2007

# Medicaid risk adjustment model with diagnosis and pharmacy-based adjusters: Does it work?

Yanen Li
*University of South Florida*

# Medicaid Risk Adjustment Model with Diagnosis and Pharmacy-Based Adjusters: Does it Work?

By

Yanen Li

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
Department of Gerontology
College of Arts and Sciences
University of South Florida

Co-Major Professor: Larry Polivka, Ph.D.
Co-Major Professor: Glenn E. Mitchell II, Ph.D.
David Chiriboga, Ph.D.
William Kearns, Ph.D.
Lawrence Schonfeld, Ph.D.

Date of Approval:
Aug 31, 2007

Keywords: capitation, payment, prescription, aged, disabled

Table of Contents

List of Tables

List of Figures

Medicaid Risk-Adjustment Model with Diagnosis and Pharmacy-Based Adjusters: Does It Work?

Yanen Li

ABSTRACT

National health expenditures will continue to grow faster than nominal gross domestic product (GDP) in the early 21st century (Heffler et al., 2002; Heffler et al., 2005). Increased Medicaid costs have spurred research to find reliable cost-saving methodologies (Kronick et al., 1996). The Medicaid administrations of some states have chosen risk adjustment as a methodology for savings (Tollen et al., 1998; Weiner et al., 1998), since it can reduce the financial burden of health care providers and distribute medical resources more efficiently. This dissertation presents a risk-adjustment model based on two types of health condition adjusters: diagnosis-based HCC adjusters and pharmacy-based RxRisk adjusters. HCC adjusters were developed from different diagnostic categories from inpatient, outpatient and long-term care data. RxRisk adjusters included diseases inferable from prescription drug usage. The underlying assumption is that using both types of health condition adjusters, rather than relying on either diagnosis-based adjusters or pharmacy-based adjusters alone can help increase predictive power and lower Medicaid's risk of reimbursing inflated medical costs for its beneficiaries. The population in this study consisted of all disabled and aged Florida adults who were eligible for Florida's Medicaid program in state fiscal year (SFY) 2002-03 and state fiscal year 2003-04. The population was broken down into two subpopulations: disabled Medicaid beneficiaries aged 64 and under and beneficiaries aged 65 or over.

The proposed regression model includes diagnostic and pharmacy-based adjusters, and this dissertation compares the proposed model with models based solely on pharmacy- or diagnosis-based adjusters.

The results presented in this dissertation demonstrate the proposed model has higher predictive power than the diagnosis-based HCC model and the pharmacy-based RxRisk model for the overall population and the subpopulations in this study. Risk-adjustment models using diagnostic and prescription drug information have higher predictive power and decrease the possibility of inappropriate gaming of the Medicaid capitation payment system.

# Chapter One

## Introduction

National health expenditures continue to grow faster than nominal gross domestic product (GDP). Some researchers (Heffler et al., 2002) have projected that national health expenditures will be $2,815.8 billion in 2011 and will constitute approximately 17 percent of GDP. This compares with 13.2 percent in 2000. Heffler and colleagues (2005) also predicted that national health expenditures would comprise 18.7 percent of the GDP in 2014.

Health spending can be divided into two categories: the public and private sectors (Levit et al., 2003). Public-sector health spending includes Medicare and Medicaid expenditures. Medicare is a federal program that pays for hospitalization, physician services, short-term nursing home care, and outpatient care expenses for the aged. It also pays for medical care for the permanently disabled. In the past century, Medicare could not cover outpatient prescription drug costs. However, Medicare beneficiaries began receiving Medicare drug benefits under the Medicare Part D program on January 1, 2006. Medicaid is funded by the federal and state governments and provides health care and health-related services to eligible low-income individuals and people with disabilities. Private health insurance and out-of-pocket expenditures constitute the private sector of national health spending. The growth of private spending was expected to slow down sharply in 2006 when the Medicare drug benefit plan was introduced. However, the growth of public-sector health expenditures could accelerate because Medicare started covering some

prescription drug costs which were paid by beneficiaries prior to the introduction of the Medicare Part D program (Heffler et al., 2005).

The growth of Medicaid expenditures may also accelerate for another reason; a slowdown in the economy may lead to more Medicaid enrollees, since Medicaid serves low-income individuals (Weil, 2003). Other factors associated with higher Medicaid costs include increased health care costs and costs associated with advanced medical technology. Medicare costs are less influenced by the economy since all people aged 65 or older are eligible for Medicare.

The acceleration of Medicaid spending has encouraged research into cost-containment methods. Some methods emphasize the control of prescription drug costs: drug formulary, a price ceiling on drugs, prescription-renewal limitations, dollar limits per prescription, and co-payment plans (Moore and Newman, 1993). Drug formularies are lists of drugs reimbursable for certain diagnoses. State Medicaid offices can establish "pharmacy and therapeutics committees" to evaluate drug formularies. The pharmacy and therapeutics committee approves drugs for Medicaid reimbursement. Moore and Newman (1993) evaluated the drug formulary methodology and suggested the restriction of prescription drug usage may yield a lower level of health benefits and fail to lower medical costs significantly. Lower drug usage may be associated with a surge in alternative therapy use, eventually increasing total health care costs. Since prescription drug costs are part of total medical expenditures, including physician and inpatient services, simply controlling prescription drug benefit alone may inadvertently result in limited and incomplete cost savings (Dranove, 1989; Moore and Newman, 1993).

Another cost-containment strategy is reducing payment rates for some physician and outpatient services. This strategy may shift the financial burden to health care providers and lower medical costs inefficiently.

Increased Medicaid costs have stimulated health care researchers and policymakers to find more feasible and reliable cost-saving methodologies. The Medicaid administrators of many states have chosen risk adjustment because it can reduce the financial burden on health care providers and distribute medical resources more efficiently. Risk-adjustment strategies use the health status of patients to predict future medical costs. Researchers review diagnoses or prescription drug usage to infer health status of beneficiaries and build risk-adjustment models based on inferred health status and other variables. Such models have been reported to have higher predictive power than models based solely on demographic factors such as age, gender, and geographic location for projecting medical expenditures (Ellis et al., 1996; Pope et al., 2000; Riley, 2000). Health plans tend to enroll healthier patients under capitation systems based on demographic factors, which provides them with greater profit margins, since predicted costs for healthier and sicker patients are equal under demographic models. Risk-adjustment models can more equitably reimburse health plans having a large proportion of sick enrollees and decrease incentives to enroll the healthiest enrollees (Ash et al., 2000; Ettner et al., 2000; Pope et al., 2000).

Some risk-adjustment models may establish perverse incentives (Fishman et al., 2003; Gilmer et al., 2001; Pope et al., 2000). For example, models based on primary inpatient diagnosis encourage more hospitalization (Pope et al., 2000). Outpatient services that tend to be less expensive than corresponding inpatient services might in some cases be inadvertently replaced with inpatient services, and models based on prescriptions have been shown to actually increase prescription drug usage (Fishman et al., 2003; Gilmer et al., 2001).

State Medicaid offices have been actively seeking a capitation payment system based on a risk-adjustment model having high predictive power and low

potential for gaming. This dissertation presents a model based on multiple types of health status predictors, which are assumed to increase predictive power and lower the risk of gaming by providers.

## Study Rationale

In the past two decades, researchers have tried to establish a risk-adjustment model with a high predictive capacity for both Medicare and Medicaid capitation payment systems. Models have been built on diagnoses (Ash et al., 2000; Ellis et al., 1996; Fishman et al., 2003; Kronick et al., 2000; McCall and Korb, 1998; Pope et al., 2000) and on prescription drug usage (Fishman et al., 2003; Meenan et al., 2003; Sales et al., 2003; Sloan et al., 2003). Both models have advantages and disadvantages, and these are discussed in the literature review section. The purpose of this dissertation is to build a new predictive model for the Florida Medicaid capitation payment system that draws on both diagnoses and prescription drug usage from patient claims files.

## Review of the Literature

Many studies have been undertaken to examine risk-adjustment methodologies for the Medicare capitation payment system. Although the goal of this dissertation is to build a new risk-adjustment model for the Medicaid capitation payment system, the literature on risk adjustment for Medicare capitation payment is reviewed since Medicare and Medicaid cover similar medical costs of their beneficiaries.

Starting in the 1980s, Medicare beneficiaries could choose managed care plans as an alternative to the fee-for-service option. The Health Care Financing Administration (HCFA) compensates managed care organizations by a predetermined payment amount for each Medicare health maintenance organization (HMO) enrollee, according to capitation rates set in advance. Before risk-adjustment models, the

adjusted average per capita cost (AAPCC) model was launched in 1985. The capitation payment of the AAPCC model is calculated by the formula: Payment = (0.95)* (County Per Capita Costs/Average County Demographic Score)*Enrollee Demographic Score. Under the AAPCC model, health plans are reimbursed according to a predetermined amount predicted by Medicaid beneficiaries' demographic characteristics only (age, gender, welfare status, and institutional status); the health status of Medicare beneficiaries is not used as a predictor in this model. Some researchers reported that Medicare HMOs could cherry-pick healthier beneficiaries from the Medicare population (Brown, 1988; Brown et al., 1993; Hill & Brown, 1990; Lichtenstein et al., 1991; Mello et al., 2003; Morgan et al., 1997) since sicker beneficiaries may have higher medical costs than predicted based on demographic models. Medicare HMO beneficiaries also have lower mortality rates than Medicare fee-for-service (FFS) beneficiaries (Brown, 1988; Brown and Langwell, 1988). The AAPCC model explained only 1 percent of the variance in total annual medical expenditures of Medicare beneficiaries (Ash et al., 1989; Ellis et al., 1996; Newhouse, 1986). Some researchers suggest an ideal risk-adjustment model could predict 20 to 25 percent of the variance in medical costs (Newhouse et al., 1989). Since the AAPCC model has much lower predictive power than an ideal model, it does not have enough capacity to predict medical costs accurately. Predicted medical costs can be significantly higher than actual medical costs for healthier patients and lower for sicker patients.

Several researchers discuss models based on other risk adjusters; self-reported health status (Fowles et al., 1996; Hornbrook and Goodman, 1996), perceived health status (Epstein and Cumella, 1988; Gruenberg et al., 1996; Hornbrook and Goodman, 1996) or functional status (Epstein and Cumella, 1988; Hornbrook and Goodman, 1996; Lichtenstein and Thomas, 1987) has been used to

build risk-adjustment models. However, some researchers argue that it is time-consuming and costly to collect self-report data (Pacala et al., 2003).

Compared to self-reported health and functional status, diagnoses in claims files have advantages as medical cost predictors: (1) they are easy to obtain from hospital records or physicians' offices; (2) they are made by physicians, and hence more objective than self-reported predictors; (3) the collection costs are lower since diagnoses are extant in hospitalization and outpatient files. Comparisons of model performance have yielded varying results: some researchers report diagnosis-based models have higher predictive power than models based on self-reported health status (Fowles et al., 1997) while opposite results were reported by Pacala and colleagues (2003). Generally, when diagnoses are available, researchers recommend using diagnosis-based models (Fowles et al., 1997; Hornbrook and Goodman, 1996).

A new Medicare capitation payment system as well as health status risk adjusters were mandated under the Balanced Budget Act of 1997; and the new system went into effect in 2000. Since diagnosis-based models have higher performance, the Health Care Financing Administration (HCFA) incorporates diagnostic information into the Medicare capitation payment system, and has funded studies on different risk-adjustment methods based on diagnostic information. These methods include the Principal Inpatients Diagnostic Cost Group (PIPDCG) model based solely on inpatient data, the multi-condition Hierarchical Condition Category (HCC) model based on inpatient and outpatient diagnoses, and the pharmacy-based RxRisk model.

The demographic AAPCC model was initially incorporated into the Medicare capitation payment system. Since the AAPCC model lacks predictive power (Ash et al., 1989; Newhouse, 1986; Pope et al., 2000) and the ability to prevent health plans gaming the Medicare capitation payment system (Brown et al., 1986; 1988;

1993; Brown and Langwell, 1987; Eggers and Prihoda, 1982; Lubitz and Prihoda, 1984; Beebe, Lubitz, and Eggers, 1985), Congress passed the Balanced Budget Act of 1997 and requested the HCFA to replace the AAPCC model with a risk-adjustment model before January 1, 2000. The PIPDCG model was chosen to be part of the Medicare capitated payment system in 2000. The PIPDCG model was developed by researchers at Boston University, Brandeis University, Harvard University, and Health Economics Research, Inc. and it was selected because it could be developed in a short time frame. As a model based solely on inpatient data, the PIPDCG model is more feasible than other models because hospitalization data are easily obtained from health plans and health care providers, and hospitalization data are associated with the severity of clinical diseases and symptoms (Pope et al., 2000). Data of short hospital stay (less or equal to one day) are excluded from the PIPDCG model because Medicare beneficiaries can be inappropriately hospitalized for higher Medicare reimbursement by health plans.

Demographic variables of the PIPDCG model are similar to the variables of the AAPCC model (Pope et al., 2000). In the AAPCC model, demographic adjusters are: age, gender, welfare status, and institutional status. Beneficiaries are divided into 20 categories, 10 for men and the other 10 for women. In the PIPDCG model, the only change is that the "age 85 or over" group in the AAPCC model was removed and the "age 85-89," "age 90-94," and "age 95 or over" categories were added to the PIPDCG model. There are 24 age-gender categories in the PIPDCG model. Medicaid status is included in the PIPDCG model, and institutional status is removed because it was found to have no significant effect on annual medical expenditures in the PIPDCG model (Pope et al., 2000).

The "Originally disabled" variable is another demographic adjuster introduced in the PIPDCG model. Disabled Medicare beneficiaries can be enrolled in the Medicare

program when they are younger than 65 years old under specific conditions. Their medical costs are significantly higher than other beneficiaries, and an "originally disabled" adjuster may predict an increase in medical costs after controlling for other variables (Pope et al., 2000). Working-age status is another demographic adjuster used in the PIPDCG model, and it refers to Medicare beneficiaries who are employed and offered private insurance when they receive Medicare benefits. When Medicare beneficiaries qualify for "working-age" status, 21 percent of their medical costs are paid by Medicare.

Researchers (Ash et al., 1989; Ellis and Ash, 1995; Ellis et al., 1996; Pope et al., 2000) developed the PIPDCG model by using the primary diagnosis in hospitalization data and categorizing diagnoses into certain diagnostic cost groups (DCGs). Since diagnoses in hospitalization data were coded according to the International Classification of Diseases, 9$^{th}$ Revision, Clinical Modification (ICD-9-CM), the DCGs are built on 14,000 ICD-9-CM codes (Ash et al., 1989; Ellis and Ash, 1995). Ellis and Ash (1996) classified 104 groups of diagnoses into 143 principal inpatient DXGROUPs, and these principal inpatient DXGROUPs were regrouped into 25 PIPDCG categories. Pope and colleagues (2000) made refinements by categorizing the ICD-9-CM codes (more than 15,000) into 172 Principal Inpatient Diagnostic Groups (PIPDxGs), and created 26 PIPDCG categories. A new PIPDCG category added was PIPDCG 4, which is "no or excluded inpatients admissions, ectopic pregnancy, miscarriage/terminated pregnancy, completed pregnancy with major complications, completed pregnancy with complications, completed pregnancy without complications (normal delivery)". Each PIPDCG category represented a group of diseases that were clinically related and required similar payments. Rare diseases were excluded when researchers developed them from a five-percent sample of Medicare's FFS enrollees. Ranks of PIPDCG categories were positively associated with

medical costs after adjustment of demographic predictors. For example, PIPDCG 29, which is the "HIV/AIDS, bloods, lymphatic cancer/neoplasms" category, is the reference patient group with the highest average medical expenditures ($30,456) in 1996 (Pope et al., 2000). PIPDCG 4 was referred to as the group of people who were not hospitalized or who were hospitalized without significant increases in medical expenditures after adjusting for demographics.

The PIPDCG model included demographic and 16 PIPDCG variables (Pope et al., 2000). As a prospective risk-adjustment model, the PIPDCG model has more predictive power than the AAPCC model (Ellis et al., 1996; Pope et al., 2000). Ellis and colleagues (1996) report that the $R^2$ value of the PIPDCG model was 5.5 percent, and Pope and colleagues (2000) report an $R^2$ value of 6.2 percent, compared with the lower value of the AAPCC model (1 percent). Consistent results can be found in other studies (McCall and Korb, 1998; Robinson and Karon, 2000; Temkin-Greener, 2001). Since Pope and colleagues (2000) re-code the PIPDXGs and make some refinements to the PIPDCG model, 6.2 percent may be closer to the predictive power of this model. Moreover, Pope and colleagues (2000) use a five-percent sample of Medicare's FFS enrollees in 1995 and 1996, consistent with the 1991 and 1992 sample (Ellis et al., 1996) used to establish the PIPDCG model. Other studies (McCall and Korb, 1998; Robinson and Karon, 2000; Temkin-Greener, 2001) have produced similar results, while samples drawn from these studies have studied different groups of Medicare beneficiaries.

Predictive ratio is another measure used to evaluate model performance of risk-adjustment models. The calculation of predictive ratio is based on the formula: predictive ratio = average medical costs predicted by risk-adjustment models divided by the average actual medical costs. Since we evaluate risk-adjustment models by their capacity to predict actual medical costs, a predictive ratio close to 1 indicates

better model performance than a value that differs significantly from 1. Consistent findings (Ellis et al., 1996; Health Care Financing Administration, 1999; Pope et al., 2000; Riley, 2000) show that the PIPDCG model has predictive ratios closer to 1 in different subgroups of Medicare beneficiaries than the AAPCC model. For example, Pope and colleagues (2000) report that the predictive ratios are 2.57, 1.88, 1.35, 0.96, and 0.47, respectively, for five subgroups of Medicare beneficiaries according to their medical costs in the AAPCC model, compared with 2.09, 1.54, 1.10, 0.84, and 0.75 in the PIPDCG model. The PIPDCG model has a better predictive ratio in all subgroups except the group whose medical costs fall within a range of 60 to 80 percent of the highest annual medical expense of Medicare beneficiaries. The PIPDCG model also outperforms the others for all chronic conditions and for frail elders with functional status difficulties. Ellis et al. (1996) and Riley (2000) found the PIPDCG model has a better predictive ratio in subgroups of Medicare beneficiaries who have impairment of 1-2, 3-4, and 5-6 ADLs. These results were later confirmed in a study by Pope and colleagues (2000).

As the first diagnosis-based model incorporated into the Medicare capitation payment system, the PIPDCG model has demonstrated more predictive accuracy than the AAPCC model. However, the $R^2$ value is far below the ideal value (Pope et al., 2000). Other models have been found to be more competitive in predicting the variance of future medical costs. For example, the HCC model $R^2$ value of 8.08 percent was found to be better than the PIPDCG model's $R^2$ value of 5.53 percent (Ellis et al., 1996).

The PIPDCG model has been criticized by researchers (Miller and Luft, 1997; Pope et al., 2000; Health Care Financing Administration, 1999) for several shortcomings: First, the data source for the development of the PIPDCG model is inpatient data only. Both ambulatory and outpatient data can have significant effects

on the prediction of future medical costs but these data are not used in the PIPDCG model. Health plans can be more profitable by hospitalizing Medicare HMO enrollees who actually only need outpatient or ambulatory services because inpatient medical care is much more expensive than outpatient services. Second, only the primary diagnosis is used to evaluate patients' health condition in hospitalization data, while other diagnoses associated with significant increases in future medical costs are not counted. The HCC model incorporates more information about diagnoses in the risk-adjustment model.

Researchers (Ellis et al., 1996; Ash et al., 1998; Pope et al., 1998) developed and examined the HCC model in the late 1990s. The HCC model, as a type of risk-adjustment model that can use information from outpatient and long-term care data, has more predictive power than the PIPDCG model based solely on inpatient data (Ellis et al., 1996).

Both the HCC model and the PIPDCG model have been developed from 15,000 ICD-9-CM codes because ICD-9-CM codes are generally used in all types of data from claims submitted by hospitals and physicians' offices. Compared with 172 PIPDxGs in the PIPDCG model, in the HCC model there are 543 DxGROUPs clustered from ICD-9-CM codes. Each DxGROUP refers to a group of clinically related diseases and symptoms. Furthermore, 543 DxGROUPs are classified into 118 Condition Categories (CC) and each CC includes DxGROUPs with similar predicted medical costs. Advantages of CCs over PIPDCG categories are: (1) there are more CCs in the HCC model than PIPDCG categories in the PIPDCG model, hence the HCC model sets a clearer criterion for coding diagnoses from raw data. Since the PIPDCG model has 26 PIPDCG categories only making it easier for health plans to intentionally change a lower PIPDCG rank to a higher PIPDCG rank linked with higher medical costs thus increasing profits. The PIPDCG model has obscure classification criteria for some

medical disorders because of its few categories. The categorization of CCs is more likely to be accurate than is the case for PIPDCG categories preventing a health plan from gaming the capitation payment system by manipulating diagnostic coding. (2) CCs are specifically designed for practical use on all types of data and therefore are more appropriate for this purpose. Since ambulatory and outpatient data are submitted by hospitals and physicians' offices, the HCC model has the advantage of using all available information. (3) The HCC model counts multiple diagnoses, while the PIPDCG model counts only the primary diagnosis. Under the capitation payment system based on the PIPDCG model, health plans that have Medicare beneficiaries with multiple health problems are significantly underpaid because only treatment for the primary medical problem is reimbursed in the PIPDCG model. Pope and colleagues (2000) report that the predictive ratios for Medicare beneficiaries with two admissions and beneficiaries with three or more admissions are 0.91 and 0.69 respectively. However, the predictive ratios for Medicare beneficiaries with zero and beneficiaries with one admission were 1.07 and 1.02 respectively, which over-predicts the true value. Although patients can be hospitalized several times with the same diagnosis, it is more likely patients are admitted to hospital with different diagnoses. Hence, evidence that health plans enrolling patients with multiple hospital admissions are underpaid indicates that the PIPDCG model underestimates actual medical costs of patients with multiple diagnoses. These results are consistent with those found in other studies (Ellis et al., 1996; Health Care Financing Administration, 1999).

Demographic adjusters of the HCC model are slightly different from those of the PIPDCG model. Age is clustered into 15 categories: 0-5 years, 6-12 years, 13-17 years, 18-24 years, 25-34 years, 35-44 years, 45-54 years, 55-64 years, 65-69 years, 70-74 years, 75-79 years, 80-84 years, 85-89 years, 90-94 years, and 95

years or over (Ash et al., 2000). In the PIPDCG model, age categories are 0-34 years, 35-44 years, 45-54 years, 55-59 years, 60-64 years, 65-69 years, 70-74 years, 75-79 years, 80-84 years, 85-89 years, 90-94 years, and 95 years or over (Pope et al., 2000). The PIPDCG model divides the 55-64 years category into two sub-categories: the 55-59 years category and the 60-64 years category. The "Medicaid status" variable is added to the HCC model. Unlike the PIPDCG model, there are five eligibility categories in the HCC model: blind/disabled, other medical problems, poverty, pregnant women, and others.

As health status adjusters of the HCC model, CCs play a different role than the PIPDCG categories do. Patients can be assigned to multiple CCs, but only one PIPDCG category. CCs are divided into subgroups and a hierarchy imposed within each subgroup. For example, the subgroup neoplasm includes seven CCs: CC6 (high-cost cancer), CC7 (moderate-cost cancer), CC8 (lower-cost cancers/tumors), CC9 (carcinoma in situ), CC10 (uncertain neoplasm), CC11 (skin cancer, except melanoma), and CC12 (benign neoplasm). CC6 ranks the highest in this subgroup, and only CC6 is counted if a patient can be categorized into CC6 and another CC in the neoplasm subgroup. Only the CC with the highest rank is counted when a patient has multiple CCs of the same subgroup according to the HCC hierarchy system. However, CCs from different subgroups can be accumulated at the same time. For example, CC1 (HIV/AIDS), CC6 (high-cost cancer), and CC19 (liver disease) can be assigned to the same patient because these CCs are in different subgroups. Meenan and colleagues (2003) refine the HCC model that has 34 demographic adjusters and 31 CC adjusters, as another version.

The $R^2$ value of the HCC model has a broad range – from 8.1 percent to 15.9 percent (Ash et al., 2000; Ellis et al., 1996; Fishman et al., 2003; Kronick et al., 2000; McCall and Korb, 1998; Sales et al., 2003). The samples in these studies are

different, however, the HCC model was found to have the greatest predictive power when it was compared to the AAPCC model, the PIPDCG model, and the pharmacy-based RxRisk model in previous studies (Fishman et al., 2003; Sales et al., 2003). The investigators (Ellis et al., 1996; Kronick et al., 2000; McCall and Korb, 1998) conclude that the HCC model has a predictive ratio closer to 1.0 than the AAPCC model, the PIPDCG model, and the RxRisk model. Ellis and colleagues (1996) report that the predictive ratios are 1.30, 1.24, 1.14, 0.99, and 0.85 for the five subgroups of Medicare beneficiaries in the HCC model, while the predictive ratios are 1.92, 1.37, 1.01, 0.78, and 0.85 in the PIPDCG model. The study is based on a national sample of Medicare beneficiaries, and the HCC model has better predictive ratios than the PIPDCG model in this study.

A disadvantage of the HCC model is that health plans can abuse the Medicare capitation payment system (Ash et al., 2000; Ellis et al., 1996). Health plans can replace lower-ranked CCs with higher ranked ones. Since multiple CCs can be counted for the same patient, health plans can inflate CC counts and receive more reimbursements.

As a means of correcting some of the problems evident in diagnosis-based models, several pharmacy-based RxRisk models have been developed in the past two decades. Two of them are based on the Chronic Disease Score (Von et al., 1992) designed by the Center for Health Studies, Group Health Cooperative of Puget Sound (GHC). Prescription drugs in current medical practice are reviewed by a panel of physicians and health services researchers, and CDS categories are created by researchers. Since the original CDS model was related only to adult chronic conditions, a pediatric CDS model has also been developed by Fishman and Shay (1999). Based on the CDS and pediatric CDS model, the RxRisk model and the RxRisk-V model have been developed and discussed in several studies (Gilmer et al.,

2001; Fishman et al., 2003; Meenan et al., 2003; Sales et al., 2003; Sloan, 2003).

The major difference between the RxRisk model (Fishman et al., 2003; Meenan et al., 2003; Sales et al., 2003) and the RxRisk-V model (Sales et al., 2003; Sloan et al., 2003) is the method of categorization. The RxRisk model consists of 28 adult RxRisk factors and 24 pediatric RxRisk factors. Sloan and colleagues (2003) review RxRisk categories and establish 45 RxRisk-V categories. Two RxRisk categories are replaced by four RxRisk-V categories: anticoagulation, antiplatelet agents, ischemic heart disease/angina, and congestive heart failure/hypertension.

Diagnosis-based models, compared with pharmacy-based models, have been reviewed by investigators (Gilmer et al., 2001; Fishman et al., 2003; Sales et al., 2003; Sloan et al., 2003). Three disadvantages of diagnosis-based models have been identified: (1) Some health plans and HMOs do not collect diagnostic data routinely while pharmacy data is recorded in all claims files. (2) Reliability of diagnoses data can be questionable because of faulty recording, while pharmacy data are more reliable. (3) Health plans can become more profitable by gaming capitation payment systems. There is a general consensus that pharmacy-based models decrease the possibility of gaming because drugs used in clinical practice must be correctly recorded by hospitals and physicians' offices.

Diagnosis-based models, however, do have certain advantages over pharmacy-based models (Fishman et al., 2003). For example, the HCC model has more variables than the RxRisk model and the RxRisk-V model, and can identify more diseases than the RxRisk model and the RxRisk-V model. A shortcoming of RxRisk and RxRisk-V factors is that they have been developed from outpatient pharmacy data, and designed for chronic conditions only.

Age and gender are major demographic variables in the RxRisk model and the RxRisk-V model. On the other hand, the "Medicaid status" variable is not included in

pharmacy-based models, unlike the HCC model and the PIPDCG model. The critical outcome of course is the proportion of variance accounted for. The $R^2$ value of the RxRisk model and the RxRisk-V model has been reported by different researchers (Fishman et al., 2003; Meenan et al., 2003; Sales et al., 2003; Sloan et al., 2003); the $R^2$ value for the RxRisk model ranges from 7.7 percent (Fishman et al., 2003) to 11.1 percent (Sales et al., 2003) while the $R^2$ values for the RxRisk-V model ranges from 10.0 percent (Fishman et al., 2003) to 12.2 percent (Sloan et al., 2003). Sloan and colleagues (2003) compare the model performance of the RxRisk model with that of the RxRisk-V model, and report that the RxRisk-V model can explain more variance of medical expenses. The HCC model has outperformed pharmacy-based models in earlier studies (Fishman et al., 2003; Sales et al., 2003).

## Hypothesis and Research Questions

In the previous studies, researchers typically built a risk adjustment model based on one class of health condition variables. Some models used HCC adjusters that utilize information on diagnoses from inpatient and outpatient files. The diagnosis-based HCC model is the most common risk adjustment model implemented in the current Medicare and Medicaid capitation payment systems. However, there are shortfalls of diagnosis-based models in clinical practice. First, diagnostic variables are just one of several factors which have effects on medical costs. Consider that two patients with the same bacterial infection may have different medical costs since one is allergic to inexpensive antibiotics and requires more expensive medicine. Consequently, the patient with the allergy will have much higher medical costs than the patient who is treated with less expensive medications. Additionally, some patients may need costly medicine while others with the same disease need only generic drugs. Obviously, a risk adjustment model with only diagnosis-based adjusters will underpredict medical costs of patients with special

needs. Second, diagnoses can be easily manipulated by health plans since diagnoses are vague even after patients have been discharged from hospitals.

Pharmacy-based models were developed as possible replacements of diagnosis-based models. However, pharmaceutical adjusters are only associated with prescription drug costs among total medical costs. Hence, it is not a good predictor for overall medical costs. However, pharmaceutical adjusters can provide additional information if they are incorporated into a risk adjustment model based on diagnostic adjusters. As mentioned in the earlier context, some patients have to be treated with more expensive medicine than other patients in the same disease group. Utilizing information on prescription drug usage can predict medical costs for these patients more accurately. Moreover, the addition variables in risk adjustment models may increase predictive power.

The hypothesis of this dissertation is that addition of pharmacy-based risk adjusters in the HCC model can improve both predictive power and performance in practical implementation. A good risk adjustment model must be easily implemented and decrease the likelihood of gaming to the maximum degree. It is assumed the proposed model having both HCC and RxRisk adjusters will predict medical costs more accurately. This model can avoid some uncertainty associated with the HCC model since prescription drug use provides more information about patients' health condition. The research questions this dissertation seeks to answer are how RxRisk adjusters should be entered into the HCC model and how to statistically and clinically evaluate model performance. The hypothesis that the proposed model can outperform the HCC model and the RxRisk model is tested and the research questions posed above are answered.

## Chapter Two

## Population

The data for this dissertation was drawn from the administrative claims generated by 2.3 million Medicaid beneficiaries eligible for the Florida Medicaid program during State Fiscal Year (SFY) 2002-03 and 2003-04. A Florida SFY begins in July and ends the following June. SFY 2002-03 is the base year data, which was used to obtain information for demographic and health status variables. The SFY 2003-04 data is the year 2 data and includes information about medical costs and eligible months.

Ash and colleagues (2000) subdivide Medicaid enrollees into five groups: (1) the blind and disabled; (2) beneficiaries with other medical problems; (3) pregnant women and children; (4) beneficiaries with poverty-related entitlement; (5) and others; most Medicaid beneficiaries age 65 or over are also eligible for Medicare. For the purposes of the research reported here, two subpopulations were identified from the SFY 2002-03 Florida Medicaid data. The first subpopulation is Medicaid beneficiaries who are age 65 and older. The second subpopulation is disabled adult Medicaid beneficiaries who are older than 18 and younger than 65 years old. There were 242,193 enrollees in the aged adult subpopulation and 144,846 enrollees in the disabled adult subpopulation. The SFY 2002-03 Florida Medicaid data is referred to as the total Medicaid population in this dissertation.

## Methods

The outcome variable is the Per Member Per Month (PMPM) cost for all services reimbursed by Medicaid for each beneficiary during SFY 2003-04. In order

to calculate PMPM, all claims for each Medicaid beneficiary were summed to obtain the annual total medical cost. This was divided by the Medicaid eligible months for the same beneficiary during SFY 2003-04. The PMPM cost includes medical expenses for hospitalization, prescription drugs, outpatient services, and other acute and long-term care services. Base year data and year 2 data were merged to build multiple linear regression models. The combined data is used to establish model structure and measure model performance. The form of regression models is:

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \ldots + \beta_n X_n ,$$

where Y represents the PMPM cost for each Medicaid beneficiary during SFY 2003-04 and $X_1$ - $X_n$ represent the various demographic and health condition variables.

The HCC model and the PIPDCG model are both diagnosis-based risk-adjustment models. The HCC model can include multiple diagnoses from different data resources, while the PIPDCG model uses only the primary diagnoses during hospitalization. Since the HCC model uses more information about diagnoses and has more predictive power, the risk-adjustment model presented in this dissertation uses the CC categories in the HCC model as the foundation for its diagnostic variables.

CC categories are developed from Diagnostic Groups (DxGroups), which are created from more than 15,000 International Classification of Diseases, 9$^{th}$ Revision, Clinical modification (ICD-9-CM) codes. There are 543 DxGroups, and each DxGroup includes a cluster of similar clinical diseases. The classification of DxGroups is made by researchers according to clinical criteria and physicians' experience. CC categories are created to consist of DxGroups with similar medical costs. Hence, diseases in a CC category usually have clinical homogeneity and similar medical costs. In this dissertation, 70 CC categories are included in the HCC model and the proposed model.

Only diagnostic codes in claims files were used to infer diagnoses for this

study. Medicaid beneficiaries can have multiple diagnoses, and hence can have multiple CCs. In the proposed risk-adjustment model, the diagnosis-based health condition variables consist only of HCC variables based on CC categories. Ash and colleagues (2000) examined the list of CC categories in order to avoid "DCG creep," a coding practice that health plans employ to intentionally change diagnoses with lower predicted capitation payment to diagnoses having higher payment. Some CC categories are excluded and hierarchies are imposed to decrease frequency of vague coding and intentional coding proliferation (Ash et al., 2000). Every HCC variable in the proposed model corresponds to a CC category, which is clearly distinguished from other CC categories. The list of HCC variables can be seen in Table 1. Hierarchies contain only the medical condition associated with the highest medical costs in risk-adjustment models when patients have several clinically related medical problems. For example, CC7 is "moderate–cost cancer" and has a higher rank than CC8 (lower–cost cancer) in the HCC hierarchy system. Hence, only moderate–cost cancer is used to predict medical expense if a patient has both moderate–cost cancer and lower–cost cancer. When a patient has several medical problems in the same hierarchy, only the highest-ranked disease is used for the prediction. For example, only CC7 will be retained in risk-adjustment models when a patient has CC7, CC8, and CC9 (carcinoma in situ). The HCC hierarchy system can be seen in Table 2.

After imposing hierarchies in risk-adjustment models, the risk-adjustment model assumes the possibility of variable inflation and inappropriate coding practice can be well-controlled. When a patient has multiple medical problems corresponding to several CCs in the same hierarchy, only the highest-ranked CC is used in risk-adjustment models. CCs with smaller numbers are higher-ranked CCs. For example, the order in the neoplasm1 hierarchy is: CC7 > CC8 > CC9 > CC10.

Table 1: The List of HCC Variables

| HCC Number | Disease Categories |
|---|---|
| HCC1 | HIV/AIDS |
| HCC2 | Septicemia/Shock |
| HCC5 | Opportunistic Infections |
| HCC7 | Metastatic Cancer and Acute Leukemia |
| HCC8 | Lung, Upper Digestive Tract, and Other Severe Cancers |
| HCC9 | Lymphatic, Head and Neck, Brain, and Other Major Cancers |
| HCC10 | Breast, Prostate, Colorectal and Other Cancers and Tumors |
| HCC15 | Diabetes with Renal or Peripheral Circulatory Manifestation |
| HCC16 | Diabetes with Neurologic or Other Specified Manifestation |
| HCC17 | Diabetes with Acute Complications |
| HCC18 | Diabetes with Ophthalmologic or Unspecified Manifestation |
| HCC19 | Diabetes without Complication |
| HCC21 | Protein-Calorie Malnutrition |
| HCC25 | End-Stage Liver Disease |
| HCC26 | Cirrhosis of Liver |
| HCC27 | Chronic Hepatitis |
| HCC31 | Intestinal Obstruction/Perforation |
| HCC32 | Pancreatic Disease |
| HCC33 | Inflammatory Bowel Disease |
| HCC37 | Bone/Joint/Muscle Infections/Necrosis |
| HCC38 | Rheumatoid Arthritis and Inflammatory Connective Tissue Disease |
| HCC44 | Severe Hematological Disorders |
| HCC45 | Disorders of Immunity |
| HCC51 | Drug/Alcohol Psychosis |
| HCC52 | Drug/Alcohol Dependence |
| HCC54 | Schizophrenia |
| HCC55 | Major Depressive, Bipolar, and Paranoid Disorders |
| HCC67 | Quadriplegia, Other Extensive Paralysis |
| HCC68 | Paraplegia |
| HCC69 | Spinal Cord Disorders/Injuries |
| HCC70 | Muscular Dystrophy |
| HCC71 | Polyneuropathy |
| HCC72 | Multiple Sclerosis |
| HCC73 | Parkinson's and Huntington's Diseases |
| HCC74 | Seizure Disorders and Convulsions |
| HCC75 | Coma, Brain Compression/Anoxic Damage |
| HCC77 | Respirator Dependence/Tracheotomy Status |
| HCC78 | Respiratory Arrest |
| HCC79 | Cardio-Respiratory Failure and Shock |
| HCC80 | Congestive Heart Failure |
| HCC81 | Acute Myocardial Infarction |
| HCC82 | Unstable Angina and Other Acute Ischemic Heart Disease |
| HCC83 | Angina Pectoris/Old Myocardial Infarction |
| HCC92 | Specified Heart Arrhythmias |
| HCC95 | Cerebral Hemorrhage |
| HCC96 | Ischemic or Unspecified Stroke |
| HCC100 | Hemiplegia/Hemiparesis |
| HCC101 | Cerebral Palsy and Other Paralytic Syndromes |
| HCC104 | Vascular Disease with Complications |

| | |
|---|---|
| HCC105 | Vascular Disease |
| HCC107 | Cystic Fibrosis |
| HCC108 | Chronic Obstructive Pulmonary Disease |
| HCC111 | Aspiration and Specified Bacterial Pneumonias |
| HCC112 | Pneumococcal Pneumonia, Emphysema, Lung Abscess |
| HCC119 | Proliferative Diabetic Retinopathy and Vitreous Hemorrhage |
| HCC130 | Dialysis Status |
| HCC131 | Renal Failure |
| HCC132 | Nephritis |
| HCC148 | Decubitus Ulcer of Skin |
| HCC149 | Chronic Ulcer of Skin, Except Decubitus |
| HCC150 | Extensive Third-Degree Burns |
| HCC154 | Severe Head Injury |
| HCC155 | Major Head Injury |
| HCC157 | Vertebral Fractures without Spinal Cord Injury |
| HCC158 | Hip Fracture/Dislocation |
| HCC161 | Traumatic Amputation |
| HCC164 | Major Complications of Medical Care and Trauma |
| HCC174 | Major Organ Transplant Status |
| HCC176 | Artificial Openings for Feeding or Elimination |
| HCC177 | Amputation Status, Lower Limb/Amputation Complications |

Note: The list of HCC categories are copied from description of 2007 CMS-HCC software downloaded from the website of Center for Medicare and Medicaid Services.

Some CC categories are excluded because they have vague clinical health conditions or did not have significant effects on the prediction of medical costs. Seventy HCC variables are included in the HCC model and are used in this dissertation's proposed model.

The creation of HCC variables is based on the 2007 version of the CMS-HCC software. The CMS version of the HCC model has been incorporated into the Medicare capitation payment system and evaluated by researchers (Pope et al., 2004).

Table 2: The HCC Hierarchy System

| Hierarchies | CC Categories |
|---|---|
| Infection | CC5, CC112 |
| Neoplasm1 | CC7, CC8, CC9, CC10 |
| Neoplasm2 | CC8, CC9, CC10 |
| Neoplasm3 | CC9, CC10 |
| Diabetes1 | CC15, CC16, CC17, CC18, CC19 |
| Diabetes2 | CC16, CC17, CC18, CC19 |
| Diabetes3 | CC17, CC18, CC19 |
| Diabetes4 | CC18, CC19 |
| Liver1 | CC25, CC26, CC27 |
| Liver2 | CC26, CC27 |
| Agina1 | CC51, CC52 |
| Psychiatric | CC54, CC55 |
| Spinal1 | CC67, CC68, CC69, CC100, CC101, CC157 |
| Spinal2 | CC68, CC69, CC100, CC101, CC157 |
| Spinal3 | CC69, CC157 |
| Arrest1 | CC77, CC78, CC79 |
| Arrest2 | CC78, CC79 |
| Heart1 | CC81, CC82, CC83 |
| Heart2 | CC82, CC83 |
| CVD1 | CC95, CC96 |
| CVD2 | CC100, CC101 |
| Vascular1 | CC104, CC105, CC149 |
| Lung1 | CC107, CC108 |
| Lung2 | CC111, CC112 |
| Urinary1 | CC130, CC131, CC132 |
| Urinary2 | CC131, CC132 |
| Skin1 | CC148, CC149 |
| Injury1 | CC154, CC75, CC155 |
| Injury2 | CC161, CC177 |

Pharmacy-based risk adjusters are included in the dissertation model in addition to the diagnosis-based adjusters. For comparison purposes, the results also include a model with only pharmacy-based variables, a version of the RxRisk model. Pharmacy-based risk adjusters are developed from the Chronic Disease Score (CDS), defined by Clark and colleagues (1995). Putnam and colleagues (2002) compare four versions of the Chronic Disease Score and report that the Clark version performs better. The CDS categories are developed from the National Drug Code (NDC). Florida Medicaid claims files have a NDC variable that can provide information about prescription drug usage. Since Medicaid beneficiaries can have multiple prescription

drug claims, they are assigned to one or several CDS categories according to the number of claims. Like CCs, CDS categories are not mutually exclusive. However, there is no hierarchy imposed on the CDS categories; only 29 CDS variables were available for the dissertation model and the RxRisk model in contrast to the 70 HCC variables.

The CDS categories and corresponding HCC categories are listed in Table 3. The work for this dissertation reviewed all 29 CDS categories and finds that 12 categories cover the same diseases with certain CC categories.

These CDS categories are not duplicated in the dissertation model. When a Medicaid beneficiary uses prescription drugs in one of these CDS categories, the value of the corresponding HCC variable is set to 1, indicating that the patient has a diagnosis related to the HCC variables, even if the original value HCC is 0. For example, the value of HCC1 (HIV/AIDS) is set to 1 when the original value of HCC1 is 0 but the Medicaid beneficiary is prescribed drugs in the A3 (AIDS) category. Six CDS categories are not used in the proposed model because they correspond to multiple CC categories. In this case, a patient in any of these CDS categories can be assigned to one of the corresponding CC categories. If we add these CDS categories to our models, the predicted medical costs might be inappropriately inflated through a coding maneuver by the health plans. The six CDS categories excluded from our models are: A10 (diabetes), A15 (heart disease /hypertension), A20 (malignancies), A7 (cardiac disease), A24 (psychotic illness), and A8 (coronary/peripheral vascular disease). Research for this dissertation created CDS variables for the 11 CDS categories without corresponding HCC variables and added them directly into the

Table 3: CDS Disease Categories with the Corresponding HCC Categories

| Diseases | CDS Category | HCC Category |
|---|---|---|
| Cystic fibrosis | A1 | HCC107 |
| Diabetes | A10 | HCC15, 16, 17, 18, 19 |
| Epilepsy | A11 | HCC74 |
| Gastric acid disorder | A12 | None |
| Glaucoma | A13 | None |
| Gout | A14 | None |
| Heart disease | A15 | HCC80, 81, 82, 83, 92 |
| Hyperlipidemia | A16 | None |
| Hypertension | A17 | None |
| Inflammatory Bowel Disease | A18 | HCC33 |
| Liver failure | A19 | HCC25 |
| End stage renal disease | A2 | HCC131 |
| Malignancies | A20 | HCC7, 8, 9, 10 |
| Pain | A21 | None |
| Pain and inflammation | A22 | None |
| Parkinson's disease | A23 | HCC73 |
| Psychotic illness | A24 | HCC54 |
| Renal disease | A25 | None |
| Rheumatoid arthritis | A26 | HCC38 |
| Thyroid disorder | A27 | None |
| Transplant | A28 | HCC174 |
| Tuberculosis | A29 | None |
| AIDS | A3 | HCC1 |
| Anxiety and tension | A4 | None |
| Asthma | A5 | None |
| Bipolar disorder | A6 | HCC55 |
| Cardiac disease | A7 | HCC80, 81, 82, 83, 92 |
| Coronary disease | A8 | HCC81, 82, 83 |
| Depression | A9 | HCC55 |

dissertation model.

The demographic parts of the HCC model, the RxRisk model, and the dissertation model include the same variables. There are 24 age-sex crossing variables that represent 24 demographic groups with different age and gender characteristics. These variables are:

- "18-24-year-old female" group
- "25-34-year-old female" group
- "35-44-year-old female" group

- "45-54-year-old female" group
- "55-64-year-old female" group
- "65-69-year-old female" group
- "70-74-year-old female" group
- "75-79-year-old female" group
- "80-84-year-old female" group
- "85-89-year-old female" group
- "90-94-year-old female" group
- "95-year-old or over female" group
- "18-24-year-old male" group
- "25-34-year-old male" group
- "35-44-year-old male" group
- "45-54-year-old male" group
- "55-64-year-old male" group
- "65-69-year-old male" group
- "70-74-year-old male" group
- "75-79-year-old male" group
- "80-84-year-old male" group
- "85-89-year-old male" group
- "90-94-year-old male" group
- "95-year-old or over male" group

Medicaid covers medically needy younger adults as well as the elderly, and the prevalence of certain diseases can differ significantly among various age-gender groups. Moreover, there can be a large gap in the average medical costs of the younger and older beneficiaries, even if they have the same medical problems. Because the dissertation model uses information about demographic characteristics

to predict the year 2 medical expenses of those who are healthy in the base year, demographic variables have an important role in risk-adjustment.

Models with only demographic variables have low predictive power (Ash et al., 1989; Ellis et al., 1996; Newhouse, 1986), which was the primary motivation for including the health condition variables mentioned above.

Multiple linear regression models were built for analyses in this dissertation. The form of the regression models is:

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \ldots + \beta_n X_n ,$$

where Y represents the total amount of medical costs in the 2003-2004 year, $X_1$ to $X_n$ represent demographic, HCC, and CDS variables from SFY 2002-03 Medicaid claims. In other words, baseline year claims are used to flag the presence of health conditions, which are used to predict year 2 claims. A regression was run on the Medicaid populations and the entire adult aged and disabled population using the version 8 of the SAS software. The HCC and RxRisk models were run for comparison purposes. Both models have the same form as the dissertation model.

## Chapter Three

## Results

The SFY 2002–03 Florida Medicaid data includes 2,849,493 Florida Medicaid beneficiaries. Eighty-one percent (N = 2,310,464) of them were eligible for Medicaid for at least one month in SFY 2003–04. The Medicaid population for this study included 2,310,464 Medicaid beneficiaries during SFY 2002-03. The young subpopulation of disabled adults consisted of 144,846 beneficiaries. The aged subpopulation had 242,193 Medicaid beneficiaries in SFY 2002-03.

In the overall Medicaid population, 59.35 percent of beneficiaries (N = 1,371,223) were female enrollees. Eighty-nine percent of beneficiaries (N = 2,007,004) were younger than 65 years old.

In the aged subpopulation, 71.44 percent of beneficiaries (N = 173,028) are female enrollees. The "70-74-year-old female" category has the largest proportion (15.36 percent), and the "75-79-year-old female" is the second-largest group (14.65 percent) among all demographic categories. Eighty-seven percent of the beneficiaries were eligible for Medicaid for 12 months in SFY 2003–04. Their PMPM total Medicaid claims was $933.01 in SFY 2003–04.

In the young subpopulation of disabled adults, 57.90 percent of beneficiaries (N = 83,861) are female enrollees. The "55-64-year-old female" category has the largest proportion (15.84 percent), and the "45-54-year-old female" is the second-largest group (14.91 percent) among all demographic categories. Ninety-one percent of the beneficiaries were eligible for Medicaid for 12 months in SFY 2003–04. Their PMPM total Medicaid claims was $1,120.09 in SFY 2003–04.

## Model Performance

To validate the predictive power of the proposed models, the $R^2$ value of the models is compared. Since the coefficient of each independent variable in the risk-adjustment model represents the average increase in the PMPM Medicaid claims, the coefficients associated with each variable are also compared. Table 4 lists the summary statistics for the proposed model, the HCC model, and the RxRisk model for the Medicaid population.

The proposed model was found to have the highest $R^2$ value among the three models. The adjusted $R^2$ value of the proposed model, which compensates for adding independent variables, is 0.28. This compares with 0.24 for the HCC model and 0.21 for the RxRisk model. The proposed model also had the lowest Root Mean Square Error (RMSE = 944.42), which indicates variance about the predicted PMPM total Medicaid claims; models with a lower RMSE make better forecasts. The HCC model was found to have a higher $R^2$ value and lower RMSE than the RxRisk model, indicating that diagnosis-based risk-adjustors are better predictors of medical costs than prescription-based risk-adjusters.

Table 4: Statistics for the Proposed Model, the HCC Model, and the RxRisk Model for the Medicaid population

| | Proposed Model | HCC Model | RxRisk Model |
|---|---|---|---|
| **Statistics** | | | |
| Number of Enrollees | 2,310,464 | 2,310,464 | 2,310,464 |
| Number of Predictors | 111 | 99 | 58 |
| $R^2$ | 0.28 | 0.24 | 0.21 |
| Validated $R^2$ | 0.28 | 0.24 | 0.21 |
| Root Mean Square Error | 944.42 | 970.67 | 991.63 |
| **Variables** | PM | PM | PM |
| Intercept | 1421.96 | 1680.57 | 1309.51 |
| **Female** | | | |
| 18-24 Years | -1162.43 | -1402.50 | -1031.62 |
| 25–34 Years | -1166.16 | -1387.31 | -1036.46 |
| 35-44 Years | -1163.63 | -1316.02 | -1019.74 |
| 45–54 Years | -1098.21 | -1150.19 | -934.94 |
| 55-64 Years | -1057.35 | -1118.08 | -889.56 |
| 65–69 Years | -1203.56 | -1292.31 | -1136.32 |
| 70-74 Years | -1127.53 | -1198.10 | -1083.46 |
| 75–79 Years | -947.78 | -993.70 | -914.85 |
| 80-84 Years | -637.78 | -635.09 | -626.84 |
| 85–89 Years | -270.28 | -221.92 | -277.77 |
| 90-94 Years | 56.62 | 127.50 | 34.29 |
| 95 Years or Over | 392.37 | 447.66 | 371.91 |
| **Male** | | | |
| 18-24 Years | -1179.93 | -1397.55 | -982.36 |
| 25–34 Years | -1080.06 | -1244.44 | -886.34 |
| 35-44 Years | -1056.33 | -1147.85 | -874.32 |
| 45–54 Years | -955.48 | -1034.14 | -750.79 |
| 55-64 Years | -883.14 | -1004.20 | -650.95 |
| 65–69 Years | -1121.33 | -1253.70 | -1029.84 |
| 70-74 Years | -1062.68 | -1181.41 | -999.03 |
| 75–79 Years | -917.62 | -1007.34 | -866.76 |
| 80-84 Years | -686.92 | -729.68 | -657.33 |
| 85–89 Years | -359.07 | -339.75 | -364.58 |
| 90-94 Years | 161.19 | -124.73 | -163.78 |
| 95 Years or over | - | - | - |
| **HCC variables** | | | |
| **Infection** | | | |
| HCC1 | 904.17 | 1074.94 | - |
| HCC2 | 566.58 | 601.91 | - |
| HCC5 | 846.56 | 845.09 | - |

| | | | |
|---|---|---|---|
| **Neoplasm** | | | |
| HCC7 | 1190.18 | 1271.69 | - |
| HCC8 | 462.58 | 543.77 | - |
| HCC9 | 218.04 | 284.68 | - |
| HCC10 | 57.77 | 82.53 | - |
| **Endocrinal Disorders** | | | |
| HCC15 | 709.10 | 804.79 | - |
| HCC16 | 306.12 | 400.57 | - |
| HCC17 | 247.93 | 289.61 | - |
| HCC18 | 214.37 | 278.80 | - |
| HCC19 | 103.87 | 139.51 | - |
| HCC21 | 656.54 | 653.71 | - |
| **Gastrointestinal Disorders** | | | |
| HCC25 | 534.77 | 642.19 | - |
| HCC26 | 228.48 | 330.40 | - |
| HCC27 | 143.71 | 247.95 | - |
| HCC31 | 381.66 | 454.35 | - |
| HCC32 | 247.78 | 273.13 | - |
| HCC33 | 67.40 | 103.81 | - |
| **Connective Tissue Disease** | | | |
| HCC37 | 195.30 | 213.85 | - |
| HCC38 | -6.19 | 83.54 | - |
| **Hematological Disorders** | | | |
| HCC44 | 1151.22 | 1162.97 | - |
| HCC45 | 396.44 | 447.59 | - |
| **Psychiatric Disorders** | | | |
| HCC51 | 0.57 | 85.51 | - |
| HCC52 | 39.63 | 126.55 | - |
| HCC54 | 440.64 | 725.80 | - |
| HCC55 | 166.79 | 364.30 | - |
| **Spinal Cord Disorders** | | | |
| HCC67 | 2416.65 | 2488.53 | - |
| HCC68 | 961.39 | 1027.30 | - |
| HCC69 | 476.58 | 506.94 | - |
| **Neurological Disorders** | | | |
| HCC70 | 1333.61 | 1301.52 | - |
| HCC71 | 127.42 | 289.57 | - |
| HCC72 | 475.80 | 672.08 | - |
| HCC73 | 494.69 | 813.15 | - |
| HCC74 | 588.76 | 815.30 | - |

| | | | |
|---|---|---|---|
| HCC75 | 1348.61 | 1334.25 | - |
| **Respiratory Disorders** | | | |
| HCC77 | 3896.12 | 3922.82 | - |
| HCC78 | 597.94 | 665.95 | - |
| **Cardiac Diseases** | | | |
| HCC79 | 410.70 | 451.56 | - |
| HCC80 | 369.33 | 414.59 | - |
| HCC81 | 117.29 | 55.77 | - |
| HCC82 | -20.23 | -15.02 | - |
| HCC83 | -154.07 | -127.56 | - |
| HCC92 | 131.78 | 129.02 | - |
| **Cerebral and Vascular Diseases** | | | |
| HCC95 | 194.98 | 191.24 | - |
| HCC96 | 457.79 | 537.97 | - |
| HCC100 | 500.44 | 524.37 | - |
| HCC101 | 1707.94 | 1777.23 | - |
| HCC104 | 420.80 | 455.21 | - |
| HCC105 | 343.28 | 395.69 | - |
| **Pulmonary Diseases** | | | |
| HCC107 | 19.53 | 1574.13 | - |
| HCC108 | -14.13 | 132.68 | - |
| HCC111 | 952.72 | 911.62 | - |
| HCC112 | 196.08 | 209.27 | - |
| **Eye Disorders** | | | |
| HCC119 | -37.00 | -22.28 | - |
| **Renal Disorders** | | | |
| HCC130 | 692.33 | 1127.53 | - |
| HCC131 | 487.36 | 536.02 | - |
| HCC132 | 183.03 | 236.06 | - |
| **Skin Disorders** | | | |
| HCC148 | 610.57 | 721.78 | - |
| HCC149 | 424.29 | 510.83 | - |
| HCC150 | 207.37 | 234.00 | - |
| **Injury** | | | |
| HCC154 | 864.19 | 926.78 | - |
| HCC155 | 140.17 | 192.38 | - |
| HCC157 | 124.15 | 170.77 | - |
| HCC158 | 357.44 | 434.85 | - |
| HCC161 | 90.82 | 181.45 | - |
| HCC164 | 387.74 | 436.66 | - |
| **Others** | | | |
| HCC174 | 256.27 | 918.76 | - |

| | | | |
|---|---|---|---|
| HCC176 | 1745.14 | 1827.30 | - |
| HCC177 | 646.77 | 699.96 | - |
| **CDS Variables** | | | |
| A1 | - | - | 18.14 |
| A2 | - | - | 1086.18 |
| A3 | - | - | 989.09 |
| A4 | 163.00 | - | 224.32 |
| A5 | 96.43 | - | 143.73 |
| A6 | - | - | -58.46 |
| A7 | - | - | 284.29 |
| A8 | - | - | 84.03 |
| A9 | - | - | 114.01 |
| A10 | - | - | 190.25 |
| A11 | - | - | 740.76 |
| A12 | 157.33 | - | 219.19 |
| A13 | 32.01 | - | 14.63 |
| A14 | 23.47 | - | -29.57 |
| A15 | - | - | -72.46 |
| A16 | -117.34 | - | -201.69 |
| A17 | 54.20 | - | 41.91 |
| A18 | - | - | 79.34 |
| A19 | - | - | 745.49 |
| A20 | - | - | 428.62 |
| A21 | -32.82 | - | 9.23 |
| A22 | -150.23 | - | -181.94 |
| A23 | - | - | 454.67 |
| A24 | - | - | 452.24 |
| A25 | 592.96 | - | 798.12 |
| A26 | - | - | 11.51 |
| A27 | 264.13 | - | 231.57 |
| A28 | - | - | 211.83 |
| A29 | 433.97 | - | 725.61 |

Notes: PM is Parameter. "-" indicates a variable that is not relevant for a particular model. Male patients age 95 or over are used as baseline in our models. The units of all medical costs are dollars.

The "age 95 years or over male" group was used as the baseline age-gender group in all three models. Future medical costs peak at age 45 to 64 years, and then decrease between age 65 and 69 years. After age 65 to 69 years, future medical costs increase again and have a positive relationship to age.

The HCC variables were divided into 18 groups and compared parameters within each group in both the proposed model and the HCC model. Three categories fell within the "infection" group: HCC1 (HIV/AIDS), HCC2 (septicemia/shock), and

HCC5 (opportunistic infections). In the proposed model, Acquired Immune Deficiency Syndrome (AIDS) was associated with the highest increase in future medical costs among the three diseases, which was $904.17 in the proposed model and $1,074.94 in the HCC model.

The "neoplasm" group includes HCC7 (metastatic cancer and acute leukemia), HCC8 (lung, upper digestive tract, and other severe cancers), HCC9 (lymphatic, heart and neck, brain, and other major cancers), and HCC10 (breast, prostate, colorectal and other cancers and tumors). Their ranks were positively associated with future medical costs. For example, metastatic cancer and acute leukemia increase PMPM medical costs by $1,190.18 in the proposed model after other factors are controlled, and by $1,271.69 in the HCC model. Breast, prostate, and colorectal cancers are associated with a lower increase in PMPM medical costs than lung, upper digestive tract, and brain cancers.

The "endocrinal disorders" group includes HCC15 (diabetes with renal or peripheral circulatory manifestation), HCC16 (diabetes with neurological or other specified manifestation), HCC17 (diabetes with acute complications), HCC18 (diabetes with ophthalmologic or unspecified manifestation), HCC19 (diabetes without complication) and HCC21 (protein-calorie malnutrition). Among all diabetes categories, diabetes with renal or peripheral circulatory manifestation was associated with the highest increase in the PMPM medical cost, which is $709.10 in the proposed model and $804.79 in the HCC model. Protein-calorie malnutrition is also associated with relatively high PMPM medical costs in the "endocrinal disorders" group.

The "gastrointestinal disorders" group consisted of six HCC variables: HCC25 (end-stage liver disease), HCC26 (cirrhosis of the liver), HCC27 (chronic hepatitis), HCC31 (intestinal obstruction/perforation), HCC32 (pancreatic disease), and HCC33

(inflammatory bowel disease). End-stage liver disease was associated with the highest increase in the PMPM medical cost among all gastrointestinal disorders, which is $534.77 in the proposed model and $642.19 in the HCC model.

The "connective tissue disease" group includes HCC37 (bone/joint/muscle infections/necrosis) and HCC38 (rheumatoid arthritis and inflammatory connective tissue disease). They were associated with a moderate increase in the PMPM medical cost.

Two HCC variables fall into the "hematological disorders" group: HCC44 (severe hematological disorders) and HCC45 (immune disorders). Severe hematological disorders are significantly associated with a high increase in the PMPM medical cost, which is $1,152.22 in the proposed model and $1,162.98 in the HCC model.

The "psychiatric disorders" group includes four categories: HCC51 (drug/alcohol psychosis), HCC52 (drug/alcohol dependence), HCC54 (schizophrenia), and HCC55 (major depressive, bipolar, and paranoid disorders). Schizophrenia is associated with the highest increase in the PMPM medical cost in this group, which is $440.64 in the proposed model and $725.80 in the HCC model.

The "spinal cord disorders" group consists of HCC67 (quadriplegia, other extensive paralysis), HCC68 (paraplegia), and HCC69 (spinal cord disorders/injuries). These are associated with a high increase in the PMPM medical cost, especially quadriplegia and paraplegia. Quadriplegia increases the PMPM medical cost by $2,416.65 in the proposed model and $2,488.53 in the HCC model.

The "neurological disorders" group includes HCC70 (muscular dystrophy), HCC71 (polyneuropathy), HCC72 (multiple sclerosis), HCC73 (Parkinson's and Huntington's Disease), HCC74 (seizure disorders and convulsions) and HCC75 (coma, brain compression/anoxic damage). Muscular dystrophy and coma are both

associated with more than a $1,300 increase in the PMPM medical cost.

The "respiratory disorders" group consists of HCC77 (respiratory dependence/tracheotomy status) and HCC78 (respiratory arrest). Respiratory dependence or tracheotomy status is associated with the highest increase in the PMPM medical cost among all diagnoses, $3,896.12 in the proposed model, and $3,922 in the HCC model.

The "cardiac diseases" group includes six categories: HCC79 (cardio-respiratory failure and shock), HCC80 (congestive heart failure), HCC81 (acute myocardial infarction), HCC82 (unstable angina and other acute ischemic heart disease), HCC83 (angina pectoris/old myocardial infarction), and HCC92 (specified heart arrhythmias). Cardio-respiratory failure and shock is significantly associated with the highest PMPM medical cost, $410.70 in the proposed model, and $451.56 in the HCC model.

The "cerebral and vascular diseases" group includes HCC95 (cerebral hemorrhage), HCC96 (ischemic or unspecified stroke), HCC100 (hemiplegia/hemiparesis), HCC101 (cerebral palsy and other paralytic syndromes), HCC104 (vascular disease with complications), and HCC105 (vascular disease). Cerebral palsy and other paralytic syndromes are associated with the highest increase in the PMPM medical cost, $1,707.94 in the proposed model, and $1,777.23 in the HCC model.

The "pulmonary diseases" group consists of four categories: HCC107 (cystic fibrosis), HCC108 (chronic obstructive pulmonary disease), HCC111 (aspiration and specified bacterial pneumonias), HCC112 (pneumococcal pneumonia, empyema, and lung abscess). Parameters of cystic fibrosis are very different for the proposed model and the HCC model – $19.53 and $1,574.13, respectively.

There is only one category in the "eye disorders" group: HCC119 (proliferative

diabetic retinopathy and vitreous hemorrhage). It is associated with a negative increase in the PMPM medical cost.

The "renal disorders" group includes HCC130 (dialysis status), HCC131 (renal failure), and HCC132 (nephritis). Dialysis is associated with the highest increase in the PMPM medical cost in this group, which is $692.33 in the proposed model and $1,127.53 in the HCC model.

The "skin disorders" group consists of HCC148 (decubitus ulcer of the skin), HCC149 (chronic ulcer of the skin, except decubitus) and HCC150 (extensive third-degree burns). Decubitus skin ulcer is associated with the highest increase in the PMPM medical cost in this group, which is $610.57 in the proposed model and $721.78 in the HCC model.

The "injury" group includes HCC154 (severe head injury), HCC155 (major head injury), HCC157 (vertebral fractures w/o spinal cord injury), HCC158 (hip fracture/dislocation), HCC161 (traumatic amputation), and HCC164 (major complications of medical care and trauma). Severe head injury is associated with a $864.19 increase in the proposed model and a $926.78 increase in the HCC model, which are higher than those of other disorders in the "injury" group.

Several HCC categories are assigned to the "others" group: HCC174 (major organ transplant status), HCC176 (artificial openings for feeding or elimination), and HCC177 (amputation status, lower limb/amputation complications). Artificial openings for feeding or elimination and amputation status are significantly associated with relatively high increase in the PMPM medical cost. Artificial openings for feeding or elimination increase PMPM medical cost by $1,745.14 in the proposed model and $1,827.30 in the HCC model.

Twenty-nine CDS variables are added into the RxRisk model and 12 of them are also independent variables in the proposed model in this dissertation. They are:

A12 (gastric acid disorder), A13 (glaucoma), A14 (gout), A16 (hyperlipidemia), A21 (pain), A22 (pain and inflammation), A25 (renal disease), A27 (thyroid disorder), A29 (tuberculosis), A4 (anxiety and tension), A5 (asthma), and A17 (hypertension). In the RxRisk model, these variables are significantly associated with a high increase in the PMPM medical cost: end-stage renal disease ($1,086.18), epilepsy ($740.76), liver failure ($745.49), renal disease ($798.12), and tuberculosis ($725.61). In the proposed model, renal disease and tuberculosis are also associated with high increases in PMPM medical costs, $592.96 and $433.97, respectively.

The aged subpopulation included 242,192 Medicaid enrollees aged 65 and above. The proposed model has the best predictive performance since its $R^2$ value ($R^2$ = 0.32) and RMSE (RMSE = 1,174.92) is the highest among all models for the aged population. The pharmacy-based RxRisk model has slightly lower $R^2$ value ($R^2$ = 0.30) and higher RMSE (RMSE = 1,187.07) than the proposed model. The HCC model has relatively poorer predictive performance with lower $R^2$ value ($R^2$ = 0.17) and higher RMSE (RMSE = 1,299.12)

In all three models for the aged subpopulation, age has a strong and linear effect on PMPM total Medicaid claims for both male and female groups. The PMPM medical costs increase significantly with a five-year increment of age. Among all age-gender groups with beneficiaries aged 80 or above only, female beneficiaries have higher average PMPM medical costs than male beneficiaries. In the proposed model, the older male beneficiaries aged under 80 have higher PMPM medical costs than their female counterparts. The results of statistics for the aged subpopulation are shown in Table 5.

Table 5: Compared Statistics for the Proposed Model, the HCC Model, and the RxRisk Model for the Aged and the Young Disabled Adults

| | Proposed Model | | HCC Model | | RxRisk Model | |
|---|---|---|---|---|---|---|
| | Aged | Disabled | Aged | Disabled | Aged | Disabled |
| **Statistics** | | | | | | |
| Number of Enrollees | 242,192 | 144,845 | 242,192 | 144,845 | 242,192 | 144,845 |
| Number of Predictors | 93 | 91 | 81 | 79 | 42 | 38 |
| R2 | 0.32 | 0.25 | 0.17 | 0.23 | 0.30 | 0.16 |
| Validated R2 | 0.32 | 0.25 | 0.17 | 0.23 | 0.30 | 0.16 |
| RMSE | 1,174.92 | 1,627.39 | 1,299.12 | 1,648.85 | 1,187.07 | 1,725.33 |
| **Variables** | PM | PM | PM | PM | PM | PM |
| Intercept | 279.11 | 524.12 | 488.57 | 563.48 | 297.02 | 705.22 |
| **Female** | | | | | | |
| 18–24 Years | - | -28.84 | - | -51.84 | - | -52.54 |
| 25–34 Years | - | 93.29 | - | 94.83 | - | -1.22 |
| 35–44 Years | - | -55.73 | - | -24.19 | - | -211.08 |
| 45–54 Years | - | -86.97 | - | -40.88 | - | -255.85 |
| 55–64 Years | - | -78.26 | - | -48.77 | - | -262.44 |
| 65–69 Years | -120.97 | - | -62.43 | - | -122.97 | - |
| 70-74 Years | -47.56 | - | 3.49 | - | -67.29 | - |
| 75–79 Years | 106.20 | - | 185.48 | - | 79.02 | - |
| 80-84 Years | 371.59 | - | 516.27 | - | 337.02 | - |
| 85–89 Years | 703.33 | - | 925.70 | - | 657.66 | - |
| 90-94 Years | 1018.77 | - | 1282.72 | - | 966.27 | - |
| 95 Years or over | 1369.69 | - | 1617.23 | - | 1316.01 | - |
| **Male** | | | | | | |
| 18–24 Years | - | - | - | - | - | - |
| 25–34 Years | - | 237.26 | - | 256.26 | - | 214.23 |
| 35–44 Years | - | 84.45 | - | 108.84 | - | 18.11 |
| 45–54 Years | - | 0.41 | - | 10.91 | - | -81.90 |
| 55–64 Years | - | 31.41 | - | 22.16 | - | -60.23 |
| 65–69 Years | - | - | - | - | - | - |
| 70-74 Years | 37.84 | - | 26.37 | - | 16.77 | - |
| 75–79 Years | 146.72 | - | 166.63 | - | 120.29 | - |
| 80-84 Years | 323.17 | - | 410.79 | - | 287.48 | - |
| 85–89 Years | 587.15 | - | 763.69 | - | 533.99 | - |
| 90-94 Years | 773.01 | - | 986.72 | - | 721.91 | - |
| 95 Years or over | 1006.31 | - | 1140.63 | - | 949.89 | - |
| **HCC variables** | | | | | | |
| **Infection** | | | | | | |
| HCC1 | 638.05 | 950.89 | 1132.92 | 1039.33 | - | - |
| HCC2 | 351.37 | 530.62 | 607.81 | 573.93 | - | - |
| HCC5 | -253.75 | 896.15 | -294.49 | 959.37 | - | - |

| | | | | | | |
|---|---|---|---|---|---|---|
| **Neoplasm** | | | | | | |
| HCC7 | 329.59 | 998.78 | 508.65 | 1153.19 | - | - |
| HCC8 | -129.70 | 495.47 | 142.42 | 578.79 | - | - |
| HCC9 | -131.86 | 69.38 | -4.85 | 128.33 | - | - |
| HCC10 | -55.24 | 32.44 | -57.23 | 66.05 | - | - |
| **Endocrinal Disorders** | | | | | | |
| HCC15 | 902.32 | 462.09 | 1110.82 | 575.41 | - | - |
| HCC16 | 279.37 | 340.70 | 369.90 | 442.04 | - | - |
| HCC17 | 133.08 | 378.67 | 160.75 | 419.29 | - | - |
| HCC18 | 289.71 | 201.77 | 410.25 | 252.35 | - | - |
| HCC19 | 107.71 | 137.68 | 128.65 | 189.46 | - | - |
| HCC21 | 769.06 | 783.46 | 827.42 | 789.10 | - | - |
| **Gastrointestinal Disorders** | | | | | | |
| HCC25 | 4.16 | 805.47 | -242.31 | 485.16 | - | - |
| HCC26 | 196.95 | 4.62 | 271.74 | 152.08 | - | - |
| HCC27 | -167.63 | 194.06 | -41.52 | 281.49 | - | - |
| HCC31 | 388.66 | 464.33 | 543.89 | 556.99 | - | - |
| HCC32 | 97.81 | 149.00 | 49.83 | 220.29 | - | - |
| HCC33 | 106.28 | -92.40 | 45.11 | -122.63 | - | - |
| **Connective Tissue Disease** | | | | | | |
| HCC37 | 94.18 | 168.74 | 216.08 | 157.26 | - | - |
| HCC38 | -49.75 | 53.30 | 14.59 | 61.24 | - | - |
| **Hematological Disorders** | | | | | | |
| HCC44 | 211.31 | 1278.71 | 277.28 | 1274.97 | - | - |
| HCC45 | 247.85 | 496.35 | 332.46 | 561.70 | - | - |
| **Psychiatric Disorders** | | | | | | |
| HCC51 | 419.61 | -11.95 | 599.24 | 8.02 | - | - |
| HCC52 | -42.91 | 121.46 | 70.93 | 183.47 | - | - |
| HCC54 | 954.61 | 440.43 | 953.02 | 676.85 | - | - |
| HCC55 | 396.07 | -81.48 | 630.69 | 191.00 | - | - |
| **Spinal Cord Disorders** | | | | | | |
| HCC67 | 992.65 | 2540.17 | 940.41 | 2619.79 | - | - |
| HCC68 | 599.40 | 895.19 | 595.42 | 936.97 | - | - |
| HCC69 | 208.78 | 381.20 | 182.82 | 422.97 | - | - |
| **Neurological Disorders** | | | | | | |
| HCC70 | 47.99 | 1710.88 | 67.68 | 1663.91 | - | - |
| HCC71 | -235.38 | 137.77 | -129.38 | 313.48 | - | - |
| HCC72 | 1482.59 | 320.41 | 1757.72 | 461.24 | - | - |
| HCC73 | 624.72 | 413.82 | 851.10 | 720.18 | - | - |
| HCC74 | 584.43 | 647.71 | 1059.17 | 1070.44 | - | - |

| | | | | | | |
|---|---|---|---|---|---|---|
| HCC75 | 490.73 | 514.08 | 625.18 | 442.88 | - | - |
| **Respiratory Disorders** | | | | | | |
| HCC77 | 937.73 | 1489.20 | 1089.04 | 1542.30 | - | - |
| HCC78 | 71.88 | 485.02 | 443.41 | 565.26 | - | - |
| **Cardiac Diseases** | | | | | | |
| HCC79 | 157.94 | 292.71 | 399.26 | 290.09 | - | - |
| HCC80 | 233.49 | 392.63 | 359.68 | 419.11 | - | - |
| HCC81 | -61.56 | 171.21 | -143.30 | 168.72 | - | - |
| HCC82 | -112.26 | 81.88 | -136.20 | 96.40 | - | - |
| HCC83 | -191.59 | -63.69 | -225.33 | -32.76 | - | - |
| HCC92 | 113.84 | 150.17 | 104.26 | 174.39 | - | - |
| **Cerebral and Vascular Diseases** | | | | | | |
| HCC95 | 56.53 | -35.43 | 125.95 | -92.44 | - | - |
| HCC96 | 582.30 | 285.71 | 778.67 | 267.36 | - | - |
| HCC100 | 506.01 | 632.72 | 690.24 | 595.09 | - | - |
| HCC101 | 1484.98 | 2167.79 | 1642.43 | 2183.22 | - | - |
| HCC104 | 346.45 | 493.64 | 435.06 | 517.01 | - | - |
| HCC105 | 537.08 | 206.29 | 680.87 | 246.66 | - | - |
| **Pulmonary Diseases** | | | | | | |
| HCC107 | -276.54 | 183.71 | 152.02 | 1520.66 | - | - |
| HCC108 | -272.45 | 198.32 | -33.77 | 315.80 | - | - |
| HCC111 | 682.89 | 736.77 | 873.28 | 686.64 | - | - |
| HCC112 | -82.59 | 221.48 | -38.93 | 232.49 | - | - |
| **Eye Disorders** | | | | | | |
| HCC119 | -46.76 | -40.67 | -100.50 | 10.05 | - | - |
| **Renal Disorders** | | | | | | |
| HCC130 | 507.48 | 1181.70 | 1050.71 | 1924.46 | - | - |
| HCC131 | 359.16 | 493.86 | 456.80 | 535.42 | - | - |
| HCC132 | 36.03 | 133.76 | -96.51 | 130.00 | - | - |
| **Skin Disorders** | | | | | | |
| HCC148 | 581.37 | 707.38 | 750.18 | 756.87 | - | - |
| HCC149 | 357.14 | 514.82 | 515.12 | 539.20 | - | - |
| HCC150 | - | -845.99 | - | -965.15 | - | - |
| **Injury** | | | | | | |
| HCC154 | - | 1153.89 | - | 1282.70 | - | - |
| HCC155 | 927.97 | 112.01 | 1238.58 | 116.40 | - | - |
| HCC157 | 198.52 | 277.55 | 346.92 | 314.13 | - | - |
| HCC158 | 334.12 | 221.28 | 560.90 | 186.56 | - | - |
| HCC161 | 130.06 | 122.16 | 247.82 | 152.31 | - | - |
| HCC164 | 321.83 | 209.06 | 474.12 | 246.58 | - | - |
| **Others** | | | | | | |
| HCC174 | 348.84 | 3.11 | 837.32 | 494.98 | - | - |

| | | | | | | |
|---|---|---|---|---|---|---|
| HCC176 | 512.00 | 1378.58 | 715.73 | 1417.84 | - | - |
| HCC177 | 451.30 | 689.39 | 434.47 | 749.75 | - | - |
| **CDS Variables** | | | | | | |
| A1 | - | - | - | - | -293.85 | 218.55 |
| A2 | - | - | - | - | 514.55 | 1283.73 |
| A3 | - | - | - | - | 631.87 | 1043.85 |
| A4 | -17.23 | 299.51 | - | - | -37.04 | 398.03 |
| A5 | 326.27 | 128.45 | - | - | 265.56 | 237.55 |
| A6 | - | - | - | - | 89.38 | 24.43 |
| A7 | - | - | - | - | 294.91 | 268.26 |
| A8 | - | - | - | - | 40.41 | 260.77 |
| A9 | - | - | - | - | 374.71 | -167.51 |
| A10 | - | - | - | - | 191.87 | 290.34 |
| A11 | - | - | - | - | 589.68 | 791.41 |
| A12 | 134.74 | 116.84 | - | - | 120.04 | 246.50 |
| A13 | -8.04 | 228.47 | - | - | -24.09 | 280.22 |
| A14 | 5.41 | 18.75 | - | - | -57.65 | 108.21 |
| A15 | - | - | - | - | -59.59 | -5.80 |
| A16 | -164.73 | -37.43 | - | - | -214.44 | -145.51 |
| A17 | 41.94 | 20.88 | - | - | -3.41 | 51.87 |
| A18 | - | - | - | - | 102.12 | 46.75 |
| A19 | - | - | - | - | 5.24 | 1371.01 |
| A20 | - | - | - | - | 254.95 | 826.56 |
| A21 | 137.14 | -167.80 | - | - | 114.67 | -22.35 |
| A22 | -177.33 | -190.27 | - | - | -186.48 | -283.17 |
| A23 | - | - | - | - | 648.48 | 349.09 |
| A24 | - | - | - | - | 978.58 | 478.58 |
| A25 | 799.74 | 1179.06 | - | - | 919.85 | 2260.73 |
| A26 | - | - | - | - | -89.21 | 91.78 |
| A27 | 210.55 | 209.75 | - | - | 185.84 | 228.29 |
| A28 | - | - | - | - | 283.19 | -101.03 |
| A29 | 704.13 | 448.37 | - | - | 766.87 | 1097.29 |

Notes: PM is Parameter. "-" indicates a variable that is not relevant for a particular model.

The young disabled subpopulation of Medicaid beneficiaries consisted of 144,845 enrollees, compared with 242,192 enrollees in the aged subpopulation. Generally, the proposed model and the RxRisk model have poorer predictive performance in the young disabled subpopulation than in the aged subpopulation as indicated by lower $R^2$ value and high RMSE in the former subpopulation. The proposed model explains 25 percent of variance in the young disabled adults' PMPM medical costs, compared with 32 percent in the aged population. The $R^2$ value of the

RxRisk model for the young disabled subpopulation is 0.16, compared with 0.30 for the aged subpopulation. However, the HCC model has higher $R^2$ value ($R^2$ = 0.23) for the young disabled subpopulation, compared with 0.17 for the aged subpopulation. The results of statistics for the young disabled subpopulation can also be found in Table 5.

RMSE had a higher value in all three models for the young disabled subpopulation than the aged subpopulation. The proposed model had the lowest RMSE (RMSE = 1,627.39) for the young disabled subpopulation among all models.

Age had a nonlinear effect on PMPM total Medicaid claims for the young disabled subpopulation in all three models. The "25-34-year-old male" group and the "25-34-year-old female" group had the highest PMPM medical costs among male and female younger disabled adults respectively, and male beneficiaries had higher PMPM medical costs than female beneficiaries.

The proposed model has better predictive performance than the HCC model and the RxRisk model. Among all infectious diseases, AIDS is associated with the highest increase in PMPM total Medicaid claims for both the aged and the young disabled subpopulation. The proposed model predicts that AIDS increases the PMPM medical cost by $638.05 in the aged subpopulation and by $950.89 in the young disabled subpopulation.

Metastatic cancers and acute leukemia are associated with higher increase in the PMPM total Medicaid cost for both the aged and the young disabled subpopulation than other diseases in the "neoplasm" group. The proposed model predicts that Metastatic cancers and acute leukemia increase the PMPM medical cost by $638.05 in the aged subpopulation and by $950.89 in the disabled subpopulation.

Diabetes with renal or peripheral circulatory manifestation increases the PMPM medical cost by $902.32 in the aged subpopulation, compared with $462.09 in

the disabled subpopulation. Protein-calorie malnutrition is associated with $769.06 of increase in the PMPM medical cost in the aged subpopulation and $783.46 of increase in the young disabled subpopulation.

End-stage liver disease increases the PMPM medical cost by $805.47 in the young disabled subpopulation and $4.25 in the aged subpopulation. Severe hematological disorders are associated with $1,278.71 of increase in PMPM total Medicaid claims for the young disabled subpopulation, and $211.31 of increase in claims for the aged subpopulation.

Schizophrenia increases the PMPM medical cost by $954.61 in the aged subpopulation and $440.43 in the young disabled subpopulation. Among neurological disorders, quadriplegia and paraplegia increase PMPM medical costs by $2,540.17 and $895.19 respectively in the young disabled subpopulation, and $992.65 and $599.40 respectively in the aged subpopulation. Muscular dystrophy increases the PMPM medical cost by $1,710.88 in the young disabled subpopulation, and multiple sclerosis increases the PMPM medical cost by $1,482.59 in the aged subpopulation.

Respiratory dependence is associated with high increase in PMPM total Medicaid claims, which is $937.73 in the aged subpopulation and $1,489.20 in the young disabled subpopulation. Among all cardiac diseases, congestive heart failure increase the PMPM medical cost by the largest margin, which is $233.49 in the aged subpopulation and $392.63 in the young disabled subpopulation.

Cerebral palsy and other paralytic syndromes increase the PMPM medical cost by $1,484.98 in the aged subpopulation and $2,167.79 in the young disabled subpopulation. Dialysis increases the PMPM medical cost by $1,181.70 in the young disabled subpopulation, and a major head injury increases the PMPM medical cost

Table 6: Predictive Ratios in the HCC model and Proposed Model for the Aged Subpopulation

| Diseases | Proposed Model PR | HCC Model PR |
|---|---|---|
| Acute Myocardial Infarction | 1.02 | 0.91 |
| Aspiration and Specified Bacterial Pneumonias | 0.97 | 0.95 |
| Cardio – Respiratory Failure and Shock | 0.93 | 0.84 |
| Chronic Obstructive Pulmonary Disease | 0.99 | 0.99 |
| Congestive Heart Failure | 0.97 | 0.93 |
| Diabetes with Renal or Peripheral Circulatory Manifestation | 0.99 | 0.94 |
| Dialysis Status | 1.00 | 0.99 |
| End-Stage Liver Disease | 0.78 | 0.74 |
| Hip Fracture/Dislocation | 1.00 | 1.00 |
| Inflammatory Bowel Disease | 0.99 | 0.99 |
| Ischemic or Unspecified Stroke | 0.96 | 0.96 |
| Major Depressive, Bipolar, and Paranoid Disorders | 0.98 | 0.98 |
| Metastatic Cancer and Acute Leukemia | 0.78 | 0.78 |
| Parkinson's and Huntington's Disease | 0.97 | 0.95 |
| Renal Failure | 0.99 | 0.97 |
| Schizophrenia | 0.90 | 0.90 |

Note: PR means predictive ratio.

by $927.97 in the aged subpopulation.

Model performance can also be evaluated by the predictive ratio. The predictive ratio is calculated by dividing predicted medical cost by actual medical cost in the year 2 data within different subgroups. The ideal value of predictive ratio is 1, which means the mean of predicted medical cost is equal to the mean of actual medical cost. Risk-adjustment models with predictive ratios closer to 1 are regarded as the models with higher predictive power. In Table 6, predictive ratios in the HCC model and the proposed model for the aged subpopulation are listed. The results show that the proposed model has better predictive ratios than the HCC model in all disease groups except chronic obstructive pulmonary disease, hip fracture, inflammatory bowel Disease, ischemic or unspecified stroke, major depression, metastatic cancers and schizophrenia.

The aged subpopulation and the young disabled subpopulation account for different proportion among all Florida Medicaid beneficiaries. In SFY 2002-03, 10.5

percent of Medicaid enrollees were aged 65 or above and 6.27 percent of Medicaid enrollees were disabled adults younger than 65. The distribution of SFY 2002-03 PMPM medical costs among two subpopulations of Medicaid enrollees is listed in Table 7. Except for the 90$^{th}$ percentile, the young disabled subpopulation has higher PMPM medical costs in all percentiles listed in Table 7. Figure 2 draws the picture of distribution of SFY 2002-03 PMPM medical costs in percentiles among two subpopulations. The blue triangle sign indicates the number of the younger disabled beneficiaries' PMPM costs in each percentile strata and the red dot sign indicates the number of the aged beneficiaries' PMPM costs in each percentile strata. The younger disabled beneficiaries, whose PMPM costs fall within top 5 percentiles, need much more reimbursements than their counterparts in the aged subpopulation.

The average PMPM costs in various disease groups are also different for the aged subpopulation and the young disabled subpopulation, as shown in Figure 1. Among the patients with psychotic diseases, younger disabled Medicaid enrollees have higher PMPM costs than older enrollees. The younger patients with renal or pulmonary diseases including renal failure and pneumonia also need much more reimbursement for their health care services than their older counterparts. However, the younger patients with chronic heart failure have similar medical costs with their older counterparts. Acute myocardial infarction is associated with higher PMPM medical cost in the young disabled subpopulation than in the aged subpopulation.

Table 7: Distribution of SFY 2002-03 PMPM Medical Costs among Two Subpopulations of Medicaid Enrollees

| Percentile | Aged | Disabled |
|---|---|---|
| Maximum | 34838.30 | 68970.28 |
| 99 | 5218.20 | 9092.51 |
| 95 | 4322.18 | 4877.27 |
| 90 | 3618.42 | 2702.80 |
| 75 | 921.56 | 1025.14 |
| 50 | 311.88 | 526.97 |
| 25 | 90.91 | 260.34 |
| 10 | 0.00 | 61.08 |
| 5 | 0.00 | 5.54 |
| 1 | 0.00 | 0.00 |

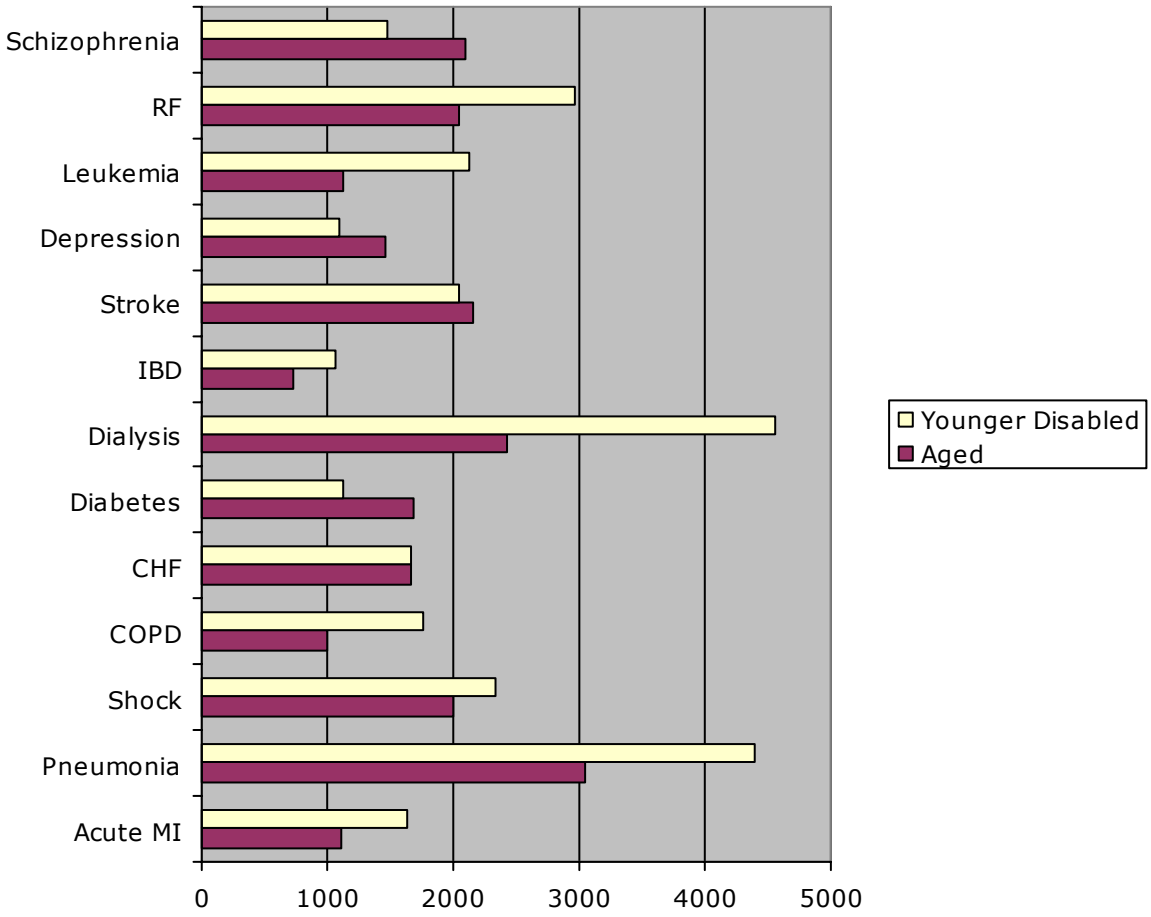


Figure 1: The Average PMPM Costs in Various Disease Groups for Two Subpopulations

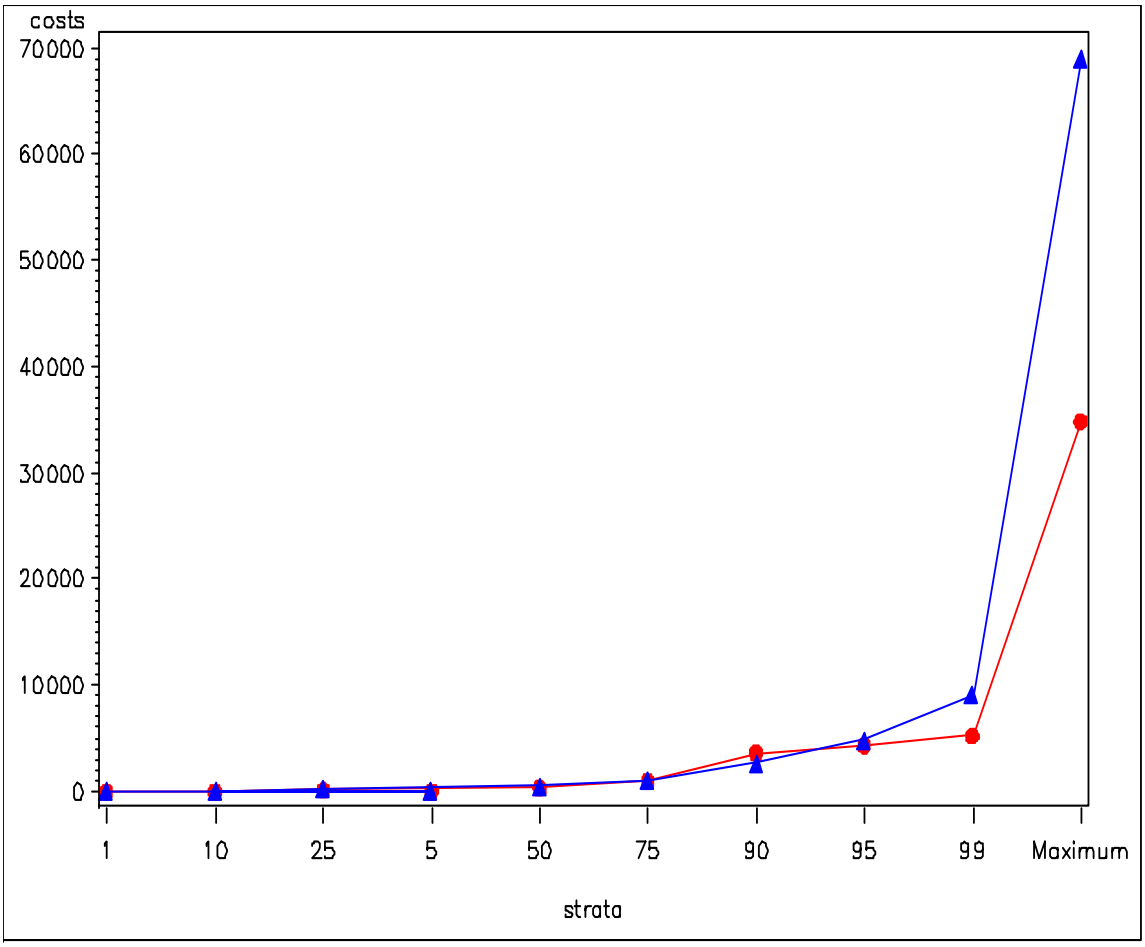


Figure 2: Distribution of SFY 2002-03 PMPM Medical Costs in Percentiles among Two Subpopulations

Note: RF is renal failure. IBD is inflammatory bowel disease. CHF is chronic heart failure. COPD is chronic obstructive pulmonary disease. MI is myocardial infarction.

# Chapter Four

## Discussion

Capitation payment systems based on risk adjustment methodologies can more equitably reimburse health plans with a large proportion of sicker beneficiaries if an effective risk adjustment model with high predictive power is developed. In former studies exploring risk adjustment methodologies, diagnoses and prescription drug usage respectively were used to build predictive models. In this dissertation, a model using information about diagnoses and prescription drug usage was built and compared with diagnosis-based and pharmacy-based models.

Since Medicaid beneficiaries other than the aged and the young disabled account for a large proportion (83.2 percent), it is necessary to compare the models for the aged and the young disabled subpopulations with the models based on the whole Medicaid population. The results show that the proposed model and the pharmacy-based RxRisk model predict the aged beneficiaries' PMPM medical costs better than they do for the whole Florida Medicaid population's PMPM costs. The variance explained for the aged beneficiaries subpopulation using the proposed model is slightly greater than that explained by the RxRisk model. The HCC model, however, predicts the general Florida Medicaid population's PMPM costs better than it does for the aged beneficiaries' costs. All models run against the aged subpopulation have higher RMSE and are therefore less accurate than the same models run against the general Medicaid population. The pharmacy-based RxRisk model had very close predictive performance in contrast to the proposed model for the aged subpopulation and it can be a good replacement for the HCC model because of its excellence in

predictive performance and practical implementation.

The proposed model has better or equal predictive ratios than the HCC model currently applied in Medicare capitation payment system, for those disease groups examined in this dissertation. Predictive ratio is an important factor used to evaluate risk adjustment model performance. Predicted PMPM costs in the proposed model are closer to the observed costs in the aged Medicaid subpopulation than predicted costs using the HCC model. The proposed model also has better performance and accuracy than the HCC model for Florida Medicaid capitation payment system.

The same disease can be associated with differential increases in PMPM costs in the aged and the young disabled subpopulations. The disease groups falling within different levels of increased costs in both subpopulations are listed in Table 8. The increases of PMPM costs for different diseases in the aged and the young disabled subpopulations are divided into five levels: less than $0, $0 to $400, $400 to $700, $700 to $1,000, more than $1,000. Some HCC categories had negative cost increments and these appear in the category <0.

Within the "more than $1,000 increase" category, two diseases can be found for the aged subpopulation and eight diseases for the young disabled subpopulation. For the aged subpopulation, two neurological diseases fell in the category with the highest PMPM cost increase: cerebral palsy and multiple sclerosis. Cerebral palsy is also associated with the second highest PMPM cost increase in the young disabled subpopulation. There is a large gap between younger and older patients' medical costs for multiple sclerosis. Multiple sclerosis increases the PMPM cost by more than $1,000 for the older patients, but falls within the "$0-$400 increase" in the young disabled adults. Multiple sclerosis is costly because it has a major lifetime impact (Whetten-Goldstein et al., 1998). The costs for personal assistant and prescription drugs are major expenses for multiple sclerosis (Henriksson et al., 2001) and older

patients may incur more medical expenditures since they have increased disability and more frequent relapse.

Several diseases fell within the "more than $1,000 increase" level for the young disabled subpopulation. Of these diseases, several can be found in the "$700-$1000 increase" level for the aged subpopulation: quadriplegia, head injury, and respirator dependence. It is likely that medical care for muscle weakness would be associated with high medical costs in both subpopulations since all five diseases mentioned above can lead to respiratory weakness and weakness in other body systems. Muscular dystrophies consist of genetic disorders characterized by progressive muscle wasting and weakness and fell within the "more than $1,000 increase" level for the young disabled subpopulation and the "$0-$400 increase" level for the aged subpopulation. A possible explanation for this finding is that muscular dystrophies can be very severe and affect the respiratory system if they begin in childhood. Older patients with muscular dystrophies have been found to have less severe symptoms and need less medical attention than their younger counterparts. Three medical conditions are associated with more than $1,000 increase in PMPM medical cost for the young disabled subpopulation but less than $700 for the aged subpopulation: artificial openings for feeding or elimination, severe hematological disorders, and dialysis status. Patients with these conditions usually need continuous transfusion of blood or feeding devices, which can be costly. The results indicate that younger disabled Medicaid enrollees need more medical care in transfusion and feeding.

Table 8: Disease Groups in Different PMPM Cost Increase Categories for the Aged and the Young Disabled Subpopulation

| PMPM Cost Increase Categories | Aged | Younger Disabled |
|---|---|---|
| >1,000 | Cerebral Palsy | Quadriplegia |
| | Multiple Sclerosis | Cerebral Palsy |
| | | Muscular Dystrophy |
| | | Respirator Dependence |
| | | Artificial Openings for Feeding or Elimination |
| | | Severe Hematological Disorders |
| | | Dialysis Status |
| | | Severe Head Injury |
| 700-1000 | Quadriplegia | Metastatic Cancer and Acute Leukemia |
| | Schizophrenia | HIV/AIDS |
| | Respirator Dependence | Opportunistic Infections |
| | Major Head Injury | Paraplegia |
| | Diabetes with Renal or Peripheral Circulatory Manifestation | End-Stage Liver Disease |
| | Protein-Calorie Malnutrition | Protein-Calorie Malnutrition |
| | | Aspiration and Specified Bacterial Pneumonias |
| | | Decubitus Ulcer of Skin |

| | | |
|---|---|---|
| 400-700 | Aspiration and Specified Bacterial Pneumonias | Amputation Status |
| | HIV/AIDS | Seizure Disorders and Convulsions |
| | Parkinson's and Huntington's Diseases | Hemiplegia |
| | Paraplegia | Septicemia/Shock |
| | Seizure Disorders | Chronic Ulcer of Skin |
| | Ischemic Stroke | Coma |
| | Decubitus Ulcer of Skin | Disorders of Immunity |
| | Vascular Disease | Lung, Upper Digestive Tract, and Other Severe Cancers |
| | Artificial Openings for Feeding or Elimination | Renal Failure |
| | Dialysis Status | Vascular Disease with Complications |
| | Hemiplegia | Respiratory Arrest |
| | Coma | Intestinal Obstruction/Perforation |
| | Amputation Status | Diabetes with Renal or Peripheral Circulatory Manifestation |
| | Drug/Alcohol Psychosis | Schizophrenia |
| | | Parkinson's and Huntington's Diseases |

| | | |
|---|---|---|
| 0-400 | Major Depressive, Bipolar, and Paranoid Disorders | Congestive Heart Failure |
| | Intestinal Obstruction | Spinal Cord Disorders/Injuries |
| | Renal Failure | Diabetes with Acute Complications |
| | Chronic Ulcer of Skin | Diabetes with Neurologic or Other Specified Manifestation |
| | Septicemia/Shock | Multiple Sclerosis |
| | Major Organ Transplant Status | Cardio-Respiratory Failure and Shock |
| | Vascular Disease with Complications | Ischemic Stroke |
| | Hip Fracture | Vertebral Fractures without Spinal Cord Injury |
| | Metastatic Cancer and Acute Leukemia | Pneumococcal Pneumonia |
| | Major Complications of Medical Care and Trauma | Hip Fracture/Dislocation |
| | Diabetes with Ophthalmologic or Unspecified Manifestation | Major Complications of Medical Care and Trauma |
| | Diabetes with Neurologic or Other Specified Manifestation | Vascular Disease |
| | Disorders of Immunity | Diabetes with Ophthalmologic or Unspecified Manifestation |
| | Congestive Heart Failure | Chronic Obstructive Pulmonary Disease |
| | Severe Hematological Disorders | Chronic Hepatitis |
| | Spinal Cord Disorders/Injuries | Cystic Fibrosis |
| | Vertebral Fractures without Spinal Cord Injury | Acute Myocardial Infarction |
| | Cirrhosis of Liver | Bone/Joint/Muscle Infections/Necrosis |

| | | |
|---|---|---|
| 0-400 | Cardio-Respiratory Failure and Shock | Specified Heart Arrhythmias |
| | Diabetes with Acute Complications | Pancreatic Disease |
| | Traumatic Amputation | Polyneuropathy |
| | Specified Heart Arrhythmias | Diabetes without Complication |
| | Inflammatory Bowel Disease | Nephritis |
| | Pancreatic Disease | Traumatic Amputation |
| | Bone/Joint/Muscle Infections/Necrosis | Drug/Alcohol Dependence |
| | Respiratory Arrest | Major Head Injury |
| | Cerebral Hemorrhage | Unstable Angina and Other Acute Ischemic Heart Disease |
| | Muscular Dystrophy | Lymphatic, Head and Neck, Brain, and Other Major Cancers |
| | Nephritis | Rheumatoid Arthritis and Inflammatory Connective Tissue Disease |
| | End-Stage Liver Disease | Breast, Prostate, Colorectal and Other Cancers and Tumors |
| | | Cirrhosis of Liver |
| | | Major Organ Transplant Status |

| | | |
|---|---|---|
| <0 | Drug or Alcohol Dependence | Drug/Alcohol Psychosis |
| | Proliferative Diabetic Retinopathy and Vitreous Hemorrhage | Cerebral Hemorrhage |
| | Rheumatoid Arthritis and Inflammatory Connective Tissue Disease | Proliferative Diabetic Retinopathy and Vitreous Hemorrhage |
| | Breast, Prostate, Colorectal and Other Cancers and Tumors | Angina Pectoris/Old Myocardial Infarction |
| | Acute Myocardial Infarction | Major Depressive, Bipolar, and Paranoid Disorders |
| | Pneumococcal Pneumonia | Inflammatory Bowel Disease |
| | Unstable Angina and Other Acute Ischemic Heart Disease | Extensive Third-Degree Burns |
| | Lung, Upper Digestive Tract, and Other Severe Cancers | |
| | Lymphatic, Head and Neck, Brain, and Other Major Cancers | |
| | Chronic Hepatitis | |
| | Angina Pectoris/Old Myocardial Infarction | |
| | Polyneuropathy | |
| | Opportunistic Infections | |
| | Chronic Obstructive Pulmonary Disease | |
| | Cystic Fibrosis | |

Note: Diseases in each PMPM cost categories are ranked by amount of costs.

Diseases in the "$700-$1000 increase" level are mostly of an infectious nature for the young disabled subpopulation. Four diseases in this level are directly related to opportunistic infections: HIV/AIDS, opportunistic infections, pneumonias, and decubitus ulcer of skin. Some diseases in this category have indirect association with infections. For example, patients with acute leukemia may have low numbers of white cells and their risk for opportunistic infections increases dramatically. End-stage liver disease and protein-calorie malnutrition lead to low protein levels and a decline in the blood antibody level coincident with a fall in the body protein level. Opportunistic infections do not increase the medical cost, as predicted by the proposed model. The results indicate that medical care for infections cost less in older ages.

For the aged subpopulation, most diseases in the "$700-1000 increase category" are neurological disorders. Schizophrenia, which is a type of mental disorder, is also associated with high medical costs. In the young disabled subpopulation, it falls within the "$400-$700" categories.

Most diseases are associated with higher predicted PMPM medical costs in the young disabled subpopulation than in the aged subpopulation. This finding may be due to the fact that younger patients suffer more severe medical conditions than older patients who have the same diseases. For example, younger patients need higher doses of prescription drugs than older patients because they more rapidly metabolize medications. Moreover, Medicaid pays for almost all medical costs for younger disabled adults, but it only reimburses the costs of prescription drugs for the aged subpopulation since older Medicaid beneficiaries are likely to be eligible for Medicare and Medicare reimburses many health care services. Hence, PMPM costs of the older Medicaid beneficiaries are likely to be underpredicted since Medicaid claims do not include all medical expenditures.

Diseases associated with high PMPM costs merit further attention from health care providers and investigators. These diseases place a heavy financial burden on the Medicaid capitation payment system. Health care providers who find cost-saving methods to manage patients with these diseases will reap significant financial returns.

The proposed model represents an improvement of 15% in variance accounted for over the current HCC model for the Medicaid capitation system for the aged subpopulation, since the proposed model $R^2$ was 0.32 while the HCC model was 0.17. However, several changes can be made which may yield further improvements in the proposed model. The proposed model has a larger RMSE value for the aged and the young disabled subpopulations than the general Medicaid population, which may indicate that additional predictors associated with medical costs can be added into the proposed model to improve performance. Some pharmacy-based adjusters were not included in the proposed model because of potential interactions among the current set of predictors, however a possible solution is the statistical transformation of these adjusters. Moreover, current RxRisk adjusters were developed for chronic conditions only; new RxRisk adjusters for acute conditions should be developed to increase the predictive accuracy of models with pharmacy-based adjusters. The procedure codes reported in Medicaid claims may contribute unique variance to the proposed model and significantly affect the prediction of medical costs.

In this dissertation, R-square value and RMSE were used to evaluate the predictive power of risk adjustment models, and they are both statistical indicators of the overall predictive performance of risk adjustment models. For example, a risk adjustment model with high R-square value and low RMSE can predict more variance of medical expenditures in the next year than models having lower R-square values and higher RMSE. The predictive ratio was used to estimate the power of models

using individual risk adjusters and the proposed model had a better predictive ratio than the HCC model for the aged subpopulation. However, both models had weak predictive performance for some risk adjusters including end-stage liver disease and metastatic cancer. In future research some improvements can be made to the proposed model. There were two types of health status predictors in the proposed model: diagnosis-based HCC adjusters and pharmacy-based RxRisk adjusters. Each HCC adjuster represented a group of diseases associated with similar clinical characteristics and costs. The distribution of each HCC disease group should be examined to find extremely high or low values (outliers). The results show that the proposed model poorly predict the costs of some disease groups, such as end-stage liver disease and metastatic cancer and a large number of outliers may play a role in poor predictive performance of the proposed model for end-stage liver diseases and metastatic cancers (Edwards et al., 2000). For example, patients with extremely high medical costs within a disease group can make the average cost of this disease group increase sharply and risk adjustment models will probably overpredict medical costs for other patients in this group. Two possible solutions are the removal of these extreme values and the addition of new variables as substitutes for these extreme values. The removal of extreme values can decrease variance within a disease group. The addition of new variables can divide medical costs for a disease group into two groups: extreme values and non-extreme values so researchers can estimate average medical costs for both groups, add them into risk adjustment models respectively, and address the need of patients with both extreme and non-extreme values. Predictive ratio is an important indicator for evaluation of predictive performance of models for individual risk adjusters, and both solutions mentioned in the early context should be attempted to make the predictive ratio as close to 1 as is possible.

Second, adding more variables improves the predictive performance of risk adjustment models. Risk adjusters based on durable medical equipment (DME), may be a new class of health condition variables. The definition of DME varies across state Medicaid systems which have set different standards for DME (www.nls.org). Generally, equipment must meet four criteria to be considered DME: (1) it can be used repeatedly over extended intervals; (2) it is serves a medical purpose; (3) it is functionally useless to persons not having the illness or injury which it was designed to ameliorate; (4) it is appropriate for usage at home. These four criteria are the only criteria for DME in Florida but different criteria are found for other states. For example, the Georgia Medicaid system requires that DME also must have a warranty. The Connecticut Medicaid system requires that DME also must be non-disposable. The criteria for DME can be found in state Medicaid manuals in different states. DME can provide more information on patients' health conditions that may be not accounted for by diagnosis or pharmaceutical adjusters. For example, wheelchair is a DME in Florida. Patients who are in need of wheelchairs may have worse health condition than patients without walking difficulties. Hence, patients using wheelchairs may incur higher medical costs than their counterparts without walking difficulties. However, all these patients are assumed to have same medical costs under risk adjustment models without wheelchair variables. In future research using this model, all DME will be reviewed and DME that have effects on future medical costs will be added into the proposed model to attempt to increase predictive power.

Future research will concentrate on improving risk adjustment models and the application of improved risk adjustment models in the Florida Medicaid capitation payment system. Improvement of risk adjustment models involves empirical reviews of statistical methods and clinical practice. An ideal risk adjustment model must have strong predictive power and easy implementation in clinical practice. As mentioned

earlier, the addition of new variables and removal of outliers may increase predictive accuracy of the proposed model. However, the proposed model's practical implementation must be addressed. The collection of diagnosis-based and pharmacy-based data may be time-consuming for health plans since both diagnoses and prescription drug records must be obtained and used to calculate predicted medical costs. Capitation systems based on the proposed model may meet resistance from health plans because of the increased workload imposed by this method. Possible solutions include the simplification of data-reporting processes for health plans and innovations in computer software that can handle more complicated medical costs calculations. Capitation payment systems based on the proposed model will have several advantages: (1) the proposed model can predict medical costs more accurately. The proposed model has more risk adjusters than the HCC model and the RxRisk model, and the results in this dissertation have shown that it had more predictive power statistically than the competing models. Predicted medical costs based on the proposed model will be closer to actual medical costs than medical costs predicted by other models; (2) the proposed model can decrease the possibility of gaming since it utilizes both diagnostic and pharmaceutical information to predict future medical costs. Prescription drug usage can provide more information on patients' health conditions that cannot be inferred from diagnoses. Hence, patients' health condition will be more easily identified using the proposed model compared with the HCC model. Health plans will have fewer opportunities to game the capitation payment system.

The RxRisk model is approximately equal to the proposed model in predictive accuracy for the aged Medicaid subpopulation. It has advantages in practical implementation, such as less work of collecting data and high reliability of data. However, it is suggested as a strategy for the aged enrollees only since the

predictive accuracy of the RxRisk model for the young disabled subpopulation was significantly lower than the proposed model in this study.

The proposed model is recommended to be the risk adjustment model for the Florida Medicaid capitation payment system because it has been shown to outperform competing alternatives. Generally, the proposed model outperformed current risk adjustment models in capitation payment systems. As national health expenditures keep rising in the near future, improvement of capitation payment systems based on risk adjustment methodology is necessary, and this dissertation proposes a better model than current ones. It can be implemented in the Florida Medicaid capitation payment system to improve predictive accuracy and save medical expenditures.

List of References

Ash S., Ellis R., Pope G., and Ayanian J. (2000). *Using diagnoses to describe populations and predict costs.* Health Care Financing Review, 21(3), 7-28.

Ash, A., Porell, F., Gruenberg, L., and others (1989). *Adjusting Medicare capitation payments using prior hospitalization*. Health Care Financing Review, 10(4):17-29.

Beebe, J., Lubitz, J., and Eggers, P. (1985). *Using prior utilization to determine payments for Medicare enrollees in health maintenance organizations.* Health Care Financing Review, 6(3), 27-38.

Brown, R. (1988). *Biased selection in Medicare HMOs.* Paper presented at the Fifth Annual Meeting of the Association for Health Services Research, San Francisco, CA, July 26-28.

Brown, R.S., Bergeron, J.W., Clement, D.G., Hill, J.W. and Retchin, S.M. (1993). *The Medicare risk program for HMOs: final summary report on findings from the evaluation*. Contract no. 500-88-006. Princeton, NJ: Mathematica Policy Research.

Brown, R.S., Clement, D.G., Hill, J.W., Retchin, S.M., and Bergeron, J.W. (1993). *Do health maintenance organizations work for Medicare*? Health Care Financing Review, 15(1), 7-23.

Brown, R.S. and Langwell, K. (1988). *Enrollment patterns in Medicare HMOs: implications for access to care*. Advances in Health Economics and Health Services Research, 9, 69-96.

Clark, D.O., von Korff, M., Saunders, K., Baluch, W.M., and Simon, G.E. (1995). *A chronic disease score with empirically derived weights.* Medical Care, 33, 783-795.

Dranove, D. (1989). *Medicaid drug formulary restrictions*. Journal of Law and Economics, 32(1), 143-162.

Edwards K, Gifford G, and Knutson D. (2000). *Risk adjustment and the impact of outliers on prospective risk models: differential methods of determining where, when and how to cut the edge.* The Academy for Health Services Research and Health Policy Meeting, 17, unknown.

Eggers, P.W. and Prihoda, R. (1982). *Pre-enrollment reimbursement patterns of Medicare beneficiaries enrolled in "at-risk" HMOs.* Health Care Financing Review, 4(1), 55-73.

Ellis, R.P. and Ash, A. (1995). *Refinements to the diagnostic cost group model*. Inquiry, 32(4), 1-12.

Ellis, R.P., Pope, G.C., Iezzoni, L.I., and others (1996). *Diagnosis-based risk adjustment for Medicare capitation payments*. Health Care Financing Review, 17(3), 101-128.

Epstein, A.M. and Cumella, E.J. (1988). *Capitation payment: using predictors for medical utilization to adjust rates*. Health Care Financing Review, 10(1), 51-69.

Ettner, S.L., Frank, R.G., Mark, T., and Smith, M.W. (2000). *Risk adjustment of capitation payments to behavioral health care carve-outs: how well do existing methodologies account for psychiatric disability?* Health Care Management Science, 3(2), 159-169.

Fishman, P.A., Goodman, M.J., Hornbrook, M.C., Meenan, R.T., Bachman, D.J., and O'Keeffe Rosetti, M.C. (2003). *Risk adjustment using automated ambulatory pharmacy data: the RxRisk model.* Medical Care, 41(1), 84-99.

Fishman, P. A. and Shay, D.K. (1999). *Development and estimation of a pediatric chronic disease score using automated pharmacy data.* Medical Care,37, 874-883.

Fowles, J.B., Weiner, J.P., Knutson, D., Fowler, E., Tucker, A.M., and Ireland, M. (1996). *Taking health status into account when setting capitation rates: a comparison of risk-adjustment methods*. JAMA, 276(16), 1316-1321.

Gilmer, T., Kronick, R., Fishman, P., and Ganiats, T.G. (2001). *The Medicaid Rx model: pharmacy-based risk adjustment for public programs.* Medical care, 39(11), 1188-1202.

Gruenberg, L., Kaganova, E., and Hornbrook, M.C. (1996). *Improving the AAPCC (adjusted average per capita cost) with health-status measures from the MCBS (Medicare Current Beneficiary Survey).* Health Care Financing Review, 17(3), 59-75.

Health Care Financing Administration, Office of Strategic Planning (1999). *Report to Congress: proposed method of incorporating health status risk adjusters into Medicare+choice payments.* Baltimore, MD.

Heffler, S., Smith, S., Won, G., Clemens, K., Keehan, S., and Zezza, M. (2002). *Health spending projections for 2001-2011: the latest outlook*. Health Affairs, 21(2), 207-218.

Heffler, S., Smith, S., Keehan, S., Borger, C., Clemens, K., and Truffer, C. (2005). *Trends: U.S. health spending projections for 2004-2014*. Health Affairs Web Exclusive.

Henriksson, F., Fredrikson, S., Masterman, T., and Jönsson, B. (2001). *Costs, quality of life and disease severity in multiple sclerosis: a cross-sectional study in Sweden*. European journal of neurology, 8(1), 27-35.

Hill, J.W. and Brown, R.S. (1990). *Biased selection in the TEFRA HMO/CMP program.* Princeton, NJ: Mathematical Policy Research.

Hornbrook, M.C. and Goodman, M.J. (1996). *Chronic disease, functional health status, and demographics: a multi-dimensional approach to risk adjustment*. Health Service Research, 31(3), 283-307.

Kronick, R., Dreyfus, T., Lee, L., and Zhou, Z. (1996). *Diagnostic risk adjustment for Medicaid: the disability payment system*. Health Care Financing Review, 17(3), 7-33.

Kronick, R., Gilmer, T., Dreyfus, T., and Lee, L. (2000). *Risk adjustment - improving health-based payment for Medicaid beneficiaries: CDPS.* Health Care Financing Review, 21(3), 29-63.

Levit, K., Smith, C., Cowan, C., Lazenby, H., Sensenig, A., and Catlin, A. (2003). *Trends in U.S. health care spending, 2001*. Health Affairs, 22(1), 154-164.

Lichtenstein, R. and Thomas, J.W. (1987). *Including a measure of health status in Medicare's health maintenance organization capitation formula: reliability issues.* Medical Care, 25(2), 100-110.

Lichtenstein, R., Thomas, J.W., Adams-Watson, J., Lepkowski, J., and Simone, B. (1991). *Selection Bias in TEFRA At-Risk HMOs.* Medical Care, 29(4), 318-331.

Lubitz, J. and Prihoda, R. (1984). *The use and costs of Medicare services in the last 2 years of life*. Health Care Financing Review, 5(3), 117-131.

McCall, N. and Korb, J. (1998). *Risk adjustment for dually eligible beneficiaries using long-term care*. Health Care Financing Review, 20(2), 71-90.

Meenan, R.T., Goodman, M.J., Fishman, P.A., Hornbrook, M.C., O'Keeffe-Rosetti, M.C., and Bachman, D.J. (2003). *Using risk-adjustment models to identify high-cost risks.* Medical Care, 41(11), 1301-1312.

Mello, M.M., Stearns, S.C., Norton, E.C., and Ricketts, T.C. 3rd. (2003). *Understanding biased selection in Medicare HMOs.* Health Services Research, 38(3), 961-92.

Miller, R.H. and Luft, H.S. (1997). *Does managed care lead to better or worse quality of care?* Health Affairs, 16(5), 7-25.

Moore, W.J. and Newman, R.J. (1993). *Drug formulary restrictions as a cost-containment policy in Medicaid programs*. Journal of Law and Economics, 36(1), 71-97.

Morgan, R. O., Virnig, B. A., DeVito, C. A., and Persily, N. A. (1997). *The Medicare HMO revolving door--The healthy go in and the sick go out.* New England Journal of Medicine, 337(3), 169-175.

Newhouse, J. P. (1986). *Rate adjusters for Medicare under capitation.* Health Care Financing Review Annual Supplement, 8, 45-56.

Newhouse, J. P. and others (1989). *Adjusting capitation rates using objective health measures and prior utilization.* Health Care Financing Review, 10(3), 41-54.

Pacala, J.T., Boult, C., Urdangarin, C., and McCaffrey, D. (2003). *Using self-reported data to predict expenditures for the health care of older people.* Journal of the American Geriatrics Society, 51(5), 609-614.

Pope, G.C., Ellis, R.P, Ash, A.S., and others (2000). *Principal inpatient diagnostic cost group models for Medicare risk adjustment.* Health Care Financing Review, 21(3), 93-118.

Putnam, K.G., Buist, D.S.M., Fishman, P., Andrade, S.E., Boles, M., Chase, G.A., Goodman, M.J., Gurwitz, J.H., Platt, R., Raebel, M.A., and Chan, K.A. (2002). *Chronic disease score as a predictor of subsequent hospitalization: a multiple HMO study.* Epidemiology, 13, 340-344.

Riley, G.F. (2000). *Risk adjustment for health plans disproportionately enrolling frail Medicare beneficiaries*. Health Care Financing Review, 21(3), 135-148.

Robinson, J. and Karon, S. L. (2000). *Modeling Medicare costs of PACE populations.* Health Care Financing Review, 21(3), 149-170.

Sales, A.E., Liu, C.F., Sloan, K.L., and others. (2003). *Predicting costs of care using a pharmacy-based measure risk adjustment in a veteran population.* Medical care, 41(6), 753-761.

Sloan, K.L., Sales, A.E., Liu, C.F., Fishman, P., Nichol, P., Suzuki, N.T., and Sharp, N.D. (2003). *Construction and characteristics of the RxRisk-V: a VA-adapted pharmacy-based case-mix instrument.* Medical Care, 41(6), 761-774.

Temkin-Greener, H., Meiners, M. R., and others (2001). *PACE and the Medicare+choice risk-adjusted payment model.* Inquiry, 38(1), 60-72.

Tollen, L. and Rothman, M. (1998). *Case study: Colorado Medicaid HMO risk adjustment*. Inquiry, 35(2), 154-170.

Von, Korff M, Wagner, E.H., and Saunders, K.A. (1992). *Chronic disease score from automated pharmacy data.* Journal of Clinical Epidemiology, 45(2), 197-203.

Weil, A. (2003). *There's something about Medicaid*. Health Affairs, 22, 13-30.

Weiner, J.P., Tucker, A.M., Collins, A.M., Fakhraei, H., Lieberman, R., Abrams, C., Trapnell, G.R., and Folkemer, J.G. (1998). *The development of a risk-adjusted capitation payment system: the Maryland Medicaid model*. Journal of Ambulatory Care Management, 21(4), 29-52.

Whetten-Goldstein, K., Sloan, F.A., Goldstein, L.B., and Kulas, E.D. (1998). A comprehensive assessment of the cost of multiple sclerosis in the United States. Multiple Sclerosis, 4(5), 419-425.

http://www.nls.org/conf2005/MA%20DME%20definitions.htm

List of Bibliography

Ash, A. and Byrne-Logan, S. (1998). *How well do models work? predicting health care costs*. Proceedings of the Section on Statistics in Epidemiology of the American Statistical Association. Dallas.

Ash, A., Ellis, R.P., Yu, W., and others (1998). *Risk adjusted payment models for the non-elderly.* Final Report to the Health Care Financing Administration under Contract Number 18-C-90462/1-02. Boston University. Boston.

Ash, A., Porell, F., Gruenberg, L, and others (1986). *An analysis of alternative AAPCC models using data from the continuous Medicare history sample*. Final Report to the Health Care Financing Administration. Health Policy Research Consortium. Brandeis/Boston Universities. Boston.

Greenwald, L.M., Esposito, A., Ingber, M.J., and Levy, J.M. (1998). *Risk adjustment for the Medicare program: lessons learned from research and demonstrations*. Inquiry, 35(2), 193-209.

Iezzoni L.I. (1997). *Risk adjustment for measuring health care outcomes*. Health Administration Press. Ann Arbor, Michigan.

Iezzoni, L.I., Ayanian, J.Z., Bates, D.W., and Burstin, H.R. (1998). *Paying more fairly for Medicare capitated care.* New England Journal of Medicine, 339(26), 1933-1938.

Ingber, M.J. (1998). *The current state of risk adjustment technology for capitation*. Journal of Ambulatory Care Management, 21(4), 1-28.

Pope, G.C., Ellis, R.P., Liu, C.F., and others (1998). *Revised diagnostic cost group (DCG)/hierarchical coexisting conditions (HCC) models for Medicare risk adjustment.* Final Report to the Health Care Financing Administration under Contract Number 500-95-048. Waltham, MA. Health Economics Research, Inc.

Pope, G.C., Kautter, J., Ellis, R.P., Ash, A.S., Ayanian, J.Z., Iezzoni, L.I., Ingber, M.J. Levy, J.M. and Robst, J. (2004). *Risk adjustment of Medicare capitation payments using the CMS-HCC model.* Health Care Financing Review, 25(4), 119-142.

Public Health Service and Health Care Financing Administration (1980). *International classification of diseases, 9th Revision, clinical modification.* U.S. Department of Health and Human Services. U.S. Government Printing Office. Washington, DC.

Van de Ven, W.P.M.M., and Ellis, R.P. (2000). *Risk adjustment in competitive health plan markets.* In Culyer, A.J., and Newhouse, J.P., eds.: Handbook in Health Economics. North Holland.

About the Author

Yanen Li received a Bachelor's Degree in Clinical Medicine from Jiangxi Medical College in 1999 and a MPH in Biostatistics from University of South Florida in 2003. He entered the Ph.D. program in Aging Studies at the University of South Florida in 2001.

While in the Ph.D. program at the University of South Florida, Mr. Li was very active in School of Aging Studies. He made several poster presentations at national meetings of the Gerontological Society of America. He works as a data analyst for the Louis de la Parte Florida Mental Health Institute's Policy and Services Data Center currently.