COMPARISON OF LOGISTIC REGRESSION AND MANTEL-HAENSZEL STATISTICAL PROCEDURES TO PREDICT LENGTH OF STAY OF FOUR DIAGNOSIS-RELATED GROUPS

by

REZA ZIAEE

DISSERTATION

Submitted to the Graduate School

of Wayne State University,

Detroit, Michigan

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

2007

MAJOR: EDUCATIONAL EVALUATION AND RESEARCH

Approved by:

Sheldon S. Sacks 9/25/07
Advisor Date

Gail Sahoome

Marc B. Rose

Julie Worth Cain

ProQuest Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

Dedication

This academic achievement is dedicated to the memory of my Professor and Advisor,

Professor Donald Marcotte,

for his support and continuous encouragement

during the demanding years of the PhD process,

and my mother,

Masoumeh Ghaemi,

an illiterate widow who devoted her life to my educational success

and used to say

“Knowledge never leaves its owner alone and destitute”

so learn and never say it is enough.

Acknowledgments

This academic endeavor would not have been possible without the help, assistance, and support of many committee members, colleagues, friends, and family members. This support came in different shape and forms.

First, special heartfelt thanks to my doctoral committee members. I am grateful to my advisor, Dr. Shlomo Sawilowsky, for his leadership, confidence and support from the beginning and to my committee members, Dr. Marc Rosa, and Dr. Leslie Monplaisir, for their precious time, suggestions and advice during this process.

I would like to recognize Kim Repac from Western Maryland Health System for all the support and encouragement during this process. I also would like to thank my current leader, Dawn Ahner from Renown Health for her understanding and support by providing me with extra time to complete my dissertation.

I especially would like to thank June Cline for her willingness to provide insight, encouragement and collegial support during the proposal and dissertation development stages. Special thanks to my great friends, James Bologna and Michael Dobra for their continuous support and encouragements when I became overwhelmed.

And finally to my loving wife Rose Cleverley, without her encouragement, patience, and support this doctoral dissertation could not have been completed.

Table of Contents

List of Tables

List of Figures

Chapter I

Introduction

Background

In the past, healthcare data were explained using simplistic descriptive statistics. One prevalent method that has been popular in analyzing healthcare data has been linear regression models. Logistic regression comes from a set of statistical tools referred to as the General Linear Model (GLM) family. GLM is the unification of both linear and nonlinear regression models that allows the incorporation of nonnormal response distributions. In a GLM, the response variable distribution must be only a member of the exponential family which includes normal, Poisson, binomial, exponential and gamma distribution as members (Neter, Nachtsheim, & Wasserman, 1996). Typical use of logistic regression produces two possible outcomes, called success or failure and is denoted by a 0 or a 1. The variable is either a contributing factor of the dependent variable or it is not. For example, when using a yes or no response in a poll, other factors such as geographic location or demographic data (e.g., family size or income) may be important to consider. A logistic regression model could be used to model the probability that a demographic or geographic factor has an effect on the overall response to the question (Rogers, 1993).

Logistic Regression

From a practical standpoint, a logistic regression method produces predictive equations. In logistic regression, the regression coefficients measure the predictive capability of each independent variables.

The response variable that characterizes logistic regression is what differs it from other types of regression analysis. With logistic regression, the response variable is an indicator of some characteristic, that is, a dichotomous variable. Logistic regression is used to determine if other measures are related to the presence of some characteristic. For example, certain blood measures

could be predictive of a particular disease. If analysis of covariance can be said to be a t test adjusted for other variables, then logistic regression can be thought of as a chi-square test for homogeneity of proportions adjusted for other variables (Holland, & Thayer, 1986).

Logistic Regression Models

The central mathematical concept that underlies logistic regression is the logit – the natural logarithm of an odds ratio. The simplest example of a logit derives from a 2×2 contingency table.

Generally, logistic regression is appropriate for describing and testing hypotheses about relationships between a categorical outcome variable and one or more categorical or continuous predictor variables. In the simplest case of linear regression for one continuous predictor, *X* (a child's reading score on a standardized test) and one dichotomous criterion variable *Y* (the child being recommended for remedial reading classes), the plot of such data results in two parallel lines, each corresponding to a value of the dichotomous outcome. Because the two parallel lines are difficult to describe with an ordinary least squares regression equation due to the dichotomous nature of the criterion variable, categories may be created for the predictor and the mean computed for the criterion variable for the respective categories.

The resultant plot of category means generally appears to be linear in the middle, typical of what one would expect to see on an ordinary scatter plot, but curved at the ends. Such a shape, often referred to as sigmoidal or S-shaped, is difficult to describe with a linear equation for two reasons. First, the extremes do not follow a linear trend. Second, the errors are neither normally distributed nor constant across the entire range of data (Peng, Manz, & Keck, 2001). Logistic regression solves these problems by applying the logit transformation to the dependent variable. In essence, the logistic model predicts the logit of *Y* from *X*. As stated earlier, the logit is the natural logarithm (ln) of odds of *Y*, and odds are ratios of probabilities (π) of *Y* happening (i.e., a student is recommended for remedial reading instruction) to probabilities ($1 - \pi$) of *Y* not occurring (i.e., a

student is not recommended for remedial reading instruction). Logistic regression can accommodate categorical outcomes that are polytomous. However, for the present study, the focus is on dichotomous outcomes. The illustration presented in this article can be extended easily to polytomous variables with ordered (i.e., ordinal-scaled) or unordered (i.e., nominal-scaled) outcomes.

The simple logistic model takes the form

$$\text{Logit}(Y) = \textit{natural}\ \log(\textit{odds}) = \ln[\pi/(1-\pi)] = \alpha + \beta X \quad (1)$$

Taking the antilog of Equation 1 on both sides, one derives an equation to predict the probability of the occurrence of the outcome of interest as follows:

$$\pi = \textit{Probability}\ (Y = \text{outcome of interest} \mid X = x, \text{ a specific value of } X) =$$

$$(e^{\alpha + \beta x})/(1 + e^{\alpha + \beta x}) \qquad (2)$$

where π is the probability of the outcome of interest or "event," such as a child's referral for remedial reading classes, α is the Y intercept, β is the regression coefficient, and $e = 2.71828$ is the base of the system of natural logarithms. X can be categorical or continuous, but $\boldsymbol{Y}$ is always categorical. According to Equation 1, the relationship between logit (Y) and X is linear. Yet, according to Equation 2, the relationship between the probability of Y and X is nonlinear. For this reason, the natural log transformation of the odds in Equation 1 is necessary to make the relationship between a categorical outcome variable and its predictor(s) linear.

The value of the coefficient β determines the direction of the relationship between X and the logit of Y. When β is greater than zero, larger (or smaller)X values are associated with larger (or smaller) logits of Y. Conversely, if β is less than zero, larger (or smaller) X values are associated with smaller (or larger) logits of Y. Within the framework of inferential statistics, the null hypothesis states that β equals zero, or there is no linear relationship in the population. Rejecting such a null hypothesis implies that a linear relationship exists between X and the logit of

Y. If a predictor is binary, then the odds ratio is equal to *e*, the natural logarithm base, raised to the exponent of the slope β (e^{β};).

Extending the logic of the simple logistic regression to multiple predictors (say X_1=reading score and X_2= gender),) one can construct a complex logistic regression for *Y* (recommendation for remedial reading programs) as follows:

$$\text{Logit}(Y) = \ln[\pi/(1-\pi)] = \alpha + \beta_1 X_1 + \beta_2 X_2 \quad (3)$$

Therefore,

$$\pi = \textit{Probability}\ (Y = \text{outcome of interest}\ X_1 = x_1, X_2 = x_2,$$

$$(e^{\alpha + \beta_1 x_1 + \beta_2 x_2})/(1 + e^{\alpha + \beta_1 x_1 + \beta_2 x_2}) \quad (4)$$

The probability of the event is once again π, α is the *Y* intercept, βs are standardized regression coefficients, and *X*s are a set of predictors. α and βs are typically estimated by the maximum likelihood (ML) method. This type of analysis is preferred over the weighted least squares approach by several authors, such as Haberman (1978) and Schlesselman (1982). The ML method was designed to maximize the likelihood of reproducing data given parameter estimates. Data are entered into the analysis as 0 or 1 coding for the dichotomous outcome, continuous values for continuous predictors, and dummy codings (e.g., 0 or 1) for categorical predictors.

The null hypothesis underlying the overall model states that all βs equal zero. A rejection of the null hypothesis implies that at least one β does not equal zero in the population, indicating that the logistic regression equation predicts the probability of the outcome better than the mean of the dependent variable *Y*. The interpretation of results can be presented using the odds ratio for both categorical and continuous predictors.

Mantel-Haenszel (MH) Procedure

The Mantel-Haenszel (MH) procedure is sensitive to one type of differential item functioning (DIF). It was not designed to detect DIF that has a nonuniform effect across trait levels. By generalizing the model underlying the MH procedure, a more general DIF detection procedure has been developed (Swaminathan & Rogers, 1990).

Sample-size restrictions limit the contingency table approaches based on asymptotic distributions, as the Mantel-Haenszel (MH) procedure, for detecting differential item functioning (DIF) in many practical applications. The detection of differences in item performance for different groups of examinees is important if tests equivalent for these groups are to be designed and maintained.

Differential item performance by Mantel-Haenszel.

According the Holland and Thayer (1986), the Mantel-Haenszel (MH) procedure (Mantel & Haenszel, 1959) is the best method for detecting this kind of item bias. The first step of the MH procedure is to identify contrasting dichotomous examinee groups (e.g., gender, ethnic group, etc.). These groups are labeled the reference group, R, chosen to provide the standard performance on the item of interest, and the focal group, F, whose differential performance, if any, is to be detected and measured.

The MH procedure requires that these groups be matched according to a relevant stratification. As external factors by which to match the strata of these groups are seldom clear, implied levels of ability are used. The ability range of the groups is divided into K score intervals (usually three to five), with these intervals used to match samples from each group. A 2 x 2 contingency table for each of these K ability intervals is constructed from the responses to the suspect item by the examinees of each group. The table of responses made by the two sample groups in the j^{th} ability interval has the form shown in Table 1.

Table 1

Table of Responses from Two Sample Groups

Response to Suspect Item			
Group j	Right (1)	Wrong (0)	Total
Reference group	A_j (P_{Rj1})	B_j (P_{Rj0})	
Focal group	C_j (P_{Fj1})	D_j (P_{Fj0})	
Total			T_j

Note: A_j is observed count. $P_{R\,j1}$ is latent probability.

The MH procedure is based on estimating the probability of a member of the reference group in interval j getting the item right, P_{Rj1}, or getting it wrong, P_{Rj0}, and similarly for a member of the focal group, P_{Fj1} and P_{Fj0}. Two statistics are derived: an estimate of the significance of the difference, and an estimate of the size of the difference. Often only the significance of the difference is reported. However, as meaninglessly small differences can be reported as significant for large samples, the discussion is concerned with the estimation of the size of the difference.

The MH estimate, α, of the difference in performance on an item between the two groups across all intervals is:

$$\alpha = \frac{\sum_{j}^{K} \frac{A_j D_j}{T_j}}{\sum_{j}^{K} \frac{B_j C_j}{T_j}}$$

in which α has a range from 0 to infinity, and a "no difference" null value of 1.

In summary, the Mantel_Haenszel Differential Item Functioning (DIF) method compares a single reference group and multiple focal groups. The Mantel and Haenszel (DIF) procedure assesses the degree of relationship of two variables, while controlling for one or more control variables. A particular situation that has received a great deal of attention is that of a relationship

between two dichotomous variables controlling for one or more categorical factors. When there is only one control variable or when the 2 x 2 relationship is examined within each combination of the levels of the control variables, the result is a 2 x 2 x K cross classification, where the K levels of the control variable or variable combinations are often referred to as strata.

For example the relationship between administration of a drug and remediation of disease effects while controlling for gender of patient, mode of administration, and/or other factors can be analyzed. A common way to assess relationships in 2 x 2 way tables is through the odds ratio. In this case one group is given a drug and another group a placebo. Following treatment, all patients are assessed for recovery, the odds ratio measures the increase (or decrease) in odds of recovery for patients given the active drug relative to those given the placebo. An odds ratio of 1 represents no effect, while a ratio greater than 1 indicates that the drug increases the odds of recovery and a ratio less than 1 indicates that it diminishes the odds of recovery. This odds ratio can be obtained for a 2 x 2 table as the Case Control Relative Risk estimate (Holland & Wainer, 1993).

Diagnosis-Related Groups

In 1965, the Social Security Act established both the Medicare and Medicaid programs. Medicare was the responsibility of the Social Security Administration (SSA), while Federal assistance to the State Medicaid programs was administered by the Social and Rehabilitation Service (SRS). SSA and SRS were agencies in the Department of Health, Education, and Welfare (HEW). In 1977, the Health Care Financing Administration was created under HEW to effectively coordinate Medicare and Medicaid. In 1980 HEW was divided into the Department of Education and the Department of Health and Human Services (HHS; CMS, 2002).

In the early 1960s and prior to the enactment of the Social Security Act, a national survey found that only 56% of those 65 years of age or older had any health insurance. President John F. Kennedy pressed legislators for health insurance for the aged. However, it was not until 1965 that

President Lyndon B. Johnson signed H.R. 6675 (The Social Security Act of 1965; PL 89-97) to provide health insurance for the elderly and the poor (CMS, 2002)

Medicare extended health coverage to almost all Americans aged 65 or older. About 19 million beneficiaries enrolled in Medicare in the first year of the program. Medicaid provided access to health care services for certain low-income persons and expanded the existing Federal-State welfare structure that assisted the poor (Sullivan & Toby, 1992).

By 2004 the Medicare program provided benefits to nearly 42 million people or14% of the total U.S. population. The Medicare population is demographically diverse and includes significant numbers of individuals who are financially and medically vulnerable. Additionally, 28% of Medicare beneficiaries report being in fair or poor health and nearly nine in 10 beneficiaries have one or more chronic illnesses. Over half report having hypertension (60%) or arthritis (58%). Roughly a quarter of all Medicare beneficiaries have a cognitive or mental impairment. (Kaiser Family Foundation 2005)

Almost immediately after the enactment of the Medicare and Medicaid programs, the cost of health care began to rise (Kotelchuck, 1976). In the attempt to identify causes of increases in healthcare services, HCFA studied and defined that about half the growth in real per capita medical spending from 1960 to 1993 and two-thirds of its growth from 1983 to 1993 resulted from either the level or the growth of insurance coverage. Dividing all factors to determine the 1960-1993 growth in real per capita medical spending into two major categories, researchers found that 70% of this growth resulted from increasing costs resulting from advances in medical services encouraged by insurance coverage levels and spending for noncommercial medical. Approximately 30% of the increase was attributable to standard factors: growth in insurance coverage, changes in age/sex mix, and growth in real per capita disposable income (Block, 1997).

A more recent study of the increases in health care costs identified similar underling

reasons for the increase in health care utilization and costs. The overall increase in commercial health care premiums between 2004 and 2005 was 8.8%, which is lower than the 13.7% increase reported in 2002. Breaking down the increase in to the component costs, increased utilization of services accounted for an estimated 43% (3.8 percentage points of the 8.8%), general inflation accounted for 27%, and price increases in excess of inflation for healthcare services accounted for the remaining 30% of the increase in health insurance premiums. The reasons for price increases in excess of inflation include movement among purchasers toward broader-access health plans, provider consolidation, increased costs of labor, and higher priced technologies. (AHIP, 2006)

Increased utilization was the most important factor in the 8.8% increase contributing 3.8 percentage points of the increase. The major factors that drive utilization are increased consumer demand (1.2%), new treatments (1.0%), and more intensive diagnostic testing (0.8%), the aging population (0.5%), and lifestyle changes (0.3%). (AHIP, 2006)

The Medicare program was implemented to meet a critical need in American society, and over its history, it has evolved into an integral part of the U.S. health care system. In the early stages, Medicare reimbursed hospitals based on “cost plus” principles which was defined as whatever hospitals spent on patient care, Medicare would pay, plus additional dollars, so this payment process developed into a cost-center based hospitals. The cost-center based culture in health care created a fertile environment for inefficiency and high expenditures that signaled a need for change in the payment methodology to reduce cost increases. The system that was developed was a Prospective Payment System (PPS) that used diagnostic-related groups (DRGs) for determining the reimbursement hospitals received for medical care. Under the PPS DRG payment system, hospitals had to bring their costs in line with the stated payments for a discharge diagnosis irrespective of the actual costs of the care provided. These payments were adjusted using the CPI (consumer Price Index) yet the financial outcomes of the hospitals and other

providers receiving these payments had mixed results (Block, 1997).

Historical experiences have found that healthcare (the management of the resources used to correct or improve health) in the United States is one of the most complex and difficult enterprises to understand. In the mid-1990s, healthcare was quickly changing due to dynamic improvements in health technology and management of care. These improvements are likely to continue for some time into the future yet, are expected to continue to increase costs which have ballooned healthcare into a $1 trillion industry, accounting for nearly 15% of the world's largest economy. U.S. statistics for benchmarks, such as infant mortality and longevity, consistently lag behind those of many other industrialized nations and some 40-million U.S. residents (about one in seven) lack sufficient health insurance. Despite this, the chance for significant healthcare reform appears to have come and gone after the defeat of President Bill Clinton's healthcare plan in late 1994 (CMS, 2000).

According to federal forecasters, by 2015 one in every five U.S. dollars is expected to be spent on health care, for total annual health-care spending of more than $4 trillion. The growth in national health spending will average 7.2% between 2005 and 2015, or 2.1 percentage points faster than GDP growth. Hospital spending growth is projected to reach $1.2 trillion in 2015, or double what it was in 2004. Similarly, spending on prescription drugs is expected to reach $446 billion in 2015, up from $188 billion in 2004. Over the decade, average annual spending growth for prescription drugs is projected to be 8.2%. (Borger, Smith, Truffer, Keehan, Sisko, Poisal et al., 2006) These projections come from the National Health Statistics Group of the Centers for Medicare and Medicaid Services and highlight the need to find analytical methods that can help to slow down the cost increases in the U.S. healthcare system.

The history of increasing costs for healthcare in the nation demonstrates that the trend is expected to continue in response to a number of outside factors. These factors can be internal or

external to the economy, such as: climate caused natural catastrophes such as floods, droughts, famines, and typhoons. Tensions and war between countries can also affect the cost of health care. For example the devolution of the former Soviet countries, lower costs of many weapons (especially conventional small arms), growing antagonistic relations along ethnic and nationalist lines, and the growing scarcity and depletion of natural resources, point to the likelihood of increased chaos and war. For healthcare, these situations can result in increases in trauma, malnutrition (as the chaos disrupts food supplies), infectious diseases, and stress-induced illnesses, as well as a diversion of resources away from healthcare toward arms and reconstruction. Also, increasing power and size of global corporations, less stable global finances, increasing influence of donor nations and central finance agencies such as the World Bank and the International Monetary Fund, along with the central government banks and finance ministries, may result in even greater constraints on resources for healthcare in many parts of the world (Heffler, Smith, Won, Clemens, Keehan, Zezzan, 2002).

Internal trends, such as increased industrial pollution, unemployment, stress, lifestyle, aging, and other factors that can affect people's health contributing to increases in healthcare costs. The aging population alone is expected to stretch all healthcare resources thinner.

Finally, certain purely medical changes endanger healthcare around the country. For example, in the ongoing war between pathogens and antibiotics, overused antibiotics appear to be losing their effectiveness against the rapidly evolving pathogens. Due to rapid increases in the number of people taking advantage of relatively cheap international travel allows new epidemics to infect upon the population. For example, the spread of HIV has gone essentially unchecked in much of the world. As of 1994, HIV was infecting 13 million adults a year. The World Health Organization (WHO) predicted that 5 million children worldwide would become infected with HIV between 1995 and 2000 (Heffler et al., 2002). Despite decreases in the rate of infection in

certain countries, the overall number of people living with HIV has continued to increase in all regions of the world, except the Caribbean. There were an additional five million new infections in 2005. The number of people living with HIV globally has reached its highest level with an estimated 40.3 million people in 2005, up from an estimated 37.5 million in 2003. (WHO, 2005)

All these changes tend to thrust national health systems into crisis and chaos. The effect is most marked at the ends of the economic spectrum in the bloated U.S. healthcare industry. Other factors that can influence the cost of healthcare include:

1. *Technical advances*: In the third millennium, new technical developments are expected with life-changing powers having similar effects that the inventions of antibiotics, antisepsis, painkillers, and X-rays brought during 1900 and 2000. The areas that show the most promise for improving people's health in long and short term are:

- Distance surgery, even though in the near future might not affect the health of most people.
- The Human Genome Project, which may isolate the genetic roots of many human diseases - including many that are not generally considered genetic
- The use of genetic markers to screen mass populations, and prevent (through diet, gene substitution, or other special therapies) the specific diseases that individuals are likely to develop
- Nanotechnology, the just-born craft of building molecular-scale machines that holds the promise of completely new types of drugs: tiny machines with the tools and intelligence to perform specific tasks, kill certain viruses, repair certain cells, and manufacture certain needed proteins or enzymes.
- New modes of pharmaceutical research that go far beyond the old blind trial-and-error techniques to actually building molecules (or evolving the bacteria) that can carry out specific tasks, lock onto specific receptor sites in the body, or defeat specific pathogens (Kahn, 2003).

Kahn indicated that the most cost savings and improved quality in healthcare are expected to come not from technical advances, but in the more effective and efficient use of the available techniques already available.

2. *Changes in the direction of healthcare*: Based on collected data, available databases, and information technologies, groundwork for fundamental changes in the organization of healthcare have begun. New technologies now allow doctors to communicate far more easily and quickly and provide healthcare managers the ability to drive their systems in real time. At the same time, they push consumer awareness about health to an entirely new level through the use of interactive cable systems, online forums, and personal health information systems in a wide variety of formats. This technology can facilitate outcomes management and provide expert systems. Outcomes measurement uses massive databases scanning millions of cases to determine which therapies actually work best in particular circumstances. As a management tool, outcomes management makes the practice of medicine more of a science, and less of a craft, resulting in lowered cost and higher quality medical services. Broadly applied, the gates to a number of highly effective and inexpensive non-medical methods that are considered "alternative" or "complementary" in the United States can be opened. The ability to measure all interventions by outcomes and costs can push all types of treatment modalities toward greater unity, bringing a wider range of therapies into official payment systems, and allowing accurate comparisons of intervention and prevention strategies. Outcomes management is spreading rapidly in the United States.

Medical knowledge is expanding faster than at any time in history. Computer programs called “expert systems” assist physicians and other health practitioners move much more rapidly and effectively through the decisions of diagnosis and therapy, isolating rare diseases, differentiating between similar syndromes, and discovering the latest research on the most effective therapies. Their widespread use is expected to substantially change the role of a doctor from knowing facts and to understanding the more human elements of the craft (i.e., making difficult judgments and helping patients change their behavior (Mun & Turner, 1999)).

3. *Change in medical treatment direction*: The future focus of healthcare is moving increasingly out of the acute-care hospitals and back to clinics to doctors' offices, and even into schools, workplaces, and the home. The focus of healthcare is changing from intervening in the acute phase of the disease to early screening, detection, and treatment, and disease prevention. This change is happening because of early diagnosis of disease is more effective and maintaining patients in their homes is far cheaper. In United States, this trend is already at work, meaning that healthcare treatment continues to move out of acute healthcare settings.

In 1982, for every thousand United States citizens, American hospitals logged 1,132 nights in a hospital bed - more than one night per citizen. By 1992, the number on nights had dropped to 607. By 1995, some states were reporting that nights in a hospital bed was as low as 225, with some specific markets (i.e., San Diego County, California) as low as 160. Health futurist, Jeff Goldsmith, estimated that within a decade, most markets are expected to only log 70 to 80 nights in a hospital per year for every 1,000 citizens. In next decade, one third to one half or more of all hospitals face closure. The remaining hospitals are expected to become smaller and care becomes more intensive. This change in the delivery of healthcare services can result in a few large-scale organizations that are bringing together hospitals, doctors, payment structures, and many other services under an oligopoly ownership. Another factor that can expedite this transition is the Congress of the United States, which is planning to enact substantial healthcare reform in the near future.

A wide array of experiments in delivery of health care continues to be noted at the local and state level. Large corporations are becoming more involved in healthcare negotiations, but in an increasingly sophisticated way as they move from bargaining for cheaper rates to working directly with doctors and hospitals to maintain quality and control costs. These corporations are working with people from government, healthcare, education, and other sectors of the society with

this trend resulting in a healthier population who have modified their behavior and cleaned up their living environment (American Hospital Association, ND).

Another method for controlling healthcare cost increases is to identify cases of efficiently provided care. Since a majority of the cost increases in the past decade have been attributed to utilization and poor quality has likewise been linked to over-utilization of services, finding efficient methods for analyzing healthcare data can have the greatest impact on holding down cost increases.

Purpose of the Study

This study tested the efficacy of two statistical analytic procedures, logistic regression versus Mantel-Haenszel differential item functioning (DIF), to predict length of stay (LOS) for pneumonia, (minor, moderate, major, and extreme complication levels) and acute myocardio infraction (minor, moderate, major, and extreme complication levels). The comparison of the two statistical procedures were assessed by model sensitivity to sample size, measure of statistical power and robustness, as well as the ease of understanding the methods and results by nonstatisticians.

Research Questions

The following research question were addressed in this study:

> Which statistical procedure, logistic regression or Mantel-Haenszel DIF, produces results that can be used to predict or explain length of stay for specific DRGs (pneumonia – minor, moderate, major, and extreme and myocardio infraction – minor, moderate, major, and extreme)?

Significance of the Study

The careful analysis of treatment patterns can provide clinical conclusions that result in consistently high quality outcomes while controlling the costs of labor, capital and supplies. This study can assist healthcare institutions by providing a statistically sound method to analyze their data and support the education of health care institutions as they strive to improve quality, control

costs and obtain repeatable positive outcomes. Society also can benefit from the use of better clinical analytical analysis through decreased allocation of scarce resources for medical care while increasing the capacity of healthcare system (providers, institutions and individuals) and permitting society to provide affordable care to more people. Finally, as new and often expensive medical technologies become available, analysis of the clinical outcomes can help allocate resources effectively and therefore allow for more efficient allocation of funds to meet urgent social needs like education, job creation and societal wellbeing for all areas of the world.

Assumptions

This research is based on following assumptions:

- Cost and quality of the same DRG can be the same regardless of the hospitals and providers.
- Similar protocols should produce similar outcomes.
- Within and between provider groups variations can be identified and assigned to specific causes.

Limitations

- The study is limited to two DRGs and four subcategories for each DRG. Results may not be generalizable to other DRGs.
- The study is limited to database information that is publicly available. The results may not reflect the most current data available.

Definition of Terms

Diagnostic Related Groups (DRG)	DRGs are used intentionally to categorize inpatient hospitalization into classes of cases that are similar cost and clinical meanings (Patients with similar illness). Since being implemented as the basis for Medicare prospective payment in 1983, DRGs have been modified and expanded in number of ways by both the Healthcare Financing Administration (HCFA) and other public and private health

	insurance groups.
Length of stay	The total number of days that a patient is receiving care as an inpatient in a hospital setting.
Discharge status	The disposition of the patient upon leaving the hospital: to home, a nursing home, or death.

Chapter 2

Review of Related Literature

Statistical Procedures

Discussion on Dichotomized Variables

Since the inception of the Medicare and Medicaid programs in mid 1960s, hospitals have been reimbursed for Medicare patients on a cost plus principle. This practice created a cost center based healthcare that nurtured inefficiency in resource utilization and cost containment. Due to persistent healthcare cost increases the U.S. Department of Health and Human Services established the Health Care Finance Administration (HCFA) to manage Medicare cost. In 2001, HFCA has been renamed to the Centers for Medicare and Medicaid Services (CMS).

In 1980, in the attempt to create a more cost competitive environment, HCFA changed the cost-plus reimbursement to a Prospective Payment System (PPS) based on standard payments on Diagnostic Related Groups (DRGs). The implementation of PPS has shown mix results. Some hospitals and providers have been making money by keeping the inpatient length of stay below the allowed days so that the standard Medicare reimbursement is greater than the cost of care for that specific patient. If a hospital consistently performs better than allowable LOS, the hospital will create a healthy operating margin that will enable the healthcare system to invest in the most recent and sophisticated technologies. This will attract the best practitioners and enable the hospital to capture a bigger share of the market.

This study examines the efficacy and appropriateness of Logistic Regression compared to Mantel-Haenszel (M-H) differential item functioning (DIF) procedure in predicting the LOS for two DRGs. As hospitals are primarily paid using a prospective payment system, the hospital receives the same payment regardless of the number of days that the patient is treated in the hospital. For the purpose of this analysis, the dependent variable, LOS, is analyzed compared to

the expected length of stay identified by CMS. Payments to the hospital are based on expected costs and expected LOS. Therefore, hospitals that treat a patient past LOS thresholds are not paid more for services and hospitals that can successfully treat a patient within the allowed LOS will generally make money on the episode of care. In comparing these two statistical procedures, the DRG LOS data needs to be dichotomized. Reasons are for using dichotomized LOS for cases in this study include:

a. Hospitals are predominately reimbursed for predetermined LOS for each DRG;
b. Higher LOS can increase the cost of healthcare, but has no effect on reimbursement;
c. Higher LOS can decrease a hospital's capacity to treat more patients;
d. Higher LOS can create a back log for Emergency, Surgery and other ancillary departments that can result in decreased hospital revenues,
e. Higher LOS can decrease available beds to be assigned to nursing units.

Therefore, the dichotomization of LOS is appropriate in the attempt to identify the most appropriate procedure to be utilized in comparing two statistical procedures (LR and M-H) to predict LOS.

Logistic Regression

In ordinary least squares regression analysis (OLS), an outcome or dependent variable Y is modeled as a linear function of a set of predictors $[X_1]$, $[X_2]$,...,$[X_K]$ using the equation $Y = ([\beta_0 X_0] + [\beta_1 X_1] + [\beta_2 X_2] + \ldots + [\beta_k X_k])$. In logistic regression, when the dependent variable is dichotomous and coded (0, 1), the model is given by $\text{logit}(Y) = \ln[\Pr(Y = 1)/\Pr(Y = 0)] = ([\beta_0 X_0] + [\beta_1 X_1] + [\beta_2 X_2] + \ldots + [\beta_k X_k])$. In contrast to OLS, where a single, commonly accepted definition has been used for the standardized coefficient, no widely accepted definition for a standardized coefficient has been accepted for logistic regression.

Reasons for the use of standardized coefficients in logistic regression are essentially the

same as those that motivate the use of standardized coefficients in ordinary least squares regression (OLS; Mayer & Younger, 1976). The first reason is that, for variables with no natural metric, a "scale-free" standardized coefficient may be more meaningful than an unstandardized coefficient. The unstandardized coefficient indicates the impact of a "one-unit" change in the independent variable on the dependent variable, but unless the "unit" itself is meaningful, a "one-unit" change has little or no meaning. For variables measured in a natural metric (centimeters, grams, numbers of people, etc.), a unit has some intrinsic meaning, but many scales used in the social and behavioral sciences (e.g., 1 = strongly disagree, 2 = disagree, 3 = neutral, 4 = agree, 5 = strongly agree) are ambiguous. Even if the reader can assume an interval scale, it can be unclear if a one-unit change, or a 0.1-unit change, represents a large or small change with respect to the scale. The use of standardized coefficients in OLS transforms the independent variable into a variable measured in "standard deviation units."

What is Logistic Regression

The mathematical concept that underlies logistic regression is the logit – the natural logarithm of an odds ratio. The very simple example of a logit is derived from a 2 × 2 contingency table. For example, the distribution of a dichotomous outcome variable (a child from an inner city school who is recommended for remedial reading classes) is paired with a dichotomous predictor variable (gender). A test of independence using chi-square could be applied. The results yield χ^2 (1) = 3.43. Alternatively, one might prefer to assess a boy's odds of being recommended for remedial reading instruction relative to a girl's odds. The result is an odds ratio of 2.33, suggesting that boys are 2.33 times more likely than girls to be recommended for remedial reading classes. The odds ratio is derived from two odds (73/23 for boys and 15/11 for girls) based on previous experience with recommendations for remedial reading. The natural logarithm (i.e., ln [2.33]) is a logit that is equal to 0.85. The value of 0.85 would be the regression coefficient of the gender

predictor if logistic regression were used to model the two outcomes of a remedial recommendation related to gender.

Who, When, and Why the Logistic Regression has been Created and Used

Usually, logistic regression is suitable for testing hypotheses and explaining outcomes about relationships among a categorical dependent variable and one or more categorical or continuous predictor variables. In the case of simple linear regression using one continuous predictor X (a child's reading score on a standardized test) and one dichotomous outcome variable Y (the child being recommended for remedial reading classes), the plot of such data results in two parallel lines, each corresponding to a value of the dichotomous outcome. Because the two parallel lines are difficult to interpret with an ordinary least squares regression equation due to the dichotomy of outcomes, categories may be created for the predictor variable and compute the mean of the outcome variable for the respective categories. The resultant plot of categories' means appears linear in the middle, to that expected on an ordinary scatter plot, but curved at the ends. Such a shape, referred to as sigmoidal or S-shaped curve, is difficult to describe using a linear equation for two reasons:

1. The extremes do not follow a linear trend.
2. The errors are neither normally distributed nor constant across the entire range of data (Peng, Manz, & Keck, 2001).

Logistic regression solves these problems by applying the logit transformation to the dependent variable. In essence, the logistic model is used to predict the logit of Y from X. As stated earlier, the logit is the natural logarithm (ln) of odds of Y, and odds are ratios of probabilities (π) of Y happening (i.e., a student is recommended for remedial reading instruction) to probabilities ($1 - \pi$) of Y not happening (i.e., a student is not recommended for remedial reading instruction). Although logistic regression can accommodate polytomous categorical variables, the focus of this

discussion is on dichotomous dependent variables only. The illustration presented in this article can be extended easily to polytomous variables with ordered (i.e., ordinal-scaled) or unordered (i.e., nominal-scaled) outcomes.

During the last decade, logistic regression has increased in popularity. The trend is evident in articles published in the *Journal of Educational Research*, as well as other higher education journals. This increase can be attributed to researchers' access to sophisticated statistical software that performs comprehensive analyses using this technique. The application of the logistic regression technique is expected to continue to increase as researchers, editors, and readers begin to understand what to expect in articles that use the logistic regression technique. For example tables, charts, or figures that should be included and assumptions that should be verified, as well as the comprehensive presentation of logistic regression results.

Use of Logistic Regression in Healthcare Data Analyses

In social science research, dependent variables (e.g., test scores) generally are continuous. However, many circumstances exist where the behavior being study is not continuous, but results in a dichotomous outcome (e.g., pass/fail, accept/reject, drop out/persist, etc.). Cases involving complex human behaviors can be difficult to measure on interval or ratio scales, although the presence or absence of a characteristic (i.e., a dichotomy) may be observed more easily. Consequently, the use of logistic regression has become the method of choice for studying many issues with dichotomous outcomes in the social sciences.

Several researchers have used logistic regression to study issues involving families. Namerow, Kalmuss, and Cushman (1993) studied factors that influenced adolescent girls when they faced difficult decisions involving the resolution of a pregnancy. In this study, the dichotomous dependent variable was the decision to choose adoption or to parent the baby. Donnelly and Finkelhor (1993) used logistic regression to identify class differences in the

determination of child custody decisions. In their study, custody (joint or sole) was the dependent variable. The authors found that parents with higher income levels, higher educational levels, those living in larger cities, and non-White parents were more likely to have joint custody.

Logistic regression also has applications to medical and other health sciences. Pokorny (1993) developed a model to predict the likelihood of psychiatric hospital inpatients committing suicide. The issue of predicting whether patients admitted to the hospital from a home setting were more likely be discharged to a nursing care facility or returned to their homes was studied by Kane and Matthias (1984). Their model helped identify high-risk background characteristics associated with placement in long-term care.

Industrial and organizational psychologists found logistic regression useful for studying turnover in social service agencies (Balfour & Neff, 1993). Land, McCall, and Parker (1994) used logistic regression as part of a program evaluation designed to assess the extent to which a new intervention functioned as planned for delinquent youth.

Logistic regression has been used to examine practical educational problems. For example, Jones and Mandeville (1990) studied risk factors associated with reading failure. Stage (1988) and Terenzini and Pascarella (1977) applied logistic regression to investigate factors associated with students' persistence or attrition from higher education. Ott (1988) used logistic regression to study factors that could predict academic dismissal in college settings.

Many educational research problems require the analysis and prediction of separate dichotomous outcomes (e.g., if a student can succeed in college, if a child should be classified as learning disabled [LD], if a teenager is prone to engage in risky behaviors, etc.). Traditionally, these research questions were addressed by using either ordinary least squares (OLS) regression or linear discriminant function analysis. Both techniques were subsequently found to be less than ideal for handling dichotomous outcomes due to their strict statistical assumptions (i.e., linearity,

normality, and continuity for OLS regression and multivariate normality with equal variances and covariances for discriminant analysis; Cabrera, 1994; Cleary & Angel, 1984; Cox & Snell, 1989; Efron, 1975; Lei & Koehly, 2000; Press & Wilson, 1978; Tabachnick & Fidell, 2001). Logistic regression was proposed as an alternative to OLS regression or discrimination function analyses in the late 1960s and early 1970s (Cabrera, 1994).

The use of logistic regression has increased in both social science (e.g., Chuang, 1997; Janik & Kravitz, 1994; Tolman & Weisz, 1995) and educational research, particularly in higher education (Austin, Yaffee, & Hinkle, 1992; Cabrera, 1994; Peng & So, 2002a; Peng, So, Stage, & St. John, 2002). With the availability of sophisticated statistical software for high-speed computers, the use of logistic regression is increasing.

An example of the use of logistic regression in health care involve patient outcomes who had a diagnosis of ventilator-associated pneumonia. Variables that were present in >10% of patients with ventilator-associated pneumonia and were significantly associated to ventilator-associated pneumonia outcomes in a univariate analysis were included in the study. These variables had been found to have a priori clinical significance and gave consideration to biological plausibility. In addition, variables that had been identified by previous investigators as risk factors for ventilator-associated pneumonia in adults (e.g., total parenteral nutrition (TPN) and steroids) were entered into multiple forward stepwise logistic regression models. Statistically significant variables that were thought to co-vary were grouped, and one variable from each group selected for entry into the model. Patient data were entered into an Access database. Data analyses were performed using SPSS Version 10 (SPSS, Chicago, IL). Categorical variables were compared using χ^2 analysis. Continuous variables were compared using the Wilcoxon rank sum test for nonnormally distributed variables, with a two-tailed $p \leq .05$ considered statistically significant. The final model was selected on the basis of biological plausibility and by selecting the logistic

regression model with the highest Hosmer and Lemeshow test of significance and the lowest -2 log likelihood function. First-order interactions among significant variables in the final model were tested, with no evidence of first-order interactions found. After the variable "transfusion" was found to be a statistically significant predictor of ventilator-associated pneumonia, new variables including transfusion before ventilator-associated pneumonia, packed red blood cells before ventilator-associated pneumonia, fresh-frozen plasma before ventilator-associated pneumonia, platelets before ventilator-associated pneumonia, and cryoprecipitate before ventilator-associated pneumonia were created and entered into the model in place of the variable "transfusion." For each type of blood product, amounts and types received after the development of ventilator-associated pneumonia were censored; thus, the variables entered into the multivariate model reflected only amounts and types of products received before infection. Both categorical and continuous variables for each type of blood product were tested in the multivariate model.

Differential Item Functioning

The origins of examining performance differences in achievement generally have been based on what were considered to be bias issues. Over the last decade, however, a considerable shift has occurred in what "biased" means. Because of connotations in the mid-1980s associated with the term, bias, a more neutral term, differential item functioning (DIF), was proposed. In contrast to differential item performance (DIP), DIF refers to items that do not function in a similar manner for comparable members of different groups. More recently, other researchers presented the concept of differential bundle functioning (DBF), a collection of DIF items with a common dimension (e.g., content) that collectively produce a bundle of items that are differentially easier for one matched group of test takers when compared to another. Methods that detect only DIF could miss an important phenomenon: DIF amplification. DIF amplification is defined as the study of a set of DIF items collectively that favors one group in comparison to

another at the test score level. However, these DIF items could show minimal or no DIF when tested as individual items.

Mantel-Haenszel Definition and History

In performing an analysis using differential item functioning (DIF) analysis, comparisons are made between a single reference group and multiple focal groups. Conducting separate tests of DIF for each focal group has several undesirable qualities: (a) the Type I error rate may exceed the intended nominal level if the alpha level for each individual test is not appropriately adjusted, (b) power may be lower than a single test that assesses DIF among all groups simultaneously, and (c) substantial time and computing resources are required. These drawbacks potentially can be avoided by using a procedure that has the capacity to assess DIF across multiple groups simultaneously. In the present study, the performance of three methods of assessing DIF across multiple demographic groups; the Mantel-Haenszel chi-square statistic with no adjustment to the alpha level, the Mantel-Haenszel chi-square statistic with a Bonferroni adjusted alpha level, and the Generalized Mantel-Haenszel statistic (GMH) that offers a single test of significance across all groups will be used. Simulations were conducted using a single reference group and 1, 2, 3, and 4 focal groups, having from 1 to all of the focal groups in a given condition experiencing DIF. Additional conditions that were varied include: group size, focal group ability distribution, and magnitude of matching criterion contamination. The results suggest that GMH is an appropriate procedure because the Type I error rate remained at the nominal level of 0.05, and its power was consistently among the higher that other types of analyses.

Concern for equality in testing during the 1960s and 1970s led to a surge in development of statistical methodology for item bias detection (Camilli & Shepard, 1994; Cole, 1993). Incipient statistical investigations into item bias were predicated on the identification of items that displayed unusually large differences in mean performance between demographic groups relative

to that observed for other test items (Angoff, 1972; Cleary & Hilton, 1968). Although these methods enabled identification of items that may have been differentially difficult for one group of examinees, they lacked the rigor needed to determine if differences in group performance on an item resulted from some form of unfairness in the item or differing levels of proficiency in groups being compared. Modern investigations into item bias control for the confounding effects of differing levels of group proficiency have been accomplished through the use of the framework of differential item functioning (DIF) that exists when examinees from different demographic groups perform differently on an item after an intervention focused on improving the ability intended to be measured by the test (Dorans & Holland, 1993). The presence of DIF may indicate the existence of a systematic invalidity of the item, placing one group at a disadvantage.

Over the past two decades, numerous DIF detection procedures have been developed for both dichotomous and polytomous items (Camilli & Shepard, 1994; Clauser & Mazor, 1998; Millsap & Everson, 1993; Penfield & Lam, 2000; Potenza & Dorans, 1995). These approaches were developed exclusively for the two-group case in which comparisons are made between a base (reference) group and a second (focal) group. However, assessing item bias for several focal groups is becoming increasingly desirable. Numerous focal groups have been identified as candidates for DIF investigation (e.g., racial groups, gender groups, and examinees with disabilities; Zieky, 1993). Linn (1993) suggested further refinement of focal group categories to distinguish among specific racial/ethnic groups (e.g., Puerto Ricans, Mexican Americans, Cubans, and other Hispanic groups). The practical need for considering multiple focal groups is highlighted by the presence of numerous studies in the literature examining DIF among multiple ethnic groups (Schmitt, 1988; Schmitt & Dorans, 1990; Zwick & Erickan, 1989) and multiple languages of administration (Angoff & Sharon, 1974; Ellis & Kimmel, 1992).

Given the prevalence of multiple-group DIF assessments, investigations into item bias

could benefit from the availability of statistical procedures that test for DIF simultaneously across multiple groups. Such procedures have three possible advantages over the traditional two-group methods: (a) the ability to detect DIF across multiple groups simultaneously may be greater than that observed in individual pairwise tests, (b) the inflated Type I error rate that may be expected when DIF is tested between multiple pairs of groups is avoided with a single procedure that tests for DIF across all groups simultaneously, and (c) a single test of DIF across all groups provides a more efficient method of assessing DIF than testing each group individually.

Common Usage of Differential Item Functioning

One of the most popular procedures for assessing DIF in dichotomous items is the Mantel-Haenszel (MH) procedure, This procedure was first developed for use in epidemiological research (Mantel & Haenszel, 1959), and was later applied to the detection of DIF by Holland and Thayer (1988). Applying the MH procedure to DIF detection begins by grouping examinees according to an estimate of ability (generally the total test score), and then forming a two-by-two contingency table crossing group membership (reference and focal) and item performance (correct and incorrect) for each level of ability. In survival analysis, a linear model often provides an adequate approximation after a suitable transformation of the survival times and possibility of the covariates. This article proposes a semiparametric regression method to estimate the regression parameter in the linear model without specifying the distribution of the random error, where the response variable is subject to case 1 interval censoring. The method uses a constructed random-sieve likelihood and constraints that combines the benefits of semiparametric likelihood with estimating equations. The estimation procedure is implemented, and the asymptotic distributions are obtained for the estimated regression parameter and the profile likelihood ratio statistic. In addition, some model diagnostics aspects are described. Finally, the small-sample operating characteristics of the proposed method are examined via simulations, with its usefulness

illustrated on datasets from an animal tumorigenicity study as well as an HIV study. (Mantel & Haenszel, 1959).

Assessing DIF across multiple groups using the MH chi-square reduces to performing individual tests for each pair of groups to be compared, leading to the problem of an increased probability of committing a Type I error. Suppose that in the course of an analysis the researcher wishes to conduct several tests for DIF, each having an associated probability of a Type I error. Because the probability of committing a Type I error over repeated significance tests is greater than the probability on any one significance test, the probability of a Type I error over all tests exceeds the intended nominal alpha level. Using the terminology of Keppel (1991), the probability of committing a Type I error for a given comparison is referred to here as the error rate per comparison, and is distinguished from the probability of at least one Type I error across all tests of significance, referred to as the familywise error rate.

Two possible alternatives to the MH chi-square procedure to assess DIF across multiple groups are proposed here. The first solution is to adjust the per comparison alpha level according to the Bonferroni inequality that asserts that the probability of making a Type I error anywhere in one of the k tests of significance is less than or equal to the sum of the Type I error rates of each test (Mendenhall, Scheaffer, & Wackerly, 1986). That is, where, in the context of DIF detection across multiple focal groups, *i* refers to the test of DIF of the *i*th group of a total of *j* focal groups.

The MH chi-square statistic can be performed for all *j* focal groups in relation to a single reference group, and the familywise error rate is guaranteed not to exceed the intended nominal Type I error rate per comparison when certain assumptions hold (i.e., equal reference and focal group ability distributions). To distinguish the MH procedure conducted with and without an adjusted alpha value, the MH test performed with the Bonferroni-adjusted alpha level is denoted by BMH, whereas the MH test performed without the adjustment is denoted simply by MH.

Although the use of a Bonferroni-adjusted alpha level can solve the problem of a spiraling Type I error rate, multiple tests of DIF are still required. However, a natural extension of MH, the Generalized Mantel-Haenszel (GMH; Somes, 1986) can be used to test for DIF across all groups simultaneously.

For example, where diagnosis (*nk*) is a (J-1)-by-(J-1) diagonal matrix with elements *nk*. Note that A_k and $E(A_k)$ are vectors of length J-1, corresponding to any J-1 of the J demographic groups. The GMH statistic is distributed as a chi-square variable with J-1 degrees of freedom under the null hypothesis of no DIF. Note that the GMH procedure has already been described as a method of DIF detection by Zwick, Donoghue, and Grima (1993), whereby it was used to assess DIF between two demographic groups for polytomous items containing J possible nominal response categories. This previous application of the GMH procedure is similar to that described here, with one exception: in the use made by Zwick et al. (1993), the group variable is dichotomous and the response variable is polytomous, whereas in the presented application the group variable is polytomous and the response variable is dichotomous.

Mantel-Haenszel as an Index of the DIF

The Mantel-Haenszel Procedure (Camilli & Shepard, 1994; Holland & Thayer, 1986; Mantel & Haenszel, 1959) is a commonly cited index of differential item functioning (DIF). It is an asymptotic chi-square statistic with 1 degree of freedom that is computed from the set of 2 x 2 x S contingency tables formed by contrasting a demographic group (e.g., male and female) with items correct and incorrect (i.e., 0 and 1) for each possible score (i.e., 0 to N for a test of N items) of the test. The statistics can be expressed as:

$$1. \qquad \Sigma A_J = \frac{nR\ NF_j\ m1_j\ m0_j}{T2_j\ (T_{j-1})}\ ; \text{ and}$$

$$2. \qquad E(A_j) = \frac{nR_j m1_j}{T_j}.$$

The squared difference of expected and actual counts of participants seen in Equation 1 also include a constant of 1/2 that is subtracted from the usual difference of actual and expected values. That constant provides a correction for continuity to increase the accuracy of this asymptotic statistic (Holland & Thayer, 1986).

Why Use Mantel Haenszel DIF Methodology versus Other Types

The question of appropriate sample size has been extensively investigated in the past several years, with particular attention placed on very small samples. Swaminathan and Rogers (1990) compared the Mantel-Haenszel procedure and logistic regression as indicators of DIF, using sample sizes of 250 and 500 in each comparison group. Penny and Bond (1992, 1995) examined the performance of the Mantel-Haenszel procedure with 5,000 participants per comparison group and encountered magnitudes of the chi-square in excess of four digits. In a follow-up study (Penny, 1996) with 500 participants per group, the chi-square, under identical levels of DIF, never exceeded three digits. Mazor, Kanjee, and Clauser (1995) compared logistic regression with the Mantel-Haenszel procedure with samples of 1,000 and 908 participants per group. Clauser, Mazor, and Hambleton (1994) examined the Mantel-Haenszel procedure with group sizes of 100, 200, 500, 1,000, and 2,000 participants. Ryan (1991) used unequal reference- and focal-group sizes ranging from a minimum of 141 observations to a maximum of 4,345 observations in a study of the Mantel-Haenszel chi-square statistic and common odds ratio. Camilli and Smith (1990) compared the performance of the Mantel-Haenszel procedure with a jackknife procedure, using 1,085 White participants and 300 Black participants. Parshall and Miller (1995) examined the Mantel-Haenszel procedure in a series of simulation studies in which

the reference group was comprised of 500 participants and the focal group size ranged from 25 to 200 participants. Those studies represent a small portion of the DIF studies that have been presented and published in the past decade and demonstrate the wide range of sample sizes that have been used. Moreover, these studies demonstrate the congruity of the several studied indices of DIF and conditions (e.g., the case of nonuniform DIF and the Mantel-Haenszel statistic) under which that congruity may be lost.

Along with the question of the appropriate sample size needed to detect DIF is the decision regarding what type and magnitude of differential functioning constitutes a practical level of aberrancy. In general, as with most other statistical studies, if a sufficient number of participants are included in the research, then DIF can be found and the problem of making a Type I error, or of identifying an item as biased when in reality it is unbiased is not of major concern to most researchers. The question that confronts researchers who assess items for deviance is the extent to which differential functioning by an item is a practical and valid method. Most items function in a slightly different manner between any two groups of people, just as one thermometer may yield slightly dissimilar temperatures for two identical fires. The concern is that researchers recognize that a trivial difference using a large sample can be of little importance regardless of the statistical significance.

Use of DIF

Researchers in the field of educational testing have developed sophisticated data analytic methods for studying measurement equivalence. In this field, student scores on specific constructs (e.g., verbal achievement, mathematical achievement, and reading comprehension) are routinely compared, with these comparisons used by college/university admissions officers to make decisions about individual applicants. In part because of the importance given to standardized test scores for individuals' life chances and because of claims of cultural bias in

these types of tests, educational test developers, working within the framework of item response theory, have developed methods to examine the equivalence of measurement instruments across diverse sociocultural groups. These methods use an explicit measurement model to represent the relations between observed behaviors (i.e., the item responses) and the construct to be quantified.

A commonly used statistical test for DIF, called Lord's (1980) chi-square, takes the above definition of operational, item, and scalar equivalence as its null hypothesis (i.e., H0: $a_{j1} = a_{j2}$, $b_{j1} = b_{j2}$) and tests it against an unspecified alternative, assuming that the estimates of a_j and b_j have a bivariate normal sampling distribution within each population. If the null hypothesis is retained for all items, then the instrument's item and scalar measurement equivalence are supported for the included groups. However, like other likelihood ratio tests, this test is highly sensitive to sample size. For large samples, the null hypothesis of no DIF is likely to be rejected even when differences among groups in the item parameter values are so small that they have no practical significance.

Researchers have proposed several alternative indices of DIF on the basis of the area between group-specific item characteristic curves (ICC). These indices are conceptually equivalent to -square, as equality of item parameters suggests that the same ICC is true for the groups being compared, in which case the area between the ICCs should be zero. The more the item parameters differ among groups, the area between the group-specific ICCs will be larger. Indices of DIF based on this reasoning are consistent with the formal definitions of item, operational, and scalar measurement equivalence and are considered to be are less sensitive to sample size than Lord's chi-square.

The role of test statistics or other indices of DIF in studies of measurement equivalence needs to be considered. In general, hypothesis tests are designed to limit Type I errors when investigators want to demonstrate the existence of some effect. In DIF analyses, however,

investigators often hope to affirm the measurement equivalence of an instrument (i.e., to demonstrate the absence of an effect of sociocultural group membership on the properties of that instrument) and leads to the retention of the null hypothesis. Researchers need to have a way to decide if items appear to function equivalently across groups. Test statistics and other indices of DIF can provide some relevant information. Comparing plots of the group-specific ICCs can also be informative. The item content should be taken considered as well in making decisions on DIF results. Investigators need to combine test statistics and DIF indices with other types of information when deciding to omit or retain a given item in an instrument.

DIF analytic methods have some advantages over traditional psychometric approaches. Indices of internal consistency (e.g., Cronbach's coefficient alpha and the Kuder-Richardson formula) essentially represent the average item slope parameter for an instrument. Yet the items in an instrument could easily have the same average slope in two groups, although on an item-by-item basis the slopes can be expected to differ. This outcome occurs if one subset of behaviors is related more strongly to the target construct for one group, while another subset is related more strongly for another group. Moreover, even if the slopes of all items were identical across groups, the threshold parameters could differ without necessarily having any impact on traditional indices of internal consistency. In contrast, DIF analytic methods could be expected to detect differences of these types and reveal sources of possible measurement discrepancies that could be masked by using traditional psychometric methods.

The DRG System and the HCFA

History/Background

The Diagnosis Related Group (DRG) system was created as a means of classifying medical conditions and the resulting hospital cases into an estimated 500 clinically specific groups. These groupings were based on criteria that integrated elements such as hospital resource

use (McMahon, Fetter and Thompson, 1986). The DRG system was developed with the intention of providing a basis for the projection of possible resource requirements and the system was later integrated into the Medicare Prospective Payment System (PPS; McMahon, Fetter & Thompson, 1986).

The DRG system is based on criteria established by Robert Fetter at Yale University in the early 1980s and backed by the Health Care Financing Administration (HCFA), now know as the Centers for Medicare and Medicaid Services (CMS). Due to the complex nature of projecting costs for health care and Medicare payment systems, the need for the DRG system was driven by the need for more effective cost projections. In essence, the application of DRGs provides a fixed fee for medical care and procedures based on the specific groupings, regardless of whether the expended costs for each individual case fluctuate from this standard (Long Island Business News, 2002).

The diagnosis related groups (DRGs) are defined by specific conditions, as well as by specific criteria that may have an effect on the health of individual patients when treating the same illness in the presence of various mitigating factors (DRG. [The ABCs of Health Care], 2002). Therefore, while diagnostic groups are related to medical diagnosis, they are further classified into categories according to medical factors that could influence patient outcomes. These determining factors include: the treatments used, the age of the patient, the gender of the patient, and patient status at discharge (DRG; The ABCs of Health Care, 2002).

Recent advancements in computer technology have led to the creation of programs that place patients in one of approximately 500 DRGs and assigns them to divisions based on diagnose codes (*The International Statistical Classification of Diseases* [ICD-9 Codes]), their age, gender, and the presence of any possible complicating factors or comorbidities (DRG. [The ABCs of Health Care], 2002). This type of classification has made it easier to implement changes in the

DRG system, as well as include new information (i.e., application of additional classification rules; Schwartz et al, 1996). As a result, the Centers for Medicare and Medicaid Services publish their DRG Definitions Manual on a yearly basis, integrating any changes into an electronic format that is readily accessible to health care administrators.

Similarly, Hoffman, Klees and Curtis (2003) publish a biannual overview of the Medicare and Medicaid programs in which they include a description of the current reimbursement system, generally referred to as the Prospective Payment System (PPS). The PPS replaced the practice of the Medicare and Medicaid systems paying what was considered the reasonable and customary cost basis of reimbursement to hospitals in 1983. Other providers like nursing homes, rehab facilities, emergency rooms, and physicians in the health care sector have different payment systems. For example, short nursing home visits covered under Medicare are reimbursed under a separate Prospective Payment System. Payments for inpatient rehabilitation, psychiatric, and home health care are currently reimbursed on a reasonable cost basis, but prospective payment systems are expected to be applied to other areas of health care in the near future (Hoffman, Klees, & Curtis, 2003; p. S1).

When a health care provider (hospital, physician, nurse practitioner or other provider) accepts assignment, that provider is agreeing to accept the Medicare rate as payment in full for coded procedures and diagnoses. The provider may not charge either the patient or supplementary insurer any additional sum. “If the provider does not take assignment, the beneficiary will be charged for the excess (which may be paid by Medigap insurance)” (Hoffman et al., 2003; p. S1).

Under the national Medicare system and by many commercial and state Medicaid health care systems, payments are made according to a specific predetermined cost based on each patient’s diagnosis related group (DRG) classification (Hoffman et al., 2003; p. S1). Placing the patient in the appropriate (i.e., most advantageous) DRG is a provider strategy that increases the

amount of reimbursement they can receive. Hospitals and other health care providers focus on ensuring that each patient's status is updated regularly to capture any changes in the diagnostic and procedure codes assigned to a patient, particularly when the patient's conditions change, to place them into a DRG that increases the payments for the provider. This process is often referred to as "up-coding" and is acceptable providing that the codes and diagnoses accurately reflect the patient's condition and the procedures and tests that the patient underwent.

Under the Medicare and Medicaid payment systems, as well as most private-sector managed care plans, patients are free to select their own health care providers. For payment under PPS, however, they must select a provider who has agreed to accept assignment for the current year. The payments that Medicare makes to providers in Medicare+Choice plans "are based on a blend of local and national capitated rates" (Hoffman et al., 2003; p. S1) and then adjusted for local economic conditions based on the demographics of the local area.

Many fewer Medicare managed care plans (Medicare+Choice) are existence than at the beginning of the 21st century. In essence, many of these plans chose not to accept assignment of Medicare reimbursement rates after 2000. The Centers for Medicare and Medicaid Services (CMS) are in the process of phasing in new "risk adjusters based on demographics and health status … to better match Medicare capitation payments to the expected costs of individual beneficiaries" (Hoffman et al., 2003; p. S1). Though CMS strives to hold Medicare costs in check, the difficulty that some patients have in finding providers willing to accept Medicare and Medicaid has forced a comprehensive review of CMS' payment systems.

The development of the DRG system resulted from a focus on cost-containment measures and the belief that both hospitals and the government could benefit from improved methods at determining cost-projections for health care conditions. Cost containment is the greatest single focus of Medicare activity in the current environment. Health care analysts voice widespread

concerns about Medicare's ability to survive as the populous baby boomer generation approaches the retirement age. Current Medicare payments are funded by collections in the form of payroll taxes from the working population. As baby boomers begin to retire and as businesses continue to downsize their workforces, the federal government can expect reduced revenues from payroll taxes. Direct capital infusion to Medicare in the form of taxes already has been reduced and is expected to be diminished still further as time passes and the American workforce ages.

As if the above situation were not enough to threaten Medicare's ability to meet the increasing health care needs of the elderly, the overall cost of health care continues to increase at rates that are making attempts at cost containment substantially more difficult. Managed care was quite effective in cost containment until the mid-1990s, when health care costs began to rise again despite managed care's best attempts to keep the cost of healthcare manageable.

Section 5101 of Public Law 105-277 stands as an example of legislative influences designed to limit cost increases. Section 5101 addresses home health agencies directly and the manner in which Medicare pays for services after October 1, 1998. The law limited acceptable charges to 105% of an agency's median charges in 1998 (Legislative Update, 1998).

This limit is indicative of the manner in which legislation is used to contain costs. It does not seek to limit what an agency can charge for services, but it does directly limit that portion for which Medicare will pay. Cost increases greater than 5% can be covered either by personal payment or payment by a Medigap or private insurance policy. DRGs, then, provide the basis for cost projections and prospective payment determinations relative to treatment for certain conditions. Rather than paying out the exact amount expended for care for each patient, Medicaid and Medicare pay out an average cost based on the condition and the DRG determination, with the belief that balance can be made over time because some cases are expected to be more costly and some cases may be less costly than the designated average cost.

DRG Data and Quality of Patient Outcomes

Care providers face challenges to operate within constraints imposed by Medicare's accepted fee levels, but care providers have little choice except to operate within Medicare's constraints because of Medicare's influence on private insurers. The scene described by Finkelstein, Silvers, Marrero, Neuhauser and Cuttler (1998) is familiar:

> There is concern in both the medical community and the general public about mechanisms of medical decision making and the interplay of physician and insurer decisions in determining access to care. (p. 663)

Medicare's purpose in setting limits on fees for services is to hold back cost increases, but CMS also realizes that cost increases will occur as a normal course of business. Therefore, Medicare automatically allows a small percentage increase each year, indirectly guaranteeing that providers' prices will increase as they seek to maximize their own revenues. A provider with its best interests in mind collects as much as it legally can from third-party payers, and as illustrated, Medicare sets the standard on what private insurers can be expected to pay for specific goods and services.

Researchers (Cantone, 1999; Finkelstein et al.,1998; Hoffman, Klees, & Curtis, 2003) have related challenges of applying DRG designations within the scope of the Medical PPS and the problems with designated cost-containment efforts in providing adequate care. Research has suggested that health care providers often are challenged with issues of ethics relative to the use of DRGs and the billing of Medicaid/Medicare, resulting in claims of medical fraud (Cantone, 1999). As a result, researchers have studied the impact of the use of DRG data and its effects on the quality of care provided by health care organizations and government funding agencies relying on DRGs.

Vaul (1998) recognized that one of the problems with the use of data collected relative to the development of DRGs was the lack of national coding norms relative to DRGs that could

support the systematic designation and billing functions relative to DRG-related data. Claims have been made that the quality of care and the function of funding systems have declined significantly as a result of this lack of coding norms (Vaul, 1998).

In terms of patient care and outcomes, DRG data has also influenced the operation and function of hospitals and determined shifts in approaches to care. Rehnquist, in a 2001 memorandum to Thomas Scully, then Administrator of the CMS, stated that the DRG-driven Medicare P.P.S. has demonstrated many reporting inaccuracies. These data inaccuracies and problems with coding issues have led to the need to create greater consistency relative to DRG designations and the development of strategies for PPS operations.

Vaul (1998) also reported that substantial issues exist relative to diagnostic groups and what has been described as "upcoding," or the repetitive coding of multiple conditions in atypical cases. Atypical cases relating to DRG pairs can cost the federal government large sums of money, as research has shown.

> The financial implications of upcoding are significant. Atypical cases of just these six DRG pairs account for more than $272 million of potential Federal government overpayments for 1996 alone, with each DRG pair offering potential recoupment of at least $5 million. These atypical cases also represent more than $500 million of potential fines, assuming that the government would levy the maximum fine of $10,000 per case. (Vaul, 1998).

In this study, the potential fines comprised approximately 70% of the total quantified impact of the DRGs that were assessed, strongly suggesting a need for greater coding consistency.

Vaul's (1998) findings, described in the "National MEDPAR DRG Benchmarking Study," suggested that inconsistencies exist in coding practices commonly used by health care

> institutions and that reducing the inconsistencies is needed when compared with the national norms. Vaul (1998) concluded that:
>
> To avoid coding problems, healthcare financial managers should ensure that their physicians are properly documenting patient records, that coding practices comply with government regulations, and that their medical record staff is well trained and

> has access to the tools necessary to ensure coding accuracy. For hospitals with a high proportion of atypical cases, careful medical record documentation can help justify higher-paying DRG coding (p. 58).

Preyra (2004) presented results that supported Vaul's outcomes.

Researchers also have argued that the quality of care and the issues of ethics relative to Medicare/Medicaid P.P.S. and billing have plagued the system based on the DRG data (Horn et al, 1985; Hsia et al, 1988; Iezzoni, 1997). Barker (2002) studied the issue relative to the use of DRGs and the need for greater consistency in the hospital setting. Because of the need for effective payment projections, the use of the DRG system requires compliance with directives and attention to the details outlined in the guide published by the CMS (Averill et al, 1992).

How Economics Shaped the DRG System

The field of economics is based on the relationship between supply and demand. The classical interpretation is that the more demand that exists for a good or service, the rarer it becomes, buyers attach a higher value, and as a result, prices increase for the good or service. Conversely, if there is an over supply of a particular good or service, the demand declines and the price or the good or service drops. While this relationship is widely accepted and successfully applied to many situations, the standard tenants of supply, demand, scarcity and other basic economic principles have not always followed these constructs in the health care industry.

While in its earliest days, economics was defined in terms of exchange and production, more recently definitions of economics tend to revolve around the concepts of supply and demand, choice and scarcity (Schenk, 2004). In 1935, economist Lionell Robbins postulated that economics is, at its heart, a social science that determines how people choose to use scarce resources in an effort to satisfy unlimited wants (Schenk, 2004). Scarcity is defined as when people demand more than what is available. Many factors can add to scarcity; for example, time and income (Schenk, 2004). Furthermore, resources such as manpower, machinery, and natural

resources are limited, which also provides a fixed maximum on the exact amount of goods and services that are available. Therefore, scarcity, along both of the above lines, requires that people need to choose which desires they are able to satisfy and which desires will be left unsatisfied.

This discussion can be directly applied to health care. One assumption is that good health is directly dependent on the strength of access to medical care. It could also be assumed that "medical care utilization increases when patients have greater access to medical care" (Robst & Graham, 2004, p. 467). These assumptions could make sense upon reading without careful analysis. For example, living near a clinic or a physician's office, or having a choice of many physicians in one area could lead to stronger preventative care, as well as stronger continuous care, especially in the event of chronic illnesses (Robst & Graham, 2004). Yet results of studies that tested these assumptions were mixed -- with some researchers claiming that medical care access may not necessarily improve health, and others (especially in rural areas) noting that there are improved health outcomes in rural areas when more physicians are available to residents (Robst & Graham, 2004). For the most part, the number of physicians per population is not necessarily a good indicator of quality health care as individuals are known to cross city and county lines to receive treatment (Robst & Graham, 2004).

However, like standard economic theory, health care economics discusses the rationing of resources and scarcity, especially when for areas such as managed care (Weinstein, 2003). In the United States' health care system, it is impossible for every individual to get every single medical service that would provide some good (Weinstein, 2003). The cost would be astronomical and it could easily be questioned whether all of the care given was required. An example of diminishing returns for more units of a service. As a matter of practical application, health care services are rationed out. Access to treatment is often hampered by availability of doctors and nurses, insurance rules, the level of insurance or ability to pay, and the willingness of providers to accept

patients. As an example, in 2003, 15% of Americans went without health care insurance (i.e., were uninsured most or all of the year; Weinstein, 2003).

These changes in the economic events in the health care industry are somewhat new and can be traced to the introduction of different forms of health care insurance. The advent of employer and government sponsored health insurance began to change the economic drivers in terms of demand, cost for services and access to care.

In the 19th and the first half of the 20th centuries, doctors and hospitals operated in a relatively unencumbered market where demand and supply directly affected the price of services for those with the means to pay and charity care was often provided by religious or county programs. (Prior to the 19th century, medical knowledge was minimal and training more of an apprenticeship than a field of study.) The early employer based health insurance programs that started in the 1920s covered a percentage of catastrophic care like hospitalizations. Noncatastrophic care (e.g., doctor's visits, nursing home care, outpatient testing) was paid directly by the consumer.

As more companies began to offer health insurance to their employees and as competition for employees increased, many employers began to offer "richer" benefit packages. These plans covered not only a percentage of catastrophic care but also a percentage of general care and sometimes preventive care. Most of these plans would not be considered "managed care" since the patient and doctor determined what care would be administered. The apparatus that placed limits on the amount spent on health care was the willingness of the patient to incur costs. Still an increase in demand for services, either induced by the provider, demanded by the patient, or due to improvements in medical knowledge and treatment, began. Since health care insurance was still a relatively inexpensive benefit to provide to employees, the richness of the plans continued to grow, less financial participation was required by the patient and providers (doctors and hospitals)

were able to increase their prices. The birth of the Medicare program offering health care coverage for those over 65, and the ability of unions to bargain for better health care coverage for their workers, increased the demand for services. By the 1980s the U.S. began to spend more for health care on a per capita basis and as a percent of GDP than all other industrialized nations.

With increases in demand and cost of services rising every year, the Federal Government, which now paid for over 33% of the U.S. health care costs through the Medicare and Medicaid programs and through coverage for federal employees and military programs, began to look for ways to cap their expenditures for health care coverage. From these pressures, "managed care" programs were born. The first programs to effectively contain costs arguably would be Health Maintenance Organizations or HMOs. In an exchange for participating in preventive care, utilizing a smaller or closed panel of selected providers, and a "gate-keeping" function that required that a primary care doctor determine the level of care that their patient would receive access to, the employer would receive a reduced premium for health insurance. It is important to note that the quality of care discussion which permeates discussions on health care and health care costs today began in earnest with the advent of HMOs.

In a backlash against real or perceived rationing of care and lack of quality under HMOs, Preferred Provider Organizations (PPOs) requiring deductibles and set copayments, but fewer restrictions on access to care and the providers that could be used began to grow in popularity. Often comparably priced to HMOs, PPOs reduced costs by requiring discounts from normal fees the doctors and hospitals. PPOs were also less regulated than HMOs where HMOs were often required to offer a prescribed set of services. Employers could determine certain parameters of services that would be paid for and payment mechanisms for the plans they offered their employees. One unexpected miscalculation with PPOs was the increased demand for services under these programs. Again the providers could have added to this increased or induced demand.

With health care expenditures continuing to escalate, cost containment programs were started in an attempt to reduce health care costs. With hospitalization being the largest percentage of costs, capitation programs aimed at hospital expenses were born. The most significant of these was Medicare's Prospective Payment System (PPS) which used diagnostic related groups or DRGs to determine hospital payments. This limited the reimbursement paid to a hospital for patient hospitalization. The DRG was assigned to the patient based on the condition presented to the hospital. As the Medicare program is federally funded, hospitals had to accept this payment. Soon after HMO and PPO plans followed suit.

Does this mean that the healthcare of those individuals on managed care or some other insurance plan are rationed? First it is important to remember that HMOs and PPOs are often lumped together under the category of "managed care." Many managed care programs limit or choose not to cover certain medical services, such as preventative screening tests (Weinstein, 2003). Health care in a managed care system is further rationed through application of co-payments and deductibles. In both the Medicare fee for service and managed care plans there are limits placed on acceptable tests and treatment given the diagnosis. The federal government determines what care is reasonable and what care is not. In commercial health plans, the employer or union typically has control over the level of deductibles, co-payments, and covered tests and procedures. State Medicaid plans ration care based on a number of factors and methods to insure that the services rendered match the funding provided for care in that calendar year. In each of these scenarios, patients are not likely to seek treatment they do not deem as necessary if the price – in terms of the deductible, co-pay, or actual cost when there are no cost constraints - is considered too high, even though it could make the patient healthier in the long run.

What is the situation in countries where health care is paid for by the government? Technically, one could assume that everyone has access to all the health care they could want.

Unfortunately, this is not the case. In these single payor, government-run systems, this scenario the rationing involves time spent. Depending on the seriousness of procedures, people might choose to forgo a treatment rather than have to wait a long time for something that is not critical.

But if health care is rationed, then how does it not fit into the general concepts of economics? Is not health care economics basically a decision of how to allocate health care decisions and resources? The main problem with placing health care into a general economic concept arises when it comes to pricing health care. Again, if pure economic theory is considered, price increases as demand increases. Because of insurance, as in the United States, and government subsidies, such as those in Canada, prices are rigidly controlled. To a certain extent, people choose health care based on price (i.e., whether a procedure is covered by their plan or not). But price alone does not drive demand for health care.

In recent years, the focus on health care economics and its payment systems, including Medicaid and the Medicare Prospective Payment System, and the integration of payment classification systems in health care billing, have played a major role in reshaping views of health care economics. The creation, integration, and continued use of DRGs have extended from the desire to create greater consistency in medical financing in the midst of cost-containment efforts. Meanwhile, critics of the DRG-based system in the United States have argued that the quality of care has declined while the cost of care has increased, in part due to increases in fraud in medical billing and with “up-coding” the process where a provider (either hospital or physician) overstates the medical condition of the patient.

Health Care Financing

Hospitals have always been under continuous pressure to contain costs and improve quality, but this pressure approached new heights in the early 1990s, and continues to grow today. Hospitals have adapted to this pressure in various ways, including the development of multi-unit

organizations (Kovner, 2002). According to experts, the two major forms of multi-unit organization exist, including:

1. The "alliance," which is when separately-owned organizations enter into a relationship primarily for strategic purposes, such as referring patients to each other for services
2. The other form is multi-corporate organizations, which are merged organizations with unified ownership. Multi-corporate organizations often boast of having advantages such as increased volume, cost savings, and greater leverage in bargaining with managed care organizations.

However, Kovner (2002) points out that these advantages have yet to be proven. Those hospitals that have remained independent argue that advocates of merged organizations have not achieved the success they expected. In each case, the form that the hospitals pursue is governed by the same needs to contain costs, while adequately funding services.

Between 1985 and 1999, around 700 hospitals in the US changed their ownership status, in many cases changing from a non-profit to a for-profit orientation (Shen, 2003). These conversions, in general, increased the hospital's profit margin, which was achieved through various means, but was associated by Shen (2003) with both staff reductions and the closing of trauma centers. Newly converted for-profit hospitals generally implemented reductions in the registered nurse-to-bed ratio, with similar reduction in other staff with nursing degrees (Shen, 2003). Proponents of conversion from non-profit to for-profit status argue that this change can facilitate health coverage of the uninsured, generating more tax revenue for the government, and frees resources for acute care that can be devoted to public health services and enhance a hospital's capability to access capital (Shen, 2003). According to Shen (2003), after conversion, these new for-profit hospitals were able to increase their efficiency and their profit margins.

Opponents to this change assert that converted hospitals increase prices, lower quality and will be less likely to accept unprofitable patients, as well as fail to provide adequate community benefits because of their focus on the bottom line (Shen, 2003). Inpatient discharges and outpatient visits were reduced by between 6 to 12 percent (Shen, 2003). Some have felt that this consolidation off providers has given hospitals additional bargaining power and the ability to obtain higher reimbursement.

In recent years, there has been a rapid increase in physician-owned specialty hospitals and ambulatory surgery centers (Casalino, Devers and Brewster, 2003). It has long been a maxim in capitalism that specialization pays. However, the traditional template for hospitals has been to offer a broad range of health care services. Some analysts feel that this model for hospitals has become an anachronism and that the introduction of market forces into health care will naturally lead to the creation of "focused factories" that will out-distance the traditional hospital (Casalino, Devers and Brewster, 2003, p. 56). This competitive edge comes from dedicating staff, equipment and management focus to the treatment of one specific type of disease, in regards to both inpatient (specialty hospitals) outpatient care (ambulatory surgery centers) (Casalino, Devers and Brewster, 2003). It is possible that such a focused approach may be better able to provide quality care at a lower cost, which would result in high patient satisfaction (Casalino, Devers and Brewster, 2003).

This development is causing some concern within the hospital industry, as well as with state legislatures and Congress. The concern is focused on three principal questions: (1) do these competing facilities provide the cost and quality benefits claimed (2) do they have a negative financial impact on general hospitals and (3) do they increase or decrease access to care (Casalino, Devers and Brewster, 2003). As these questions demonstrate, while there is concern for health issues, a major concern is financially oriented. Policymakers are concerned that these facilities

threaten the financial stability of hospitals, which are an integral part of the health care delivery system.

Gosden, Pedersen and Torgensen (1999) point out that how individuals are paid has a direct effect on how they work. It is this observation that has led to the belief that the manner in which doctors are rewarded for their services has a direct impact on a health care system as a whole. A principal method by which doctors are paid in the United States is defined as "fee-for-service." The term "fee for service" refers to the practice of doctors setting a price for a specific service and then the patient either paying that price personally or through an insurer (Thorpe and Knickman, 2002). It has been assumed by many insurers in recent years that allowing doctors to charge patients directly for services encourages them to provide more, and possibly unnecessary, services (Gosden, Pedersen and Torgersen, 1999).

An alternative to direct fee-for-service payment is the capitation system. Capitation refers to the practice of paying the doctor, or group practice, according to a fixed fee per month for each HMO member who lists that physician/practice as their primary care physician (Smits, 2002). This method of payment has been praised by policymakers as a mechanism for keeping down health costs, as it rewards a physician for keeping patients healthy (Smits, 2002). On the other hand, while the capitation system does not lead to an oversupply of services, it may not be as effective as other system if patients are unable to assess the quality of care that they need (Gosden, Pedersen and Torgensen, 1999). A third system of rewarding doctor is through salaries. However, little is known concerning whether this system of payment leads to more efficient and equitable systems than either fee payment systems or capitation systems (Gosden, Pedersen and Torgensen, 1999).

To address this question Gosden, Pedersen and Torgensen (1999) conducted a review of empirical research on the topic. The review found that available evidence was limited and not

consistent in quality, which made it difficult for the researchers to draw clear conclusions concerning the effectiveness of the different methods for paying doctors. While some studies examined the link between payment type and quality of services, few examined the impact that payment type had on the health of patients, and no studies addressed whether or not salaried doctors treated their patients differently (Gosden, Pedersen and Torgensen, 1999).

Under the DRG system, hospitals are paid a pre-established amount per case treated, with the payment rates varying according to the case (Thorpe and Knickman, 2002). DRGs measure hospital performance by, first of all, classifying patients into one of 23 major diagnostic categories (MDCs). The MDCs are then divided further into 47 diagnostic groups based on various factors. This concept maintains that the best measure of a hospital's output is the diagnosis, rather than individual services provided or the length of the hospital stay. Also DRG payments are determined prospectively and are fixed (Thorpe and Knickman, 2002). The use of DRGs has been a significant factor--but not the only factor--affecting the decline in in-patient hospital utilization that has become a trend since the mid-1980s. However, this reduction in hospital use has also had side effects, such as sharp increases in post-hospital use of services, which includes home health care and nursing home care (Thorpe and Knickman, 2002).

Thorpe (2002) lists the increased use of DRGs as the "most important change in hospital rate setting during the 1980s" (p. 440). Medicare's experience with the DRG system has produced mixed results. At the onset, hospitals responded quickly to the altered incentives created by the DRG payment system, as the system provided an incentive to reduce costs as any savings stemming from reductions could be retained by the hospital (Thorpe, 2002). Falling lengths of stay, admissions and employment levels result in slower hospital expenditure growth. However, with cost growth slower than increases in revenue, hospital operating margins increased. Also, as the growth of Medicare inpatient hospital spending slowed, outpatient

hospital spending increased (Thorpe, 2002). Medicare provides payment for outpatient services through a variety of means. Over the past decade, outpatient charges have risen at a higher rate than Medicare payments, which has placed an increasing amount of cost sharing on Medicare beneficiaries (Thorpe, 2002). For instance, during the federal fiscal year of 1995, Medicare payments for hospital outpatient services totaled $19.4 billion (Thorpe, 2002). Thirty-seven percent of this was financed by beneficiary overpayment (Thorpe, 2002). The sharp increase in outpatient payments has caused some experts to propose a Prospective Payment Program for outpatient care also. One proposal states that the Secretary of Health and Human Services should be responsible for establishing such a program. The design of such a program would have to embrace the difficult problem of defining the clinically and financial relevant DRG equivalents (Thorpe, 2002).

The principal motivation of both insurers and government agencies in using capitation systems of payment has been to control costs (Catalano, et al, 2000). The argument is that fee-for-service encourages physicians to increase their financial benefits by dispensing expensive treatments. According to Newhouse (2002), the method that a particular insurer uses to contract with physicians depends largely upon the size of the plan, as well as the size of the physician group with which it is contracting. Plans that predominantly used fee-for-service (FFS) as a basis for contracting with physicians tended to be smaller. Likewise, practices contracting using FFS were more often solo practices, as large group practices tend to go with capitation plans (Newhouse, 2002). Larger physician groups are better able to bear the risk that occurs from random variation. A given enrollment, the “larger the plan’s provider network, the smaller will be its share of any given provider’s business and thus the less leverage it will have with that provider” (Newhouse, 2002, p. 207).

Robinson and Casalino (2002) pointed out that the relationship between insurers and

physicians can differ markedly according to geographical area. These researchers specifically looked at the states of New York and California that historically have stood at opposite ends of the managed care spectrum. Traditionally, New York physicians retain the model of solo practice and fee-for-service reimbursements while their Californian counterparts form medical groups that receive payment through capitation. In both cases, Robinson and Casalino (2002) found a departure from the experience of the 1990s, including a reduction in the scope of prepayment and a "rethinking" of the contractual relationship of capitation.

Despite the widespread perception that physician networks are broad, in 1998, the majority of enrollees were not in plans that gave them access to virtually any primary care provider (Newhouse, 2002). Only 35% of enrollees in HMOs had access to 75% or more of the principal care physicians in their market area.

Disease management protocols are designed to improve quality of care, but Newhouse (2002) found that their use varied according to the clinical problem presented. The term "disease management" refers to "efforts by health plan and/or specialized subcontractors to establish a prospective approach to managing a complex disease with the aim of reducing unnecessary use of costly services" (Smits, 2002, p. 300). Use of this term can be confusing as doctors have been managing disease since the inception of their profession. Within the realm of managed care, disease management refers specifically to:

1. integration of all components of care using the data available to the managed care plan;
2. emphasis on prevention and education; and
3. use of guidelines endorsed by expert professionals to guide the process of care (Smits, 2002, p,. 300).

Disease management, in this context, began as a means for handling such health problems as refractory asthma, diabetes and heart disease. It has been shown to decrease emergency room

use, improve quality of life, and reduce long-term complications (Smits, 2002). The trend toward capitated payments is expected to create strong financial incentives for managed care plans to solve some noneconomic and quasi-economic obstacles to care, such as the prevention of hospital admissions for diseases that can be managed through timely and effective outpatient management (Billings & Cantor, 2002).

Chapter 3

Methodology

The methods that were used to address the research questions posed by this study are presented in this chapter. The topics that were discussed include a restatement of the problem, the research design, variables in the study, data collection, and data analysis.

Restatement of the Problem

This study tested the efficacy of two statistical analytic procedures, logistic regression versus Mantel-Haenszel differential item functioning (DIF), to predict length of stay (LOS) for pneumonia, (minor, moderate, major, and extreme complication levels) and acute myocardio infraction (minor, moderate, major, and extreme complication levels). The comparison of the two statistical procedures was assessed by model sensitivity to sample size, measure of statistical power and robustness, as well as the ease of understanding the methods and results by nonstatisticians.

Research Design

A retrospective, correlational research design was used in this study. The use of two different statistical procedures was used to determine if DRG LOS outcomes could be predicted from selected patient variables. This type of research design is appropriate when doing a secondary analysis of previously collected data to determine specific outcomes or make comparisons between variables that had not been explored previously. The major problem with the use of previously collected data is the lack of control the researcher has over data collection.

Variables in the Study

The variables that were included in the study are located in publicly available data bases from Premier Group Purchasing, Inc. The data base did not provide identifying information about the patients (patient names and social security numbers), but did include the following information

that was used in this study:

- Patient demographic data (race, age, marital status)
- Hospital data (Admission type, Length of stay, DRG code, APR DRG [risk adjusted], and variable costs).

Data Collection Procedures

The data managers from Western Maryland Health Systems were contacted to obtain permission to use their data bases for the four selected DRGs. Personal patient information was eliminated from the data base to protect the identity of patients. The remainder of the information was obtained on compact disk.

Data Analysis

The data were transferred to a data file that was compatible with SPSS – Windows, ver. 15.0. The data analysis was divided into two sections. The first section provided a description of the variables to provide baseline data for the reader. The second section used logistic regression and Mantel-Haenszel statistical procedures to address the research questions developed for this study. All decisions on the statistical significance were made using an alpha level of .05. Figure 1 presents results of the statistical analyses.

Merits of Logistic Regression Vs. Mantel-Haenszel Statistic

The Mantel-Haenszel procedure is a commonly cited index of differential item functioning (DIF) (Camilli & Shepard, 1994; Holland & Thayer, 1986; Mantel & Haenszel, 1959). It is an asymptotic chi-square statistic with 1 degree of freedom that is computed from the set of 2 x 2 x S contingency tables formed by contrasting a demographic group (e.g., male and female) with items correct and incorrect (i.e., 0 and 1) for each possible score (i.e., 0 to N for a test of N items) of the test. The statistics can be expressed as

$$(1)\ \Sigma A_J = \frac{nR\ NF_j\ m1_j\ m0_j}{T2_j\ (T_{j-1})}$$

and

$$(2)\ E(Aj) = \frac{nRjm1j}{Tj}$$

The squared difference of expected and actual counts of participants seen in Equation 1 also includes a constant of 1/2 that is subtracted from the usual difference of actual and expected values. That constant provides a correction for continuity that increases the accuracy of this asymptotic statistic (Holland & Thayer, 1986).

An important feature of the Mantel-Haenszel statistic is that participants are equated before they are compared, in much the same manner as is found in analysis of covariance. That equating is performed at the level of test scores that can run from, say, 0 to 100 for a 100-item test in which the score is the number correct. The denominator of the first equation is a variance, given in Equation 2, computed from the marginal totals for the 2 x 2 contingency table at the j^{th} test score level. That is, for a given level, j, of test score, nR_j is the number of participants in the reference group; nF_j is the number of participants in the focal group; m_{1j} and m_{0j} are the number of correct and incorrect responses, respectively, by the participants who scored at the j^{th} level; and T_j is the total number of participants.

The null hypothesis of no DIF is frequently written in terms of a ratio that is known as an odds ratio. If p and q represent the probabilities that members of the focal and reference groups will answer the j^{th} item correctly or not, then the null hypothesis that is tested by the Mantel-Haenszel procedure may be expressed as

$$(3)\ H_0: \quad \frac{pR_j}{qR_j} = \frac{pF_j}{qF_j}$$

and the alternative hypothesis can be written as

$$(4)\ H_1: \quad \frac{pRj}{qRj} = \acute{\alpha}\ \frac{pF_j}{qFj}$$

Alpha ($\acute{\alpha}$) is the common odds ratio. One can compute alpha by averaging the S odds ratios, one for each level of ability. By inspection, alpha = 1 represents the null hypothesis. Because the odds ratio is an average of the individual odds ratios for each score, one can achieve an average, or common, value by having some odds ratios that are above 1 and others that are below 1 in such numbers and magnitudes that they average to 1. In such an event, the common odds ratio would support the null hypothesis when in fact the alternative is true, although in opposite directions for different score levels with one direction cancelling the other. For that reason, an underlying assumption of the Mantel-Haenszel test is that the odds ratio is constant across the levels of test score. Moreover, when the condition of a constant odds ratio is met, the Mantel-Haenszel test provides a powerful unbiased test of the null hypothesis against the alternative (Holland & Thayer, 1986).

If the condition of an invariant odds ratio is not met, then certain characteristics of the underlying item response theory model that describe the item can be inferred (Fischer, 1993, 1995). For example, when the odds ratio varies across the levels of ability that are represented by the test score, the IRT item discrimination index differs between the two demographic groups that are being compared. This type of DIF is known as nonuniform because the direction, as well as the magnitude, of the differential functioning is not uniform, or consistent, across the levels of ability over which the comparison is being made. Moreover, the Mantel-Haenszel procedure may not be robust in regard to the violation of the assumption of invariant odds ratios (Penny, 1996; Zwick, 1990), resulting in situations in which the procedure may produce excessive rates of Type I and Type II errors.

Appropriate Sample Size

The question of appropriate sample size attends studies of differential item functioning and has been extensively investigated in the past several years, with particular attention placed on very small samples. Swaminathan and Rogers (1990) compared the Mantel-Haenszel procedure and logistic regression as indicators of DIF, using sample sizes of 250 and 500 in each comparison group. Penny and Bond (1992, 1995) examined the performance of the Mantel-Haenszel procedure with 5,000 participants per comparison group and encountered magnitudes of the chi-square in excess of four digits. In a follow-up study, Penny (1996) using a sample of 500 participants per group, the chi-square, under identical levels of DIF, never exceeded three digits. Mazor, Kanjee, and Clauser (1995) compared logistic regression with the Mantel-Haenszel procedure with samples of 1,000 and 908 participants per group. Clauser, Mazor, and Hambleton (1994) examined the Mantel-Haenszel procedure with group sizes of 100, 200, 500, 1,000, and 2,000 participants. Ryan (1991) used unequal reference- and focal-group sizes ranging from a minimum of 141 to a maximum of 4,345 in a study of the Mantel-Haenszel chi-square statistic and common odds ratio. Camilli and Smith (1990) compared the performance of the Mantel-Haenszel procedure with a jackknife procedure, using 1,085 White participants and 300 Black participants. Parshall and Miller (1995) examined the Mantel-Haenszel procedure in a series of simulation studies in which the reference group was composed of 500 participants and the focal group size ranged from 25 to 200 participants. Those studies represent a small portion of the DIF studies presented and published in the past decade and demonstrate the wide range of sample sizes that have been used. Moreover, they, as well as many other studies, have demonstrated the congruity of the several studied indices of DIF and the conditions (i.e., the case of nonuniform DIF and the Mantel-Haenszel statistic) under which congruity could be lost.

Robustness

Connected with the question of the sample size needed to detect DIF is the decision of which type and what magnitude of differential functioning constitutes a practical level of aberrancy. In general, as with most other statistical studies, if enough observations are used, then DIF can be found. However, the problem of making a Type I error or identifying an item as aberrant when in reality it is fair, is a real concern to most researchers. Hence, an important question that faces those who would assess items for aberrancy is just how much differential functioning by an item constitutes a practical difference. It is unlikely that any item functions in exactly the same way between any two groups of people, just as one thermometer is likely to yield slightly different temperatures for two identical fires. Trivial differences need to be recognized as such, although that difference may be statistically significant because of a large sample.

When the question of practical significant differences can be answered, then an appropriate sample size can be estimated. The Mantel-Haenszel chi-square statistic is sensitive to sample size (Cohen, 1988; Penny, 1996), increasing in power as the sample grows. It is also a good indicator of uniform DIF, the kind of DIF that is present when the odds ratio is constant across the levels of the test score, especially when the data fit the IRT one-parameter logistic model (1PL), or Rasch model (Zwick, 1990; i.e., that is, when the item discrimination parameters and pseudo-guessing parameters are 1 and 0, respectively). Hence, with a sufficiently large sample, any magnitude of group difference in the IRT item difficulty parameter, no matter how minuscule, conceivably can be detected by the Mantel-Haenszel chi-square statistic. Where, then, does a difference between item difficulty seen in two groups grow from a magnitude that is considered representative of simple random variability to a magnitude that is considered representative of a systematic difference between the two groups?

The question of practical levels of aberrancy is addressed by associating DIF, as defined by between-groups differences in IRT item difficulty,

$$(5)\ DIF = b_R - b_F,$$

with Cohen's effect size, w (Cohen, 1988), where

$$(6)\ \chi^2 = nw^2,$$

and b_R and b_F are estimates of item difficulty seen in the reference and focal groups. Items that fit two special cases of the general three-parameter IRT model (3PL) are[:

$$(7)\ P(\theta) = c + 1 - \frac{c}{1 + e^{-1.7a\,(\theta - b)}}$$

where “a” represents the item discrimination parameter, “b” represents the item difficulty parameter, “c” represents the item pseudo-guessing parameter, and Theta (θ) represents the ability of the person who is answering the item. P(θ) represents the probability that a person with ability θ correctly answers an item with IRT parameters (a, b, c). The value 1.7 is a scaling constant that renders the logistic curve numerically similar to the normal curve.

The first special case of the 3PL is known as the 1PL or Rasch model and can be written as

$$(8)\ P(\theta) = \frac{1}{1 + e^{-1.7(\theta - b)}}$$

where the difference between the 3PL and 1PL is that c = 0 and a = 1.

The second special case of the 3PL presented in the present study does not have a special name but is achieved by using a = 1 in the 3PL. Such a restricted model is useful when one is describing items in which guessing may be a factor in student responses, but in which the discrimination parameter is approximately 1. That model can be written as

$$(9)\ P(\theta) = c + 1 - \frac{c}{1 + e^{-1.7(\theta - b)}}$$

The central mathematical concept that underlies logistic regression is the logit – the natural logarithm of an odds ratio. The simplest example of a logit is derived from a 2 × 2 contingency table. Consider an instance in which the distribution of a dichotomous outcome variable (a child from an inner city school who is recommended for remedial reading classes) is paired with a dichotomous predictor variable (gender). Example data are included in Table 2.

Table 2

Example of Logistic Regression Contingency Table

Remedial reading instruction	Gender		Total
	Boys	Girls	
Recommended (coded as 1)	73	15	88
Not recommended (coded as 0)	23	11	34
Total	96	26	122

To test for differences between boys and girls, a chi-square test of independence could be applied. The results yield $\chi^2_{(1)} = 3.43$. Alternatively, one might prefer to assess a boy's odds of being recommended for remedial reading instruction relative to a girl's odds. The result is an odds ratio of 2.33, suggesting that boys are 2.33 times more likely to be recommended for remedial reading classes when compared to girls. The odds ratio is derived from two odds (73/23 for boys and 15/11 for girls); its natural logarithm [i.e., ln(2.33)] is a logit that equals 0.85. The value of 0.85 would be the regression coefficient of the gender predictor if logistic regression were used to model the two outcomes of a remedial recommendation for gender.

Logistic Regression

Generally, logistic regression is well suited for describing and testing hypotheses about relationships between a categorical outcome variable and one or more categorical or continuous predictor variables. In the simplest case of linear regression for one continuous predictor X (a

child's reading score on a standardized test) and one dichotomous outcome variable Y (the child being recommended for remedial reading classes), the plot of such data results in two parallel lines, each corresponding to a value of the dichotomous outcome. Because describing the two parallel lines can be difficult with an ordinary least squares regression equation due to the dichotomy of outcomes, categories may be created for the predictor variable (scores on standardized reading test) by computing the mean for each of the respective categories. The resultant plot of means for each of the categories appears to be linear in the middle of the scatterplot, much like what one would expect to see on an ordinary scatter plot, but curved at the ends. Such a shape, often referred to as sigmoidal or S-shaped, is difficult to describe with a linear equation for two reasons. First, the extremes do not follow a linear trend. Second, the errors are neither normally distributed nor constant across the entire range of data (Peng, Manz, & Keck, 2001). Logistic regression solves these problems by applying the logit transformation to the dependent variable. In essence, the logistic model predicts the logit of Y from X. As stated earlier, the logit is the natural logarithm (ln) of odds of Y, and odds are ratios of probabilities (π) of Y occurring (i.e., a student is recommended for remedial reading instruction) to probabilities ($1 - \pi$) of Y not occurring (i.e., a student is not recommended for remedial reading instruction). Although logistic regression can accommodate categorical outcomes that are polytomous, in this study, the focus is on dichotomous outcomes.

Summary

Unlike discriminant function analysis, logistic regression does not assume that predictor variables are distributed as a multivariate normal distribution with equal covariance matrix. Instead, it assumes that the binomial distribution describes the distribution of the errors that equal the actual Y minus the predicted Y. The binomial distribution is also the assumed distribution for the conditional mean of the dichotomous outcome. This assumption implies that

the same probability is maintained across the range of predictor values. The binomial assumption may be tested by the normal z test (Siegel & Castellan, 1988) or may be taken to be robust as long as the sample is random; thus, observations are independent from each other.

Figure 1

Statistical Analysis

Research Question	Variables	Statistical Analysis
1. Which statistical procedure, logistic regression or Mantel-Haenszel DIF, produces results that can be used to predict or explain length of stay for specific DRGs (pneumonia – minor, moderate, major, and extreme and myocardio infraction – minor, moderate, major, and extreme)?	<u>Dependent Variable</u> Length of stay for four DRGs • Pneumonia – minor • Pneumonia – moderate • Pneumonia – major • Pneumonia - extreme • Myocardio infarction – minor • Myocardio infarction – moderate • Myocardio infarction – major • Myocardio infarction - extreme <u>Independent Variables</u> Age Race Marital status Admission type Variable Costs	Logistic regression was used to determine which of the independent variables are predictors of length of stay for the four specific diagnoses. Mantel-Haenszel DIF procedures were used to determine the efficacy of using differential item functioning to predict the length of stay for four DRGS. Length of stay was dichotomized into two groups – those that exceed the allowed length of stay and those that are less than the allowed length of stay.

Chapter IV

Results of Data Analysis

This chapter presents results of the data analysis used to address the research question posed for this study. The chapter is divided into two sections. The first section compares the five included independent variables (patient demographics: age, marital status, and race; admission type; and variable costs). The second section presents the results of the Mantel-Haenszel differential item function (DIF) procedures and binary logistic regression used to determine which of the variables provide the best information regarding the influence of the independent variables on the dependent variable, length of stay.

The data for these analyses were obtained for two diagnostic-related groups (DRGs), pneumonia and myocardio infarction from closed records drawn from two hospitals in an eastern state. All identifying information on the records was removed prior to completing any analysis.

To prepare the dependent variables for use in the statistical analyses, the DRGs for both pneumonia and myocardio infarction were each categorized into four specific DRGs based on the severity of the illness. Using usual and customary length of stay (LOS) for the four DRGs, length of stay was dichotomized into less than or equal to or greater than the usual and customary LOS. The dichotomized LOSs for the four DRGs for each type of diagnosis (pneumonia/myocardio infarction) were then combined to develop the dependent variables that categorized the LOS for each patient as either less than or equal to or greater than the respective LOS for their DRGs.

The independent variables measuring the demographic characteristics (age, marital status, and race) of the patients and hospital variables (admission type and variable costs) were dichotomized for use in both the logistic regression and Mantel-Haenszel DIF procedures. Comparisons of the independent variables were made between the two hospitals to determine if

significant differences would require separate analyses by hospital. Generally, age and variable costs would be compared using t-tests for independent samples to determine the existence of statistically significant differences. However, chi-square tests for independence were used with the dichotomized variables as that scaling was used for both types of analyses.

Age was dichotomized using a median split that was determined using descriptive statistics. The reported marital status of the patients was dichotomized into married/not married. Race was dichotomized into White/minority. Hospital variables (admission type and variable costs) also were dichotomized. The admission types were combined into emergency/urgent care and elective. A median split was used to divide variable costs into high and low.

The dichotomized variables were compared between the two hospitals to determine if statistically significant differences were present that would require the hospitals to be examined separately. The results of the chi-square tests for independence for the pneumonia DRG are presented in Table 3.

Table 3

Crosstabulations
Demographic Characteristics by Hospital – Pneumonia Diagnoses

Demographic Characteristic	Hospital				Total (N = 1,766)	
	A (n = 976)		B (n= 790)			
	n	%	n	%	N	%
Age						
≤ 74	551	56.5	356	45.1	907	51.4
>74	425	43.5	434	54.9	859	48.6
χ^2 (1) = 22.23, p ≤ .001						
Marital status						
Married	617	63.2	431	54.6	1048	59.3
Single	359	36.8	359	45.4	718	40.7
χ^2 (1) = 13.22, p ≤ .001						
Race						
White	956	98.0	777	98.4	1733	98.1
Minority	20	2.0	13	1.6	33	1.9
χ^2 (1) = .20, p = .656						
Admission Type						
Emergency/Urgent Care	951	97.4	771	97.6	1722	97.5
Elective	25	2.6	19	2.4	44	2.5
χ^2 (1) <.01, p = .834						
Variable Costs						
≤ $2,249	467	47.8	416	52.7	883	50.0
> $2,249	509	52.2	374	47.3	883	50.0
χ^2 (1) = 3.85, p = .050						

χ^2 was computed using the continuity correction for 1 degree of freedom

Three statistically significant results were obtained for the chi-square tests for independence. The patients admitted for a pneumonia diagnosis were younger (≤ 74 years of age, n = 551, 56.5%) in Hospital A than those in Hospital B (≤ 74 years of age, n = 356, 45.1%), χ^2 (1) = 22.23, p ≤ .001. The association between marital status and hospital also was statistically significant, χ^2 (1) = 13.22, p ≤ .001. A greater percentage of patients in Hospital A were not married (n = 617, 63.2%) than in Hospital B (n = 431, 54.6%). A statistically significant association was found for variable costs, with Hospital A (n = 467, 47.8%) having fewer cases under the median for variable costs ($2,289) than Hospital B (n = 416, 52.7%), χ^2 (1) = 4.04, p =

.050. The remaining independent variables (race and admission type) were not associated with the hospitals.

A second set of analyses were used with the data for myocardio infarction. The same five variables (age, marital status, race, admission type, and variable costs) were crosstabulated by hospital to determine if an association existed between the hospital and demographic characteristics. The results of these analyses are presented in Table 4.

Table 4

Crosstabulations
Demographic Characteristics by Hospital – Myocardio Infarction Diagnoses

Demographic Characteristic	Hospital				Total (N = 645)	
	A (n = 383)		B (n= 262)			
	n	%	n	%	n	%
Age						
≤ 76	217	56.7	127	48.5	344	53.3
> 76	166	43.3	135	51.5	301	46.7
χ^2 (1) = 3.87, p = .049						
Marital status						
Married	196	51.2	146	55.7	342	53.0
Single	187	48.8	116	44.3	303	47.0
χ^2 (1) = 1.12, p = .291						
Race						
White	378	98.7	260	99.2	638	98.9
Minority	5	1.3	2	0.8	7	1.1
χ^2 (1) = .07, p = .790						
Admission Type						
Emergency/Urgent Care	372	97.1	256	97.7	628	97.4
Elective	11	2.9	6	2.3	17	2.6
χ^2 (1) =.04, p = .839						
Variable Costs						
≤ $3,275	193	50.4	129	49.2	322	49.9
> $3,275	190	49.6	133	50.8	323	50.1
χ^2 (1) = .04, p < .835						

χ^2 was computed using the continuity correction for 1 degree of freedom

One variable, age, produced a statistically significant chi-square value, χ^2 (1) = 3.87, p =

.049. This result indicated that an association existed between hospital and the age of the patients. Patients in hospital A (age $\leq$ 76; n = 217, 56.7%) tended to be younger than patients in hospital B (n = 127, 48.5%). The chi-square analyses for the remaining independent variables were not statistically significant, indicating that the hospital was not associated with marital status, race, admission type, or variable costs.

As a result of obtaining statistically significant outcomes on some of the independent variables for both pneumonia and myocardio infarction diagnoses, the Mantel-Haenszel DIF and logistic regression statistical procedures were conducted separately for each hospital.

Research Question

The research question posed for this study is:

> Which statistical procedure, logistic regression or Mantel-Haenszel DIF, produces results that can be used to predict or explain length of stay for specific DRGs (pneumonia – minor, moderate, major, and extreme and myocardio infraction – minor, moderate, major, and extreme)?

Pneumonia Diagnoses

Hospital A

The first analysis that is presented to address this research question is logistic regression. The dependent variable in this analysis is length of stay for pneumonia diagnoses. The independent variables for this analysis include age, martial status, race, type of admission, and variable costs. Results of the correlation analysis for Hospital A are presented in Table 5.

Table 5

Pearson Product Moment Correlations
Pneumonia – Hospital A (N = 976)

Measure	1	2	3	4	5	6
1. Length of Stay – Pneumonia	--					
2. Age	.20**					
3. Marital Status	.02	-.02				
4. Race	-.02	-.13**	-.05			
5. Admission Type	.02	-.01	.04	-.02		
6. Variable Costs	.56**	.29**	.05	-.05	<.01	

Length of Stay: Less than or equal to usual and customary allowable inpatient days is coded as a 1; Age" Less than or equal to 74 coded as a 1; Marital status: Not married is coded as a 1; Race: White is coded as a 1; Admission type: Emergency is coded as a 1; Variable costs less than or equal to $2,249 coded as a 1.
**p ≤ .01; *p ≤ .05

Two independent variables, age, r = .20, p < .001; and variable costs, r = .56, p < .001, were significantly related to the dependent variable, length of stay. Intercorrelations between age and race, r = -.13, p < .001 and age and variable costs, r = .29, p < .001 also were statistically significant. The remaining correlations were not statistically significant.

These variables were used in a logistic regression analysis, with all independent variables entered simultaneously. Results of this analysis are presented in Table 6.

Table 6

Logistic Regression
Pneumonia Diagnoses – Hospital A (n = 976)

Variable	B	SE	Odds Ratio	Wald Statistic
Age	.01	<.01	1.01	4.94*
Marital Status	.18	.17	1.19	1.07
Race	.19	.62	1.21	.09
Admission Type	-.43	.48	.65	.80
Variable Costs	<.01	<.01	1.00	175.41**

Note: Hosmer and Lemeshow Test = χ^2 (8) = 22.30, p < .004
**p ≤ .01; *p ≤ .05

The results of the logistic regression provided support that the five independent variables in the model were not a good fit for predicting LOS for pneumonia diagnoses, χ^2 (8) = 22.30, p < .004. As determined by the Wald statistic, two independent variables were found to be statistically significant, age, 4.94, p = .026, and variable costs, 175.41, p < .001. In examining the odds ratios, older patients (> 74 years of age), unmarried patients, minority patients were more likely to have longer LOS. Patients who were admitted through emergency rooms were more likely to be in the hospital longer and variable costs were higher for patients who stayed in the hospital longer.

Mantel-Haenszel statistical procedures were used to determine if length of stay (less than or equal to the allowable days and greater than allowable days) was influenced by the five characteristics (age, marital status, race, admission type, and variable costs) of patients admitted to the hospital for pneumonia diagnoses. The results of the crosstabulations for LOS by the five characteristics are presented in Table 7.

Table 7

Mantel-Haenszel Statistical Procedures
Length of Stay (Pneumonia) Hospital A

Patient Characteristic	Length of Stay				Total (N = 976)	
	≤ Criteria for DRG (n = 560)		> Criteria for DRG (n = 416)			
	N	%	N	%	N	%
Age						
≤ 74 years of age	365	65.2	186	44.7	551	56.5
> 74 years of age	195	34.8	230	55.3	425	43.5
Mantel-Haenszel = 39.80, p < .001						
Marital Status						
Unmarried	359	64.1	258	62.0	617	63.2
Married	201	35.9	158	38.0	359	36.8
Mantel-Haenszel = .36, p = .547						
Race						
White	547	97.7	409	98.3	958	98.0
Minority	13	2.3	7	1.7	20	2.0
Mantel-Haenszel = .22, p = .640						
Admission Type						
Emergency/Urgent Care	547	97.7	404	97.1	951	97.4
Elective	13	2.3	12	2.9	25	2.9
Mantel-Haenszel = .12, p = .730						
Variable Costs						
≤ $2,249	403	72.0	64	15.4	467	47.8
>$2,249	157	28.0	352	84.6	509	52.2
Mantel-Haenszel = 303.63, p ≤ .001						

**p ≤ .01; *p ≤ .05

A statistically significant result was obtained on the Mantel-Haenszel analysis for age, M-H statistic = 39.80, p < .001. This result indicated that an association exists between LOS and age. The odds ratio estimate of 2.32 (95% CI = 1.79 to 3.00) indicated that older people with a pneumonia diagnosis were more than twice as likely to be in the hospital longer than the criteria allowable LOS for their specific DRGs than those who were younger.

The second statistically significant result obtained on the Mantel-Haenszel procedures was variable costs, M-H statistic = 303.63, p < .001. The odds ratio estimate of 14.12 (95% CI = 10.21 to 19.52) indicated that patients whose variable costs were greater than the median were more

likely to exceed the criteria allowable LOS for their specific DRGs than those whose variable costs were less than the median.

The remaining Mantel-Haenszel analyses for marital status, race, and admission type were not statistically significant. These findings support the lack of association between LOS and these independent variables.

Hospital B

The first step in the logistic regression process was to obtain an intercorrelation matrix of all variables included in the analysis. Table 8 presents results of this analysis.

Table 8

Pearson Product Moment Correlations
Pneumonia – Hospital B (N = 790)

Measure	1	2	3	4	5	6
1. Length of Stay – Pneumonia	--					
2. Age	.09*					
3. Marital Status	-.04	-.21**				
4. Race	.03	< -.01	-.08*			
5. Admission Type	.03	.01	.02	-.02		
6. Variable Costs	.51**	.14**	-.12**	.02	-.03	

Length of Stay: Less than or equal to usual and customary allowable inpatient days is coded as a 1; Age: Less than or equal to 74 years coded as a 1; Marital status: Not married is coded as a 1; Race: White is coded as a 1; Admission type: Emergency is coded as a 1, Variable costs less than or equal to $2,249 coded as a 1
**p ≤ .01; *p ≤ .05

Two statistically significant correlations were obtained between LOS (dichotomized into less than or equal to criteria LOS for DRGs and greater than criteria LOS) and age, r (790) = .09, p = .011 and variable costs (dichotomized into less than or equal to median [$2,249] variable costs and greater than median [$2,249] variable costs), r (790) = .51, p < .001. statistically significant correlations also were obtained between age and marital status, r (790) = -.21, p < .001 and

variable costs, r (790) = .14, p < .001. Marital status was significantly correlated with race, r (790) = -.08, p = .028 and variable costs, r = -.12, p = .001. The remaining correlations were not statistically significant.

A logistic regression analysis was used to determine if the five independent variables (age, marital status, race, admission type, and variable costs) could be used to predict the dependent variable, length of stay. Table 9 presents results of this analysis.

Table 9

Logistic Regression
Pneumonia Diagnoses – Hospital B (n = 790)

Variable	B	SE	Odds Ratio	Wald Statistic
Age	<-.01	.01	1.00	.04
Marital Status	.14	.19	1.14	.52
Race	-.32	.72	.73	.20
Admission Type	-.95	.51	.39	3.47
Variable Costs	<.01	<.01	1.00	140.53**

Note: Hosmer and Lemeshow Test = χ^2 (8) = 40.93, p < .001
**p ≤ .01; *p ≤ .05

The results of the logistic regression used to examine the relationships between the five independent variables and length of stay for pneumonia diagnoses at Hospital B were statistically significant, χ^2 (8) = 40.93, p < .001, indicating the five independent variables in the model were not a good fit for predicting LOS for pneumonia diagnoses. One independent variable, variable costs was a statistically significant predictor of length of stay, Wald = 140.53, p < .001. The associated odds ratio was 1.00, which indicated that the variable costs associated with pneumonia were equally likely for lengths of stays that were less than or equal to or less than the allowable criteria for each pneumonia DRG. The remaining variables were not statistically significant predictors of LOS in Hospital B.

The same five independent variables were crosstabulated by the dichotomized length of stay. Mantel-Haenszel statistical tests were used to determine if the association between the variables were statistically significant. Table 10 presents results of this analysis.

Table 10

Mantel-Haenszel Statistical Procedures
Length of Stay (Pneumonia) Hospital B

Patient Characteristic	Length of Stay				Total (N = 790)	
	≤ Criteria for DRG (n = 524)		> Criteria for DRG (n = 266)			
	N	%	N	%	N	%
Age						
≤ 74 years of age	253	48.3	103	38.7	356	45.1
> 74 years of age	271	51.7	163	61.3	434	54.9
Mantel-Haenszel = 6.13, p= .013						
Marital Status						
Unmarried	278	53.1	153	57.5	431	54.6
Married	246	46.9	113	42.5	359	45.4
Mantel-Haenszel = 1.24, p = .265						
Race						
White	517	98.7	260	97.7	777	98.4
Minority	7	1.3	6	2.3	13	1.6
Mantel-Haenszel = .44, p = .507						
Admission Type						
Emergency/Urgent Care	513	97.9	258	97.0	771	97.6
Elective	11	2.1	8	3.0	19	2.4
Mantel-Haenszel = .29, p = .588						
Variable Costs						
≤ $2,249	370	70.6	46	17.3	416	52.7
>$2,249	154	29.4	220	82.7	374	47.3
Mantel-Haenszel = 198.81, p ≤ .001						

**p ≤ .01; *p ≤ .05

The results of the Mantel-Haenszel analysis for age was statistically significant, M-H statistic = 6.13, $p < .013$. The odds ratio estimate for age was 1.48 (95% CI = 1.09 to 2.00). A statistically significant result was obtained on the Mantel-Haenszel procedures for variable costs, M-H statistic = 198.81, $p < .001$ in Hospital B. The odds ratio estimate of 11.49 (95% CI = 7.94 to

16.62) indicated that patients whose variable costs were greater than the median were more likely to exceed the criteria allowable LOS for their specific DRGs than those whose variable costs were less than the median.

The remaining Mantel-Haenszel analyses for marital status, race, and admission type were not statistically significant for Hospital B. These findings support the lack of association between LOS and these independent variables.

Myocardio Infarction

Hospital A.

The first set of analyses for myocardio infarction diagnoses was logistic regression. The dependent variable was length of stay, with age, marital status, race, admission type, and variable costs used as the independent variables. An intercorrelation matrix was developed as the first step of this analysis. Table 11 presents results of these analyses.

Table 11

Pearson Product Moment Correlations
Myocardio Infarction – Hospital A (N = 383)

Measure	1	2	3	4	5	6
1. Length of Stay – Myocardio Infarction	--					
2. Age	.19**					
3. Marital Status	-.09	-.30**				
4. Race	.04	-.10*	-.07			
5. Admission Type	-.06	-.02	-.11*	-.02		
6. Variable Costs	.42**	-.08	-.05	.02	-.01	

Length of Stay: Less than or equal to usual and customary allowable inpatient days is coded as a 1; Age: Less than or equal to 76 years coded as a 1; Marital status: Not married is coded as a 1; Race: White is coded as a 1; Admission type: Emergency is coded as a 1; Variable costs: Less than or equal to $3,275 coded as a 1.
**$p \leq .01$; *$p \leq .05$

Two statistically significant correlations were obtained between length of stay and age, r = .19, p < .001 and variable costs, r = .42, p < .001. Age was significantly correlated with marital status, r = -.30, p < .001 and race, r = -.10, p = .049. The correlation between marital status and admission type, r = -.11, p = .039 also was statistically significant. The remaining correlations in the analysis were not statistically significant.

A logistic regression analysis was used to determine which of the five independent variables (age, marital status, race, admission type, and variable costs) were statistically significant predictors of LOS for myocardio infarction dichotomized into less than or equal to the criteria allowable days and greater than allowable days. The results of this analysis are presented in Table 12.

Table 12

Logistic Regression
Myocardio Infarction – Hospital A (N = 383)

Variable	B	SE	Odds Ratio	Wald Statistic
Age	.03	.01	1.03	9.82**
Marital Status	.15	.24	1.16	.39
Race	-1.17	1.08	.31	1.16
Admission Type	.72	.74	2.05	.93
Variable Costs	<.01	<.01	1.00	43.11**

Note: Hosmer and Lemeshow Test = χ^2 (8) = 15.06, p = .058
**p ≤ .01; *p ≤ .05

The results of the logistic regression were not statistically significant, χ^2 (8) = 15.06, p = .058, indicating the five independent variables were providing a good fit for the model predicting LOS for myocardio infarctions in Hospital A. However, two independent variables, age, Wald = 9.82, p = .002, and variable costs, Wald = 43.11, p ≤ .001 were found to be statistically significant predictors of LOS. The Wald statistics for the remaining independent variables were not

statistically significant, indicating they were not predictors of LOS. The odds ratio of 2.05 for admission type indicated that patients who were admitted through the emergency room or urgent care were twice as likely to have longer lengths of stay as patients who had elective admissions.

The second type of analysis was the Mantel-Haenszel. The dependent variable was length of stay dichotomized into two groups, less than or equal to allowable days and greater than allowable days. The independent variables were age, marital status, race, admission type, and variable costs. The results of these analyses are presented in Table 13.

Table 13

Mantel-Haenszel Statistical Procedures
Length of Stay (Myocardio Infarction) – Hospital A

Patient Characteristic	Length of Stay				Total (N = 383)	
	≤ Criteria for DRG (n = 216)		> Criteria for DRG (n = 167)			
	N	%	N	%	N	%
Age						
≤ 76 years of age	140	64.8	77	46.1	217	56.7
> 76 years of age	76	35.2	90	53.9	166	43.3
Mantel-Haenszel = 12.64, p < .001						
Marital Status						
Unmarried	102	47.2	94	56.3	196	51.2
Married	114	52.8	73	43.7	187	48.8
Mantel-Haenszel = 2.74, p = .098						
Race						
White	214	99.1	164	98.2	378	98.7
Minority	2	.9	3	1.8	5	1.3
Mantel-Haenszel = .08, p = .772						
Admission Type						
Emergency/Urgent Care	208	96.3	164	98.2	372	97.1
Elective	8	3.7	3	1.8	11	2.9
Mantel-Haenszel = .64, p = .424						
Variable Costs						
≤ $3,275	149	69.0	44	26.3	193	50.4
> $3,275	67	31.0	123	73.7	190	49.6
Mantel-Haenszel = 66.61, p ≤ .001						

**p ≤ .01; *p ≤ .05

Two independent variables, age of patient, Mantel-Haenszel statistic = 12.64, $p \leq .001$, and variable costs Mantel-Haenszel statistic = 66.61, $p \leq .001$ were significantly associated with LOS. The obtained odds ratio for age of patient was 2.15 (95% CI = 1.43 to 3.25). The obtained odds ratio for variable costs was 6.22 (95% CI = 3.97 to 9.74). These odds ratios indicated that higher age and variable costs were associated with lengths of stay in the hospital that exceeded the criteria allowable length of stay. The association between LOS and the remaining independent variables were not statistically significant.

Hospital B.

An intercorrelation matrix was constructed to determine the extent to which the dependent (length of stay for myocardio infarction diagnoses) and independent variables (age, marital status, race, admission type, and variable costs) were related. Table 14 presents results of these analyses.

Table 14

Pearson Product Moment Correlations
Myocardio Infarction – Hospital B (N = 262)

Measure	1	2	3	4	5	6
1. Length of Stay – Myocardio Infarction	--					
2. Age	.02					
3. Marital Status	-.07	-.10				
4. Race	-.08	-.01	.01			
5. Admission Type	-.03	-.06	-.03	-.01		
6. Variable Costs	.56**	.05	-.09	-.09	-.05	

Length of Stay: Less than or equal to usual and customary allowable inpatient days is coded as a 1; Age: Less than or equal to 76 years coded as a 1; Marital status: Not married is coded as a 1; Race: White is coded as a 1; Admission type: Emergency is coded as a 1; Variable costs: Less than or equal to $3,275 coded as a 1.
**$p \leq .01$; *$p \leq .05$

One independent variable, variable costs, $r = .56$, $p < .001$, were significantly related to length of stay for myocardio infarctions. The remaining independent variables were not related to

length of stay. The intercorrelations among the independent variables also were not statistically significant.

To determine if the five independent variables, age, marital status, race, admission type, and variable costs, could predict length of stay for myocardio infarctions at Hospital B. Table 15 presents results of these analyses.

Table 15

Logistic Regression
Myocardio Infarction – Hospital B (N = 262)

Variable	B	SE	Odds Ratio	Wald Statistic
Age	.03	.02	1.03	5.08*
Marital Status	.14	.33	1.15	.17
Race	19.06	27861.92	1.90E+008	.00
Admission Type	-.26	1.08	.77	.06
Variable Costs	<.01	.00	1.00	55.76**

Note: Hosmer and Lemeshow Test = χ^2 (8) = 10.30, p = .245
**p ≤ .01; *p ≤ .05

The results of the Hosmer and Lemeshow Test was not statistically significant, χ^2 (8) = 10.30, p = .245. This result indicated that the five independent variables in the model were a good fit with the length of stay (LOS) for myocardio infarctions. Two independent variables; age, Wald = 5.08, p = .024 and variable costs, Wald = 55.76, p < .001; were significant predictors of LOS. Because of the unequal distribution between African American (n = 2) and White patients (n = 260) the odds ratio for race was out of range and should not be considered as a predictor of LOS for myocardio infarction patients at Hospital B.

The second set of analyses used the Mantel-Haenszel statistical procedures to determine the association between LOS for myocardio infarction patients and the five independent variables (age, marital status, race, admission type, and variable costs). The results of this analysis are

presented in Table 16.

Table 16

Mantel-Haenszel Statistical Procedures
Length of Stay (Myocardio Infarction) – Hospital B

Patient Characteristic	Length of Stay				Total (N = 262)	
	≤ Criteria for DRG (n = 152)		> Criteria for DRG (n = 110)			
	N	%	N	%	N	%
Age						
≤ 76 years of age	75	49.3	52	47.3	127	48.5
> 76 years of age	77	50.7	58	52.7	135	51.5
Mantel-Haenszel = .04, p = .837						
Marital Status						
Unmarried	80	52.6	66	60.0	146	55.7
Married	72	47.4	44	40.0	116	44.3
Mantel-Haenszel = 1.12, p = .291						
Race						
White	150	98.7	110	100.0	260	99.2
Minority	2	1.3	0	0.0	2	0.8
Mantel-Haenszel = .24, p = .626						
Admission Type						
Emergency/Urgent Care	148	97.4	108	98.2	256	97.7
Elective	4	2.6	2	1.8	6	2.3
Mantel-Haenszel = .00, p = .987						
Variable Costs						
≤ $3,275	111	73.0	18	16.4	129	49.2
> $3,275	41	27.0	92	83.6	133	50.8
Mantel-Haenszel = 79.42, p ≤ .001						

**p ≤ .01; *p ≤ .05

One independent variable, variable costs, produced a statistically significant result on the Mantel-Haenszel analysis, M-H statistic = 79.42, $p < .001$. The odds ratio estimate for this variable was 13.84 (95% CI = 7.45 to 25.70). The remaining variables did not provide any evidence of statistically significant associations between each independent variable and the LOS.

Summary

The results of the data analysis used to address the research question developed for the

study has been presented in this chapter. Conclusions and recommendations made for this study are included in Chapter V.

Chapter 5

Summary, Conclusions, and Recommendations

Summary

The purpose of this study was to determine if either logistic regression or Mantel-Haenszel statistical procedures could be used to introduce a more accurate, simpler to understand, and statistically sound approach to identifying specific patient variables that could influence LOS. With these types of statistical analysis, hospital administrators and government officials responsible for decision making in health care could have valid and reliable information on which to base their decisions.

Logistic Regression

Logistic regression is a method for determining the relationship between independent variables and a dichotomously-coded dependent variable. Modelling with logistic regression allows one to contrast different theoretical sets of predictor variables. Logistic regression methods are analogous to multiple linear regression methods when the dependent measure is dichotomous (coded into variables of 0 and 1). A common way to assess the influence of an independent variable on the dependent variable is to look at the odd-ratio that is an index of how likely it is that the client scored either of the two alternatives given values of the independent variable.

Logistic regression, being a special class of regression models, is appropriate for the study of categorical dependent variables, such as staying in or dropping out from college. This technique is increasingly applied in educational research. A search of the ERIC database indicated that between January 1988 and December 1999, a keyword search on terms such as “logistic regression,” “logit,” “probit,” “normit,” or “tobit” produced 90 abstracts that investigated education issues, out of total of 233 (or 38.63%). The proportion of higher education

related articles increased noticeably beginning in 1992. In recent years, the application of logistic regression has been found more frequently in the annual meeting programs of the Postsecondary Education Division of the American Educational Researcher Association than in those of the other 11 divisions. The trend in higher education is for researchers to recognize limitations of ordinary least squares (OLS) regression and turn increasingly to logistic regression for explaining relationship between a categorical dependent variable and a mixture of continuous and categorical independent variables. This trend is primarily motivated by complex data and categorical outcome measures, for example, enrolment/matriculation, retention, and graduation that are of interest to higher education researchers.

Despite the popularity of logistic regression in recent years, confusion continues to exist over terms, concepts, practices, and interpretations. Logistic regression results have been reported in terms of logit, odds, odds ratio, relative risk, predicted probability, marginal probability (also called marginal effect, partial effect, or partial change), and change in predicted probability (also called delta-P). The present study used logistic regression to determine the relationship between set of independent variables (age, marital status, race, admission type and variable costs) and dichotomized length of stay for two sets of DRGs (pneumonia and myocardio infarction).

The first use of logistic regression is prediction of group membership. As logistic regression calculates the probability of success over the probability of failure, results of the analysis generally are in the from an odds ratio. For example, logistic regression is often used in epidemiological studies where the result of the analysis is the probability of developing cancer after controlling for other associated. Logistic regression also can provide knowledge of the relationships and strengths among the variables (e.g., smoking 10 packs a day puts you at a higher risk for developing cancer than working in an asbestos mine). The process by which

coefficients are tested for significance for inclusion or elimination from the model involves several different techniques.

Wald test.

A Wald test is used to test the statistical significance of each independent variable in the model. A Wald test calculates a z statistic. This z value is then squared, yielding a Wald statistic with a chi-square distribution. However, several authors have identified problem with the use of the Wald statistic. Menard (1995) warned that for large coefficients, standard error can be inflated, lowering the Wald statistic (chi-square) value. Agresti (1996) argued that the likelihood-ratio test is a more reliable test for small sample sizes than the Wald test.

Hosmer-Lemshow Goodness of Fit test.

The Hosmer-Lemshow statistic evaluates the goodness-of-fit by creating 10 ordered groups of subjects and then compares the number actually in the each group (observed) to the number predicted by the logistic regression model (predicted). Thus, the test statistic is a chi-square statistic with nonsignificance the desired outcome, indicating that the model prediction does not significantly differ from the observed.

The 10 ordered groups are created based on their estimated probability; those with estimated probability below 0.1 from one group, and so on, up to those with probability 0.0 to 1.0. Each of these categories is further divided into two groups based on the actual observed outcome variable (success, failure). The expected frequencies for each of the cells are obtained from the model. If the model is good, the most of the subjects with success are classified in the higher deciles of risk and those with failure in the lower deciles of risk.

Mantel-Haenszel Statistical Procedure

The Mantel-Haenszel statistic has been designed to test the association between two dichotomous variables using information from several 2 x 2 tables. This statistic is extended here

for incidence data on a rare outcome arising at two or more levels of an exposure variable.

One of the most popular procedures for assessing differential item functioning (DIF) in dichotomous items is the Mantel-Haenszel (MH) procedure. This procedure was developed first for use in epidemiological research (Mantel & Haenszel, 1959), and later applied to the detection of DIF by Holland and Thayer (1988). Differential item functioning (DIF) is present when examinees from different groups have a different probability or likelihood of answering an item correctly, after conditioning on ability.

Applying the MH procedure to DIF detection begins by grouping examinees according to an estimate of ability (generally the total test score), and then forming a two-by-two contingency table crossing group membership (reference and focal) and item performance (correct and incorrect) for each level of ability. Let us denote a particular level of ability by k, where k = 1, 2, . . . m. Then, the MH chi-square statistic can be used to assess the association between group membership and item performance across all m levels of the estimated ability. The MH chi-square is distributed approximately as a chi-square variate with one degree of freedom (Mantel & Haenszel, 1959).

Conditioning on ability is an important step because it ensures that examinees are matched on a common measure before they are compared. To produce accurate results, the conditioning variable must also provide a valid measure of the construct of interest for both groups (Ackerman & Evans, 1994; Clauser, Nungester, Mazor, & Ripkey, 1998). Although many DIF methods are available, a relatively small number of these methods are “preferred” because of their theoretical and empirical strengths (Clauser & Mazor, 1998). One of the preferred methods that used frequently to detect items with DIF is Mantel-Haenszel (MH).

Health Care Issues

United States health care expenditures have continued to escalate during the past 40 years at a rate substantially higher than the rate of inflation. This escalation coincides with the growth of public health insurance (i.e., Medicare and Medicaid programs) and the subsequent growth of employer paid insurance programs. This growth has led to cost containment programs that are intended to reduce continued growth of total health care costs. Capitation programs, where a pre-determined payment is provided, were initiated to control the cost of hospitalization, the largest percentage of total costs. Ironically, the first successful large scale capitation program was Medicare's Prospective Payment System (PPS) which used Diagnostic Related Groups (DRGs) to determine hospital payments. The PPS identified the inpatient hospitalization payment by DRG that Medicare program would pay for care. The DRG was assigned based on the patient's discharge diagnosis. Since Medicare is a federally-funded program, hospitals had to accept this payment. Soon after, commercial HMO and PPO plans followed suit.

Governmental agencies have many reasons to keep health care costs down. First and foremost, they end up paying for a large portion of the nation's healthcare through programs such as Medicaid and Medicare. The Center for Medicaid and Medicare Services (CMS), which oversees these programs, has been analyzing clinical outcomes by DRG for more than 20 years. The government is also responsible for paying for care received through the Veterans Affairs (VA), the Department of Defense Tricare program responsible for medical care for families of military members, the Military Health System (MHS) that provides care to individuals in the military and federal employees. Each of these programs is massive in the number of people they treat and the cost for providing this care. Analysis of DRG-based data is important for governmental organizations to determine the effectiveness of cost containment strategies.

Vaul (1998) recognized that one of the problems with the use of data collected relative to the development of DRGs is the lack of national coding norms relative to DRGs that support the systematic designation and billing functions relative to DRG-related data. Claims have been made that the quality of care and the function of funding systems have declined significantly as a result of the lack of coding norms (Vaul, 1998). Vaul also reported that there are substantial issues relative to diagnostic groups and what has been described as "upcoding," or the repetitive coding of multiple conditions in atypical cases. For example, atypical cases relating to DRG pairs can substantially increase costs to the federal government. Over time, both government agencies who review DRG based data and the hospital staff who code the discharge diagnosis have improved their techniques for identifying upcoding and their basic coding procedures.

The principal motivation for both insurers and government agencies in using capitation systems of payment has been to control costs (Catalano, et al., 2000). The argument is that previously used fee-for-service (FFS) reimbursement encourages physicians to increase their financial benefits by dispensing expensive treatments or providing more services.

Along with the capitation systems, more stringent medical utilization review and cost containment programs have been used to supplement the capitation systems. Utilization review programs strive to monitor and challenge the use of expensive or redundant treatments. The primary goal of cost containment, utilization review, and capitation is to reduce costs, but the secondary objectives are to increase access and improve the quality of health care.

To better determine the effectiveness of cost containment, many hospitals and medical providers, as well as government and private insurance groups are beginning to use statistical analysis as a way to summarize data on DRGs and associated variables that impact efforts to decrease costs while improving health care delivery.

Typical methods that health care administrators have been used to summarize data included simple descriptive statistics (e.g., means, standard deviations, medians, etc.) while displaying results as simple histograms and bar charts. While descriptive statistics can be useful for specific reporting, basing important decisions on these types of statistics is neither wise nor appropriate. Inferential statistics were generally not used due to the perceived complexity and difficulty in interpreting the results. Within the health care industry, conveying information in a way that could be understood by all stakeholders is important. The methods used to examine hospital data need to be valid and reliable, as well as acceptable to physicians and other health care personnel.

One of the most important variables in hospitals is length of stay (LOS). Reimbursement for in hospital medical services is based on an average LOS for specific DRGs. Hospitals that are able to release a patient in fewer days than the LOS designated by the DRG make money on the patient, while patients who remain in the hospital for longer periods than allowed by the DRG result in losses to the hospital. Therefore, doctors and other health care professionals are being pressured to discharge patients in a timely manner.

Methods

To test the applicability and appropriateness of the logistic regression and Mantel-Haenszel statistical procedures for the healthcare data, a health system in one eastern state was contacted. After receiving approval from the executive leadership of the healthcare system, the data administrator was contacted. To ensure the data accuracy and completeness a detailed data dictionary was developed and provided to the data administrator. The data administrator reviewed and modified the data dictionary based on healthcare data sharing protocols and guidelines related to patient privacy and confidentiality. The data administrator then provided three years of data on the variables related to the LOS (inpatient length of stay) for two major

DRGs (Pneumonia and Myocardio Infraction) which were categorized into four sub-groups based on the patient risk and acuity. Data for the three years (2003, 2004, and 2005) were provided in ASCII format, with all patient identifiers removed to ensure privacy and confidentiality of the patient information. The three years of data were combined into a single data file, which was reformatted into an Excel™ file for use with SPSS – Windows, ver. 15.0.

The dependent variable, length of stay, for use in the statistical analyses, the DRGs for both pneumonia and myocardio infarction were each categorized into four specific DRGs based on the severity of the illness. Using usual and customary length of stay (LOS) for the four DRGs, length of stay was dichotomized into less than or equal to or greater than the usual and customary LOS. The dichotomized LOSs for the four DRGs for each type of diagnosis (pneumonia/myocardio infarction) were then combined to develop the dependent variables that categorized the LOS for each patient as either less than or equal to or greater than the respective LOS for their DRGs.

The independent variables measuring the demographic characteristics (age, marital status, admission type, variable costs, and race) of the patients were dichotomized for use in the logistic regression and Mantel-Haenszel DIF procedure. Age was dichotomized using a median split that was determined using descriptive statistics. The reported marital status of the patients was dichotomized into married/not married. Race was dichotomized into White/minority. Hospital variables (admission type and variable costs) also were dichotomized. The admission types were combined into emergency/urgent care and elective. A median split was used to divide variable costs into high and low. The continuous variables (age and variable costs) were dichotomized for each diagnosis (pneumonia and myocardio infarction) separately.

The independent variables, age, marital status, race, admission type, and variable costs, were compared between the two hospitals that provided data for the study. As statistically

significant differences were noted for age, marital status, and variable costs for pneumonia DRGs and age for myocardio infarction, data from each hospital were analyzed separately.

The research questions posted for this study was:

> Which statistical procedure, logistic regression or Mantel-Haenszel DIF produces results that can be used to predict or explain length of stay for specific DRGs (pneumonia – minor, moderate, major, and extreme and myocardio infarction – minor, moderate, major, and extreme).

Pneumonia Diagnoses

Hospital A. The results of the Pearson product moment correlations used to examine the relationships between the dependent variable and the independent variables provided evidence of statistically significant relationships between LOS for pneumonia diagnoses with age and variable costs. The significant result for the Hosmer and Lemeshow test provided evidence that the model using the five independent variables was not a good fit for predicting LOS. In examining the Wald statistics, age and variable costs were statistically significant predictors of LOS for pneumonia diagnoses. The odds ratios associated with these independent variables were 1.01 and 1.00 respectively. The other three independent variables were not statistically significant predictors of LOS for pneumonia diagnoses in Hospital A.

Results of the Mantel-Haenszel statistical procedures crosstabulated LOS for pneumonia diagnoses in Hospital A. Results of these analyses produced two statistically significant outcomes for age and variable costs. In examining the outcomes, it appeared that younger patients were more likely to have shorter LOS and older patients were more likely to exceed the criteria LOS for DRGs. Younger patients were more likely have lower variable costs for pneumonia diagnoses, while those who were older tended to be associated with variables costs that were greater than the median. The other independent variables did not appear to be associated with LOS.

Hospital B. Results of the Pearson product moment correlations used to examine the relationships between the dependent variable and the independent variables provided evidence of

statistically significant relationships between LOS for pneumonia diagnoses with age and variable costs. The correlations between age and marital status and variable costs were significant, as was the correlations between marital status and race and variable costs. Based on the statistically significant results for the Hosmer and Lemeshow test, the five independent variables in the model predicting LOS were not a good fit. By further examining the analysis results based on the Wald statistics, variable costs was statistically significant predictors of LOS for pneumonia diagnoses at hospital B. The odds ratios associated with the independent variable (variable costs) were 1.00 that indicated variable cost did not have an increasing effect on the pneumonia LOS at Hospital B. The other four independent variables were not statistically significant predictors of LOS for pneumonia diagnoses in Hospital B.

The results of the Mantel-Haenszel statistical procedures crosstabulated LOS for pneumonia diagnoses in Hospital B produced two statistically significant outcomes for age and variable costs. In examining the outcomes, it appeared that younger patients were more likely to have shorter LOS and older patients were more likely to exceed the criteria LOS for DRGs. Younger patients were more likely have lower variable costs for pneumonia diagnoses, while those who were older tended to be associated with variables costs that were greater than the median. The other independent variables did not appear to be associated with LOS.

Myocardio Infarction Diagnoses

Hospital A. To examine the effect of the independent variables, age, marital status, race, admission type, and variable costs on the LOS for myocardio infarction diagnoses, an intercorrelation matrix using Pearson product moment correlations was developed. The results of the Pearson product moment correlations provided statistically significant relationships between LOS for myocardio infarction diagnoses with age and variable costs. A statically significant negative correlation was found between age and marital status and race were identified. The

nonsignificant results for the Hosmer and Lemeshow test indicated that the model comprised of the five independent variables were a good fit to predict LOS. By further reviewing the analysis results based on the Wald statistics, age and variable costs were statistically significant predictors of LOS for myocardio infarction diagnoses at hospital A. The odds ratios associated with these independent variables were 1.03 and 1.00 respectively. The other three independent variables were not statistically significant predictors of LOS for Myocardio infarction diagnoses in Hospital A. Although no statistically significant relationships were obtained between admission type and LOS, the odd ratio of 2.05 for admission type inferred that emergency and urgent admission had an increasing effect (by factor of 2.05) on LOS for myocardio infarction diagnoses in Hospital A.

The Mantel-Haenszel statistical analyses produced two statistically significant outcomes for age and variable costs. In examining the outcomes based on the produced odd ratios, it appeared that younger patients were more likely to have shorter LOS and older patients were more likely to exceed the criteria LOS for DRGs. Younger patients were more likely to have lower variable costs for Myocardio infarction diagnoses, while those who were older tended to be associated with variables costs that were greater than the median. The other three independent variables (Marital status, Race and Admission type) did not appear to be associated with LOS.

Myocardio Infarction Diagnoses – Hospital B

To examine the effect of the independent variables, age, marital status, race, admission type, and variable costs on the Myocardio infarction LOS, an intercorrelation matrix using Pearson product moment correlations was developed. Statistically significant relationships were obtained between LOS for myocardio infarction diagnoses with variable costs. The nonsignificant result for the Hosmer and Lemeshow test, found that the model using the five independent variables was a good fit for predicting LOS. By further reviewing the Wald

statistics, age and variable costs were statistically significant predictors of LOS for myocardio infarction diagnoses at Hospital B. The odds ratios associated with these independent variables were 1.03 and 1.00 respectively. The other three independent variables were not statistically significant predictors of LOS for myocardio infarction diagnoses in Hospital B. Due to the unequal distribution between patients who were White or minority, an out of range odd ratio was produced that should not be considered as a predictor of LOS for myocardio infarction diagnoses in Hospital B. The other two independent variables were not statistically significant predictors of LOS for Myocardio infarction diagnoses in Hospital B.

The results of the Mantel-Haenszel statistical procedures crosstabulated LOS for Myocardio infarction diagnoses in Hospital B. The Mantel-Haenszel statistical procedures analysis produced one statistically significant outcome for variable costs. The odd ratio of 13.84 for the variable costs states that higher LOS for myocardio infarction diagnoses in hospital B associated with the higher variable costs. The other four independent variables (age, marital status, race and admission type) did not appear to be associated with LOS for myocardio infarction diagnoses in Hospital B.

Conclusion

This study compared the performance of Mantel-Haenszel (MH) statistical procedures and the logistic regression (LR) procedure in analyzing healthcare length of stay (LOS) data. The results of the study indicated that the logistic regression may be better suited for this purpose, although results appeared to be similar for both types of analyses.

The Mantel-Haenszel (MH) procedure is sensitive to only one type of differential item functioning (DIF) and was not designed to detect DIF that has a non uniform effect across trait levels. The MH procedure can be used to evaluate one item at a time, so it is time intensive and multiple analyses are likely to increase the α value, making Type 1 errors a possibility. While the

MH results may be easier to explain to nonstatistician audiences, the limitation of MH relating to evaluation of a single item, with the uniformity of the effect across trait levels limiting the use of MH DIF in analyzing healthcare LOS data.

Logistic regression (LR) allows analysis of the independent variables simultaneously or in a stepwise procedure. Although the dependent variable must be dichotomous, the results can be interpreted as model fitting. When interpreting the results of the LR procedure, inferences can be drawn regarding which variables should be included in prediction model and which can be excluded from further analyses. Additional advantages of the logistic regression are as follows:

1. A linear relationship between the dependent and the independent variables is not required for using Logistic regression. Logit link function in logistic regression is nonlinear so it can handle nonlinear effects, even when exponential and polynomial terms are not explicitly added as additional independents.
2. The normal distribution of the dependent variable is not necessary but it needs to be assumed as an exponential distribution family (such as normal, Poisson, binomial, gamma).
3. The homoscedasticity of dependent variable for each level of the independents is not required; that is, variances need not be equal within categories.
4. Error terms are not assumed to be normally distributed.
5. Logistic regression allows both continuous and categorical independent variables.
6. Logistic regression does not require that the independent variables be unbounded.

Limitations

The following limitations have affected the generalizations of this study:

The study used retrospective data from two hospitals. These hospitals were located in a community with a majority White population in a single state. The data were limited to five

variables (length of stay, age, marital status, race, admission type, and variable costs) for patients. Additional variables (e.g., gender, educational level, income level, etc.) may have provided a better fit for the study. In addition, the use of two specific groups of DRGs may have provided results that would not be replicated with other types of DRGs.

Recommendations for Implementation

While the data used in this study are not unique, the analyses of these data using different statistical procedures generally have not been used in health care settings. Based on the study findings, the following recommendations for hospital administrators are proposed:

1. The logistic regression (LR) method may have advantages over MH statistical procedures. To enhance decision making by hospital clinical and nonclinical staff, LR in LOS data analysis is recommended.
2. The use of statistical software packages that can accommodate LR analyses is recommended for hospitals and other healthcare delivery systems to ensure that decision making is based on accurate and appropriate data and assumptions.
3. Hospitals or health systems should use complete data sets to identify statistically significant relationships between dependent and independent variable.
4. Hospitals/healthcare systems can develop LR model for dominant DRGs (based on volume or cost) and periodically update the model based on current available data.
5. Hospitals should test their developed model on an ongoing basis to ensure the accuracy and reliability of the model.

Recommendations for Future Research

The following recommendations should be used to continue research on appropriate statistical procedures that can be used to develop predictive models for healthcare:

1. Conduct logistic regression for the same DRGs at other hospitals and health systems to determine if results are similar to the present study.
2. Use a Monte Carlo statistical procedure to evaluate the efficacy of using LR or MH to analyze the effects of demographic and hospital variables on LOS.
3. Apply different statistical procedures to develop models that can be used for forecasting hospital programs to reduce costs and improve quality of care.
4. Create comparative analysis between hospitals' LOS for similar DRGs to identify benchmarks that hospitals can use to develop appropriate protocols and guidelines.
5. Evaluate different statistical analysis procedures that could be use to analyze healthcare data and identify the most appropriate statistical procedures for healthcare data.

Appendix A

Human Investigation Committee Approval

WAYNE STATE UNIVERSITY

HUMAN INVESTIGATION COMMITTEE
101 East Alexandrine Building
Detroit Michigan 48201
Phone: (313) 577-1628
FAX: (313) 993-7122
http://hic.wayne.edu

CONCURRENCE OF EXEMPTION

To: Reza Ziaee

Deans Office Education

From: Ellen Barton, Ph.D. Nevra Nahan for E

Chairperson, Behavioral Institutional Review Board (B3)

Date: December 21, 2006

RE:

HIC #:	127706B3X
Protocol Title:	A Comparison of Two Statistical Techniques in Predicting Length of Stay for Two Diagnostic Related Groups
Sponsor:	
Coeus #:	0612004439

Risk Level: No greater than minimal risk.

The above-referenced protocol has been reviewed and found to qualify for **Exemption** according to paragraph #1 of the Department of Health and Human Services Code of Federal Regulations [45 CFR 46.101(b)].

This proposal has not been evaluated for scientific merit, except to weigh the risk to the human subjects in relation to the potential benefits.

- Exempt protocols do not require annual review by the IRB.
- All changes or amendments to the above-referenced protocol require review and approval by the HIC **BEFORE** implementation.
- Adverse Reactions/Unexpected Events (AR/UE) must be submitted on the appropriate form within the timeframe specified in the HIC Policy (http://www.hic.wayne.edu/hicpol.html).

NOTE:

1. Forms should be downloaded from the HIC website at each use.
2. Submit a Closure Form to the HIC Office upon completion of the study.

References

Agency for Health Research and Quality (AHRQ). Statistics are from the HCUP-3 Nationwide Inpatient Sample for 1994: Diagnosis-Related Groups. Retrieved on January 5, 2006 from http://www.ahrq.gov/data/hcup/94drga.htm#103.

Aldrich, J., & Nelson, F. (1984). *Linear probability, logit, and probit models*. Beverly Hills, CA: Sage.

American Hospital Association (2003). *National Hospital Indicators Survey: 1997 – 2002.* Unpublished Report, Chicago.

America's Health Insurance Plans (AHIP), (2006) *The Factors Fueling Rising Healthcare Costs 2006* Retrieved on January 22, 2007 from http://www.ahip.org/redirect/PwCCostOfHC2006.pdf

Armas, G. (2002). People 65-plus increase by 300%. 420 million seniors worldwide retire earlier, live longer. *Associated Press*. Retrieved November 17, 2004 from: http://detnews.com/2001/health/0112/13/a04-365979.htm

Atchley, R. (2002, February). *Continuity Theory: How Elders Find Wisdom in Spite of It All*. Retrieved February 3, 2005 from: http://www.asaging.org/at/at-214/continuity.html

Averill, R. F., McGuire, T. E., Manning, B. E., Fowler, D. A. Horn, S. D. Dickson, P. S., Coye, M. J., Knowlton, D. L., Bender, J. A. (1992). A Study of the Relationship between Severity of Illness and Hospital Cost in New Jersey Hospitals. *Health Services Research 27*(5): 587-617.

Baker, J. (2002). Medicare Payment System for Hospital Inpatients: Diagnosis-Related Groups. *Journal of Healthcare Finance 28,* (3), 1-13.

Balfour, D. L., & Neff, D. M. (1993). Predicting and managing turnover in human service agencies: A case study of an organization in crisis. *Public Personnel Management, 22*, 473-486.

Berger, K. (2000). *The Developing Person through the Life Span*. Palgrave.

Billings, J. & Cantor, J.C. (2002). Access. In A. R. Kovner, & S. Jonas, S. (Eds.). *Health Care Delivery in the United States*. New York, NY: Springer Publishing Company.

Block, L. E. (1997). Evolution, growth, and status of managed care in the United States, *Public Health Review (3)*, 195-200.

Borger, C., Smith S. Truffer, C., Keehan, S., Sisko, A., Poisal, J. et al. (2006) Health spending projections through 2015: Changes on the horizon. *Health Affairs. 25*(2), w61-w73

Boyer, C., McAlpine, D., Pottick, K., & Olfson, M. (2000). Identifying Risk Factors and Key Strategies in Linkage to Outpatient Psychiatric Care. *American Journal of Psychiatry 157*(10), 1592.

Budgell, G. R., Raju, N. S., & Quartetti, D. A. (1995). Analysis of differential item functioning in translated assessment instruments. *Applied Psychological Measurement, 19*, 309-321.

Camilli, G., & Shepard, L. (1994). *Methods for identifying biased test items*. Thousand Oaks, CA: Sage.

Camilli, G., & Smith, J. K. (1990). Comparison of the Mantel-Haenszel test with a randomized and a jackknife test for detecting biased items. *Journal of Educational Statistics, 15*, 53-67.

Cantone, L. (1999, January). Corporate Compliance: Critical To Organizational Success. *Nursing Economics 17*, 15.

Casalino, L.P., Devers, K.J. & Brewster, L. R. (2003). Focused factories? Physician-owned specialty facilities? *Health Affairs, 22*(6), 56.

Cassas, M., & Wiley, M.M., eds. *Diagnosis Related Groups in Europe*. Berlin, Germany: Springer, 1993.

Catalano, R., Libby, A., Snowden, L. & Cuellar, A.E. (2000). The effect of capitated financing on mental health services for children and youth, the Colorado Experience. *American Journal of Public Health, 90*(12), 1861-1865.

Center for Medicare and Medicaid Services (CMS), Office of Research, Development and Information. (2000). *Data development by Planning and Policy Analysis Group, data from 1992, 1994, 1996, and 1998 HCFA 64*. Washington, DC: Author.

Center for Medicare and Medicaid Services (CMS), Office of Research, Development, and Information: Analysis of Data (2002). *Historical prospective of the Medicare and Medicaid programs*. Washington, DC: Author.

Cheah T S. (1998). The impact of clinical guidelines and clinical pathways on medical practice: effectiveness and medico-legal aspects. *Annals Academy of Medicine. 27*(4), 533-539.

Christensen, K.T. (2005). Ethically important distinctions among managed care organizations. In J.F. Monagle, & D.C. Thomasma, (Eds.) *Health Care Ethics, Critical Issues for the 21st Century.* Sudbury, MA: Jones and Bartlett Publishers.

Clauser, B., & Mazor, K. M. (1998). Using statistical procedures to identify differentially functioning test items. *Educational Measurement, Issues, and Practice, 17*(1), 31-44.

Clauser, B., Mazor, K. M., & Hambleton, R. K. (1994). The effects of score group width on the Mantel-Haenszel procedure. *Journal of Educational Measurement, 31*, 67-78.

Cleary, P. D., & Angel, R. (1984). The analysis of relationships involving dichotomous dependent variables. *Journal of Health and Social Behavior, 25*, 334-348.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum.

Collett, D. (1991). *Modeling binary data*. New York: Chapman Hall.

Cost Efficiency vs. Effective Care: Understanding What Works (2000*). American Society on Aging*. Retrieved June 26, 2005 from http://www.asaging.org/at/at-219/RT4.html.

Cox, D. R., & Snell, E. J. (1970). *Analysis of binary data*. New York: Chapman Hall.

Dalton, K., E. C. Norton, & K. Kilpatrick. (2001). A Longitudinal Study of the Effects of Graduate Medical Education on Hospital Operating Costs. *Health Services Research 35*(6), 1267-1291.

Deber, R. B. (2003). Health care reform: lessons from Canada. *American Journal of Public Health,* 93(1), 20.

Deming, W. E. (1986). *Out of the Crisis*. Cambridge, MA: MIT Press.

Donnelly, D., & Finkelhor, D. (1993). Who has joint custody? *Family Relations, 42*, 57-60.

Dougherty, C. J. (2005). Equality and inequality in American health care. J.F. Monagle, & D.C. Thomasma, (Eds.) *Health Care Ethics, Critical Issues for the 21st Century.* Sudbury, MA: Jones and Bartlett Publishers.

DRG. (The ABCs of Health Care). (2002, July 26). *Long Island Business News*. 49(31), S54.

Efron, B. (1975). The efficacy of logistic regression compared to normal discriminant analysis. *Journal of the American Statistical Association, 70*, 892-898.

Efron, B. (1980) Logistic regression, survival analysis, and the Kaplan-Meier curve, *Journal of the American Statistical Association, 83*, 414-425.

Eliason, S. R. (1993). *Maximum likelihood estimation: Logic and practice*. Newbury Park, CA: Sage.

Feder, P. I. (1975) The log likelihood ratio in segmented regression, *The Annals of Statistics, 3*, 84-97.

Federal Legislative History. (2004). Retrieved 2 March 2004 from www.nashia.org/ pdocfiles/Federal%20Legislative%20History.doc

Feldstein, P. J. (1996). The Politics of Health Legislation : An Economic Perspective . *Health Administration Press*.

Ferguson J H. (1991). Research on the delivery of medical care using hospital firms. *Med Care*; 29:JS1.

Findlay, S. (2003, May 14). To reform health care, bust myths. *USA Today*, 13a. Retrieved February 23, 2004 from www.elibrary.com.

Finkel, S. (2003, April). International psychogeriatrics: multiple disciplines come together to share knowledge of medicine. *Geriatrics*. Retrieved March 2, 2004 from www.elibrary.com

Finkelstein, B. S., Silvers, J.B., Marrero, U., Neuhauser, D. & Cuttler, L. (1998,). Insurance coverage, physician recommendations, and access to emerging treatments: growth hormone therapy for childhood short stature. *The Journal of the American Medical Association 279*(9), p. 663.

Fischer, G. H. (1993). Notes on the Mantel-Haenszel procedure and another chi-square test for the assessment of DIF. *Methodika, 7*, 88-100.

Fischer, G. H. (1995). Some neglected problems in IRT. *Psychometrika, 60*, 459-487.

Fowlkes, E. B. (1987). Some diagnostics for binary logistic regression via smoothing. *Biometrika, 74*, 503-515.

Friedman, J. H., & Silverman, B. W. (1989) Flexible parsimonious smoothing and additive modeling, *Technometrics, 31*, 3-39.

Gallant, A. R. (1975) Nonlinear regression, *The American Statistician, 29*, 73-81.

Ginsburg, P. B., & Carter, G. M. (1986). Medicare Case-Mix Index Increase. *Health Care Financing Review 7*(4), 5-65.

Goldsmith, J. (2003) The road to meaningful reform. A conversation with Oregon's John Kitzhaber. *Health Affairs 22*(1), 114.

Gosden, M. Pedersen, L. & Torgersen, D. (1999). How should we pay doctors? A systematic review of salary payments and their effect on doctor behavior. *Center for Health Economics, University of York*. Retrieved July 11, 2004 from http://www.eldis.org/static/DOC11142.htm.

Goss, S, S. (2000). Care: What lies ahead? (Interview). *Derma Sciences, Inc*. Retrieved June 26, 2005 from http://www.dermasciences.com/clinical/CareWhatLiesAhead.html.

Griffiths, D. A., & Miller, A. J. (1973) Hyperbolic regression--a model based on two-phase piecewise linear regression with a smooth transition between regimes, *Communications in Statistics, 2*, 561-569.

Grimshaw J M, Russell I T. (1993). Effect of clinical guidelines on medical practice: a systematic review of rigorous evaluations. *Lancet, 342*,1317-1322.

Guglielmo, W.J. (2004). A physician's guide to EMTALA rules: the rules prohibiting patient dumping by hospitals have been revised--and on-call doctors are directly affected. *Medical Economics 81*(22), 34-35.

Haberman, S. (1978).Analysis of qualitative data Vol. 1. New York: Academic Press.

Hackler, C. (2005). Is rationing of health care ethically defensible? In J. F. Monagle, & D. C. Thomasma, (Eds.) *Health Care Ethics, Critical Issues for the 21st Century.* Sudbury, MA: Jones and Bartlett Publishers.

Haddad, A. & Martin, S.S. (1997). What would you do? *RN 60*(9), 21-23.

Hagle, T. M., & Mitchell, G. E. (1992). Goodness-of-fit measures for probit and logit. *American Journal of Political Science, 36*, 762-784.

Hammer, P.J. & Sage, W.M. (2003). Critical issues in hospital antitrust law. *Health Affairs* 22(6), 88.

Hardy, M. A. (1993). *Regression with dummy variables*. Newbury Park, CA: Sage.

Harnisch, D. L. (1991). *Techniques for assessing differential item performance on achievements*. In Proceedings of the sixteenth annual SAS Users Group International Conference (pp. 1503-1508). Cary, NC: SAS Institute.

Harrel, F. E., Lee, K. L. & Pollock, B. G. (1988) Regression models in clinical studies: determining relationships between predictors and response, *Journal of the National Cancer Institute, 8*, 1198-1202.

Havel, J. (1993a). PASARR: What you should know about the new mental health requirements. Nursing Homes. Retrieved 2 March 2004 from www.elibrary.com

Havel, J. (1993b). Will this new government regulation be a boon or a bane? Nursing *Homes*. Retrieved 2 March 2004 from www.elibrary.com

Heffler, S., Smith, S., Won, G., Clemens, M., Keehan, S., Zezzan, M. (2002). Health spending projection for 2000-2011: The latest outlook. *Health Affairs, 21*(2), 208-212.

Herbert, B. (1996, May 6). Back to the streets. *The New York Times 145*, A15.

Hill-Burton Act (2003). Retrieved August 1, 2004 from http://www.infoline.org/InformationLibrary/Documents/Hill-Burton%20Act%20cw.asp.

Hoel, D. G., Wakabayashi, T., & Pike, M. C. (1983) Secular trends in the distributions of the breast cancer risk factors--menarche, first birth, menopause, and weight--in Hiroshima and Nagasaki, Japan. *American Journal of Epidemiology, 118*, 78-89.

Hoffman, E.D., Klees, B.S. & Curtis, C.A. (2003). Overview of the Medicare and Medicaid Programs. *Health Care Financing Review 24*(3), p. S1.

Holland, P. W., & Thayer, D. T. (1986). Differential item performance and the Mantel-Haenszel procedure. In H. Wainer & H. Braun (Eds.), *Test validity* (pp. 147-169). Hillsdale, NJ: Erlbaum.

Holland P., & Wainer H. (Eds.; 1993). *Differential Item Functioning*. Hillsdale, NJ: Lawrence Erlbaum Associates

Holleman, W.L., & Holleman, M.C. (1997). Are ethics and managed care strange bedfellows or a marriage made in heaven? *Lancet 349*(9048), 350-351.

Horn, S. D., Bulkey, G. Sharkey, P. D., Chambers, A. F. Horn, R. A., & Schramm, C. (1985). Interhospital Differences in Severity of Illness: Problems for Prospective Payment Based on Diagnosis-Related Groups. *New England Journal of Medicine 313*(1), 20-24.

Hosmer, D. W., & Lemeshow, S. (1989). *Applied logistic regression*. New York: Wiley.

Hosmer, D. W., Taber, S., & Lemeshow, S. (1991). The importance of assessing the fit of logistic regression models: A case study. *American Journal of Public Health, 81*, 1630-1635.

Hsia, D. C., & Ahern, C. (1992). Medicare Reimbursement Accuracy Under the Prospective Payment System, 1985 to 1988. *American Medical Association Journal 268*(7), 873-896.

Huberty, C. J., & Barton, R. M. (1989). An introduction to discriminant analysis. *Measurement and Evaluation in Counseling and Development, 22*, 158-168.

Hudson, D. J. (1966) Fitting segmented curves whose join points have to be estimated, *Journal of the American Statistical Association, 61*, 1097-1129.

Ilminen, Gary R. (2001) Nursing home survey reveals mixed consumer perceptions. *Nurse Week.* Retrieved January 12, 2004 from http://www.nurseweek.com/ industrypulse/ nursinghome.html

Iezzoni, L. I. (1997). Assessing quality using administrative data. *Annals of Internal Medicine, 15*(127, part 2), 666-674.

Johnson, D. H. Jr. (1999). Reforming Medicare for the New Century. The World & I. Retrieved 11 June 2005 from: www.elibrary.com

Jones, M. M., & Mandeville, G. K. (1990). The effect of age at school entry on reading achievement scores among South Carolina students. *Remedial and Special Education, 11*(2), 56-62.

Kahn, J. (2003). The end of cancer. *Wired Magazine, 11*, 8.

The Henry J. Kaiser Family Foundation, (2005). *The Medicare Chart Book (3rd Ed.), Summer 2005 - Section* . Retrieved on September 28, 2006 from http://www.kff.org/medicare/upload/Medicare-Chart-Book-3rd-Edition-Summer-2005-Section-1.pdf

Kane, R. L., & Matthias, R. (1984). From hospital to nursing home: The long-term care connection. *The Gerontologist, 24*(6), 604-609.

Knickman, J.R (2002) Futures. *Health Care Delivery in the United States*. Kovner, A.R. & Jonas, S.., Eds. New York, NY, Springer Publishing Company.

Kotelchuck, D. (1976). *Prognosis negative: Crisis in the health care s*ystem, Vantage Book, New York: Random House.

Kotelchuck, D. (Ed.), 1976. *Prognosis Negative: Crisis in the Health Care System*. Health Policy Advisory Center, Inc. Vantage.

Kovner, A.R. (2002). Hospitals. In A. R. Kovner, & S. Jonas, (Eds.) *Health Care Delivery in the United States*. New York, NY: Springer Publishing Company.

Land, K. C., McCall, P. L., & Parker, K. F. (1994). Logistic versus hazards regression analyses in evaluation research. *Evaluation Review, 18*, 411-437.

Landwehr, J. M., Pregibon, D., & Shoemaker, A. C. (1984). Graphical methods of assessing logistic regression models. *Journal of the American Statistical Association, 79*, 61-83.

Leapfrog Group, (2003). *Profiles of Organizations Using Quality Incentives*.. Retrieved on September 28, 2006 from http://www.leapfroggroup.org/RewardingResults/pdf/profiles.pdf.

Leary, R. S.; Johantgen, M. E; Farley D., Thane Forthman M., Dwight, J. & Wooster L. (1997) All-payer severity-adjusted diagnosis-related groups: A uniform method to severity-adjust discharge data. *Top Health Inform Manage, 17*(3), 60-71.

Legislative Update. (1998, Fall). *Health Care Financing Review 20*, 133.

Leichtman, M., Leichtman, M, Cornsweet, C., & Neese, D. (2001, April). Effectiveness of Intensive Short-Term Residential Treatment with Severely Disturbed Adolescents. *American Journal of Orthopsychiatry 71*(2), 227.

Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Hillsdale, NJ: Erlbaum

MacKinney, A. Clinton M.D., McBride, T. D., Shambaugh-Miller, M, D. & Mueller, K. J. (2003). Medicare Physician Payment: Practice Expense. *Rural Policy Brief, 8*(9) RUPRI Center for Rural Health Policy Analysis

Mantel, N., & Haenszel, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the National Cancer Institute, 22*, 719-748.

Marascuilo, L. A., & Levin, J. R. (1983). *Multivariate statistics in the social sciences: A researcher's guide*. Belmont, CA: Brooks/Cole.

Marek, K. D., Rantz, M. J., Fagin, C. M., & Krejci, J. W. (1996). OBRA '87: Has it resulted in better quality of care? *Journal of Gerontological Nursing, 22*(10), 28-36.

Mazor, K. M., Kanjee, A., & Clauser, B. E. (1995). Using logistic regression and the Mantel-Haenszel with multiple ability estimates to detect differential item functioning. *Journal of Educational Measurement, 32*, 131-144.

McAuley, W. J.; & Travis, S. S. (1997). Positions of influence in the nursing home admission decision. *Research on Aging*, 26(20).

McKinley, R. L. (1989). Methods, plainly speaking: An introduction to item response theory. *Measurement and Evaluation in Counseling and Guidance, 22*, 37-57.

McMahon, L., Fetter, R., & Thompson, J. (1986). Hospital matrix management and DRG-based prospective payment. *Hospital & Health Services Administration 31*(1), 62-75.

Menard, S. (1995). *Applied logistic regression analysis*. Thousand Oaks, CA: Sage.

Mercola, J. (2002, February*). $100 Billion for Medicare Prescriptions Will Further Devastate US Medicine.* Retrieved on June 26, 2005 from http://www.mercola.com/2002/feb/20/us_medicine.htm.

Montgomery D, Peck E., & Vining G. (2006). *Introduction to Linear Regression Analysis*. Hoboken, NJ: Wiley Interscience.

Morgan, S. P., & Teachman, J. D. (1988). Logistic regression: Description, examples, and comparisons. *Journal of Marriage and the Family, 50*, 929-936.

Mullins, L. C.-Elston, C. H, et al. (1996). Social determinants of loneliness among older Americans. *Genetic, Social & General Psychology* Monographs *122*, 455.

Mun, S. K., & Turner, J. W. (1999). Telemedicine: Emerging e-medicine. Department of Radiology, Georgetown University, Washington DC. Annual Review. *Biomed Engineering, 1.* 595- 608

Muse, W. V.; Burkhalter, B. B. (1998, July). Restructuring brings quality improvements to Auburn University. *Total Quality Management 9*(4-5).

Namerow, P. B., Kalmuss, D. S., & Cushman, L. F. (1993). The determinants of young women's pregnancy-resolution choices. *Journal of Research on Adolescence, 3*, 193-215.

Neter J, Kutner M., Nachtsheim C., & Wasserman W. (1996) *Applied Linear Statistical Models*. New York: WCB/McGraw-Hill.

Newhouse, J. (2002). Managed care: an industry snapshot. *Inquiry-Excellus Health Plan, 39*(3), 207.

Ott, M. D. (1988). An analysis of predictors of early academic dismissal. *Research in Higher Education, 28*, 34-48.

Ozar, D.T. (2005). The social obligations of health care practitioners. In J. F. Monagle, & D. C. Thomasma, D.C. (Eds.) *Health Care Ethics, Critical Issues for the 21st Century.* Sudbury, MA: Jones and Bartlett Publishers.

Parshall, C. G., & Miller, R. (1995). Exact versus asymptotic Mantel-Haenszel DIF statistics: A comparison of performance under small-sample conditions. *Journal of Educational Measurement, 32*, 302-316.

Peace of mind never gets old (2001, December). *Business Week 3761*, 95E2.

Pedhazur, E. J. (1982). *Applied regression in behavioral research: Explanation and prediction (2nd ed.)*. New York: Holt, Rinehart and Winston.

Pellegrino, E.D. (2005). Rationing health care: the ethics of medical gatekeeping. In J. F. Monagle, & D. C. Thomasma, (Eds.) *Health Care Ethics, Critical Issues for the 21st Century.* Sudbury, MA: Jones and Bartlett Publishers.

Penny, J. (1996). *Identifying IRT defined DIF using logistic regression and the Mantel-Haenszel procedure*. Paper presented at the annual meeting of the National Council on Measurement in Education, New York.

Penny, J. A., & Bond, L. (1992). *Using the area between two item response to index differential item functioning: A generalized approach*. Paper presented at the annual meeting of The Psychometric Society, Columbus, OH.

Penny, J., & Bond, L. (1995). *Using the area between two item response functions to index differential item functioning: A generalized approach using Monte-Carlo and GMAT data*. Paper presented at the annual meeting of The Psychometric Society, Minneapolis, MN.

Penny, J., & Johnson, R. L. (1996). *A study of the empirical relationship between group difference and Type I error rate when using logistic regression and the Mantel-Haenszel procedure to detect differential item functioning*. Paper presented at the annual meeting of the American Educational Research Association, New York.

Pneumothorax. (2004). MedicineNet.com. Retrieved December 29, 2004 from http://www.medicinenet.com/pneumothorax/article.htm.

Pokorny, A. D. (1993). Suicide prevention revisited. *Suicide and Life-Threatening Behavior, 23*(1), 1-10.

Pregibon, D. (1981). Logistic regression diagnostics. *Annals of Statistics, 69*, 705-724.

Press, J. S., & Wilson, S. (1978). Choosing between logistic regression and discriminant analysis. *Journal of the American Statistical Association, 73*, 699-705.

Preyra, C. (2004). Coding response to a case-mix measurement system based on multiple diagnoses. *Health Services Research 39*(4), 1027-1046.

Randolph, Frances L. (1995). Improving Service Systems Through Systems Integration: The ACCESS Program. *American Rehabilitation 21*(1), 36-37.

Rehnquist, Janet (2001). Memorandum to Thomas Scully, Administrator, Centers for Medicare and Medicaid Services: "Medicare inpatient hospital prospective payment system transfers incorrectly reported as discharges." *Department of Health & Human Service, Office of the Inspector General*. Retrieved September 27, 2004 from http://oig.hhs.gov/oas/reports/region6/60000041.pdf.

Robinson, J.C., & Casalino, L.P. (2002). Reevaluation of capitation contracting in New York and California. *Health Affairs, 21*, W11.

Robst, J. & Graham, G. G. (2004). The relationship between the supplier of primary care physicians and measures of health. *Eastern Economic Journal 30*(3), 467-487.

Rodts, M. F. (2000). They cared so much. *Orthopaedic Nursing*, *14*(1).

Rodwin V.G. (2003). The health care system French national health insurance: lessons for health reform in the United States. *American Journal of Public Health, 93*(31),31.

Rodwin, V.G. (2002). Comparative health systems. In A. R. Kovner, & S. Jonas, (Eds.) *Health Care Delivery in the United States*. New York, NY: Springer Publishing Company.

Rogers HJ, Swaminathan H. (1993). A comparison of logistic regression and Mantel-Haenszel procedures for detecting differential item functioning, *Applied Psychological Measurement 6* vol. 17: pp. 105 - 116.

Russell, L., Gold, M., Siegel, J., Daniels, N., & Weinstein, M. (1996, October). The role of cost-effectiveness analysis in health and medicine. *The Journal of the American Medical Association 276*(14), 1172-1178.

Ryan, K. E. (1991). The performance of the Mantel-Haenszel procedure across samples and matching criteria. *Journal of Educational Measurement, 28*, 325-337.

Samuelson, R. J. (1999, Apr 12). *The aging of America and the reinvention of retirement. The New Republic*. Retrieved 2 March 2004 from www.elibrary.com

SAS Institute. (1985). *SAS(r) user's guide: Basics, Ver 5*. Cary, NC: Author.

Schenk, R. E. (2004, December). *Definitions of economics*. Retrieved on June 27, 2005 from http://ingrimayne.saintjoe.edu/econ/.

Schlesselman, J. J. (1982).Case control studies: Design, control, analysis. New York: Oxford University Press.

Schroeder, L. D., Sjoquist, D. L., & Stephan, P. E. (1986). *Understanding regression analysis: An introductory guide*. Beverly Hills, CA: Sage.

Sheehy, G. (1996). *New passages :mapping your life across time*. New York: Ballantine Books.

Shen, Y. (2003). Changes in hospital performance after ownership conversions. *Inquiry-Excellus Health Plan, 40*(3), 217.

Smith, P. L. (1979). Splines as a useful and convenient statistical tool. *The American Statistician, 33*, 57-62.

Smits, H.L. (2002). Managed care. In A. R. Kovner, & S. Jonas, (Eds.) *Health Care Delivery in the United States*. New York, NY: Springer Publishing Company.

Social support may reduce impairment in depressed elderly, (2001) *Mental Health Weekly 11*(45), 8.

Spang, H.R., Bazzoli, G.J. & Arnould, R.J. (2001). Hospital mergers and savings for consumers: exploring new evidence. *Health Affairs 20*(4), 150.

Stage, F. K. (1988). University attrition: LISREL with logistic regression for the persistence criterion. *Research in Higher Education, 29*, 343-357.

Strom-Gottfried, K. (1998). Is 'ethical managed care' an oxymoron? *Families in Society: The Journal of Contemporary Human Services 79*(3), 297-307.

Sullivan L. W., & Toby W. (1992). *Medicare Hospital Information: 1992, Technical Supplement*. Washington DC: U.S. Department of Health and Human Services, Healthcare Financing Administration.

Swaminathan, H., & Rogers, H. J. (1990). Detecting differential item functioning using logistic regression procedures. *Journal of Educational Measurement, 27*, 361-370.

Taylor, M. (2001). 10-4, Five-0. *Modern Healthcare 31*(40), 24.

Terenzini, P. T., & Pascarella, E. T. (1977). Voluntary freshman attrition and patterns of social and academic integration in a university: A test of a conceptual model. *Research in Higher Education, 6*, 25-43.

Thakre, R., M.D. (n.d.). *Meconium stained amniotic fluid delivery, to intubate or not?* Retrieved December 30, 2004 from http://www.neoclinic.net/Artcl/msaf.htm.

Theories of Aging. (2002, February). Retrieved 2 March 2004 from http://www.clas.ufl.edu/users/ardelt/Aging/Theories.htm

Thorpe, K.E. & Knickman, J.R. (2002). Financing for health care. A. R. Kovner, & S. Jonas, (Eds.). *Health Care Delivery in the United States*. New York, NY: Springer Publishing Company.

Tishler, A. & Zang, I. (1981a) A maximum likelihood method for piecewise regression models with a continuous dependent variable. *Applied Statistics, 30*, 116-124.

Tishler, A. & Zang, I. (1981b). A new maximum likelihood algorithm for piecewise regression. *Journal of the American Statistical Association, 76*, 980-987.

U.S. Census Bureau. (2002, January). *Statistical brief. Sixty-five plus in the United States. May 1994*. Economist and Statistics Administration, US Department of Commerce. Retrieved 17 November 2003 from: http://www.census.gov/population/socdemo/statbriefs/agebrief.html

U.S. Newswire. (1998, Mar 24). Analysis shows Alzheimer's disease epidemic will bankrupt medicare as boomers reach age of risk. *US Newswire*.

Van Der Linde, C. H. (1998). Clinical supervision in teacher evaluation: A pivotal factor in the quality management of education. *Education 119*(2): 328(1).

Vaul, J. (1998). DRG benchmarking study establishes national coding norms. *Healthcare Financial Management 52*(5), 52-58.

Wagner, J., King, C., & Kesselman Burns, A. (1983). *Diagnosis Related Groups (DRGs) and the Medicare Program: Implications for Medical Technology.* U.S Congress, Office of Technology Assessment (OTA) Report OTA-HA-H-17. July 1983.

Wainess, F. J. (1999l). The ways and means of national health care reform, 1974 and beyond. *Journal of Health Politics, Policy and Law, 24*(12), 305-333.

Walsh, A. (1987). Teaching understanding and interpretation of logit regression. *Teaching Sociology*, *15*, 178-183.

Warner, R. S. (2003, December 13). Radical middle: conservative means toward liberal ends. *The Christian Century*. Retrieved February 23, 2004 from www.elibrary.com.

Watts, D. G. & Bacon, D. W. (1974) Using an hyperbola as a transition model to fit two-regime straight-line data. *Technometrics, 16*, 369-373.

Weaver, Tyler (1992). *Total Quality Management*. ERIC Clearinghouse on Educational Management, Eugene, OR. Sponsoring Agency: Office of Educational Research and Improvement (ED), Washington, DC. Report No: EDO-EA-92-6.

Wegman, E. J. & Wright, I. W. (1983) Splines in statistics, *Journal of the American Statistical Association*, *78*, 351-365.

Weinstein, Milton C. (2003, June 1). We ration health care: Better to do it rationally. *The Washington Post*, B3.

Wennberg, J. E.; Gittelsohn, A.; & Shapiro, N.. (1975) Health Care Delivery in Maine III: Evaluating the Level of Hospital Performance. *The Journal of the Maine Medical Association*. *66*(11), 298-306

Whitener, M., Cox, R., & Maglich, A. (1998). Use of theory to guide nurses in the design of health messages for children. (Development and Aging). Vol. 20, *Advances in Nursing Science, 21*(15).

Wing, K.R. (2003). *The Law and the Public's Health,(6th Ed).* Chicago, IL: Health Administration Press.

Woolf, S H. (1990) Practice guidelines: a new reality in medicine. *Archives of Internal Medicine, 150*, 1811-1818.

WVMI. (2004). *Nursing Facility/Aged Disabled Home and Community Based Waiver Services.* Retrieved March 2, 2004 from http://www.wvmi.org/priorauth/review3.htm

World Health Organization (WHO; 2005) *HIV infection rates decreasing in several countries but global number of people living with HIV continues to rise.* Retrieved January 22, 2007 from http://www.who.int/hiv/epiupdate2005/en/

Abstract

COMPARISON OF LOGISTIC REGRESSION AND MANTEL-HAENSZEL STATISTICAL PROCEDURES TO PREDICT LENGTH OF STAY OF FOUR DIAGNOSIS-RELATED GROUPS

by

REZA ZIAEE

December 2007

Advisor: Dr. Shlomo Sawilowsky

Major: Educational Evaluation and Research

Degree: Doctor of Philosophy

The purpose of this research was to test the appropriateness of logistic regression and Mantel-Haenszel statistical procedures in analyzing health care patient data to examine the relationship between the dependent variable, dichotomized length-of-stay (LOS) and factors that influence the behavior of the LOS. Hospital LOS is a traditional cost driver in health care.

Logistic regression defines the relationship between a dichotomously coded dependent variable and one or more independent variables. This statistical technique frequently is used to assess the influence of independent variables on a dependent variable by evaluating odds-ratios. The odds-ratio demonstrates an index of the likelihood of an expected outcome of the two alternatives given values of the independent variables.

The Mantel-Haenszel statistical procedure examines the relationship between two dichotomous variables using information from 2 x 2 contingency tables. Developed for use in epidemiological research, Mantel-Haenszel was later introduced in the detection of bias in educational research. The Mantel-Haenszel Differential Item Functioning (DIF) is one of the most popular procedures for identifying bias in dichotomous variables.

In this study, the author examined the appropriateness of these two techniques in analyzing

hospital LOS to identify factors influencing the length of hospitalization for two sets of Diagnostic Related Groups, pneumonia and myocardio infarction. The results demonstrated that both techniques effectively identified factors that influenced LOS.

The Mantel-Haenszel procedure was sensitive to one type of differential item functioning and was not designed to detect DIF that has a nonuniform effect across trait levels. The MH procedure evaluates one item at a time requiring multiple analyses which could increase the likelihood of Type 1 errors. While the MH results may be easier to explain to a nonstatistician, the limitations of MH restrict the use of MH DIF in analyzing healthcare LOS data.

Logistic regression (LR) allows simultaneous analysis of independent variables. When interpreting results of the LR procedure, inferences can be drawn regarding which variables should be included in the prediction model and which can be excluded from further analyses. Consequentially, the logistic regression may be better suited for this type of healthcare data analysis.

Autobiographical Statement

Reza Ziaee

Education	2007 – Doctor of Philosophy Wayne State University, Detroit, Michigan Major: Educational Evaluation and Research
	1986 – Master of Science in Engineering University of Michigan, Ann Arbor, Michigan Major: Industrial and Systems Engineering
	1979 – Master of Arts University of Detroit, Detroit, Michigan Major: Economics
	1971 – Bachelor of Science University of Tehran, Tehran, Iran Major: Management
Professional Experience	2004 to 2007 Western Maryland Health System, Cumberland, Maryland System Director, Process Engineering,
	2002-2004 Blue Cross Blue Shield of Michigan, Detroit, Michigan Manager – Process Engineering and Human Performance
	1998 – 2002 Banner Health System, Phoenix, Arizona Senior Management Engineer
	1997 – 1998 Henry Ford Health System, Detroit, Michigan Principal Analyst
	1987 – 1997 Oakwood Healthcare System, Dearborn, Michigan Senior Operations Consultant
Professional Organizations	Healthcare Information System Society
Presentations	Management Engineers, Are you worth your weight in gold. Presented at HIMS National Conference in 1990 Executive Information System, Practical vision through practical Experience. Presented at HIMS National Conference in 1991.