; also PROC GLIMMIX (in newest versions of SAS, <http://support.sas.com/rnd/app/da/glimmix.html>) and the GLMMIX macro(similar functionality for older versions) for generalized (logistic, Poisson, etc.) mixed models. <http://www.sas.com/>

See article by Judith Singer at <http://gseweb.harvard.edu/~faculty/singer/sasprocmixed.pdf> (also included in handouts for this tutorial).

MIXOR/MIXREG: Free standalone package by Hedeker and Gibbons)  
<http://www.uic.edu/~hedeker/mix.html>

Stata: gllamm procedure. (Fairly general function operating within this general-purpose statistical package.) <http://www.Stata.com>, <http://www.gllamm.org>