DEVELOPMENT AND VALIDATION OF RISK ASSESSMENT MODELS TO PREDICT PHARMACY EXPENDITURES FOR BOTH COMMERICAL AND MEDICAID POPULATIONS

by

CHRISTOPHER RON CANTRELL

(Under the Direction of Bradley C. Martin)

ABSTRACT

Numerous risk assessment models have been developed to predict overall healthcare costs, utilization, mortality and hospital length of stay, however, there have been no models to date developed specifically to predict prescription expenditures. The objective of this research was to empirically develop a claims-based risk assessment model to predict prescription expenditures for both a commercial and Medicaid population.

The models were developed using three years, 1998 through 2000, of MEDSTAT MarketScan data (commercial) and California Medicaid data (Medicaid). Both datasets are claims-based data that include medical and pharmacy claims and enrollment information in a linkable format. The MarketScan training sample used to develop the commercial models included over 1.3 million lives after the inclusion/exclusion criteria were satisfied. The California Medicaid (MediCal) training sample used to develop the Medicaid models included over 138 thousand lives after the inclusion/exclusion criteria were satisfied. A random sample of each dataset was used to validate the models.

Ordinary Least Squares (OLS) was utilized to estimate the model coefficients. The primary model for this research, the Rx Cost Model (RxCost), is a diagnostic-based model that was empirically developed using diagnostic information. Another model, the Mixed Rx Cost Model (MRxCost), is a diagnostic and drug-based model that was developed to explore the gain in predictive power of supplementing the RxCost Model with drug information.

The MarketScan validation sample was utilized to compare the performance of the models developed for the commercial population to each other as well as a Demographic-only model and the commercially available DCG-HCC model. The MediCal validation sample was used to compare the performance of the models developed for the Medicaid population to each other as well as a Demographic-only model, a Demographic and Medicaid eligibility model and the CDPS model.

The R-square values for the commercial RxCost Model, the MRxCost Model and the DCG-HCC using the validation sample were 0.22, 0.34 and 0.16 respectively. The R-square values for the Medicaid RxCost Model, the MRxCost Model and the CDPS using the validation sample were 0.24, 0.30 and 0.04 respectively. The RxCost Model for both the commercial and Medicaid population performed better then the DCG-HCC and the CDPS in terms of R-square. The MRxCost Model for each population also performed well and resulted in a substantial gain of predictive power in terms of R-square.

INDEX WORDS: Risk Assessment, Risk Adjustment, Prescription Costs, Drug Costs, Risk Models

DEVELOPMENT AND VALIDATION OF RISK ASSESSMENT MODELS TO PREDICT PHARMACY EXPENDITURES FOR BOTH COMMERICAL AND MEDICAID POPULATIONS

by

CHRISTOPHER RON CANTRELL

BACHELORS OF SCIENCE, UNIVERISITY OF GEORGIA, 1997

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2004

DEVELOPMENT AND VALIDATION OF RISK ASSESSMENT MODELS TO PREDICT PHARMACY EXPENDITURES FOR BOTH COMMERICAL AND MEDICAID POPULATIONS

by

CHRISTOPHER RON CANTRELL

Major Professor: Bradley C. Martin

Committee: Jeffrey A. Kotzan
Jeffrey H. Dorfman
Joseph T. Dipiro
William E. Wade

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
May 2004

ACKNOWLEDGEMENTS

I would like to thank Dr. Martin for his continued efforts in improving the quality of this research and providing guidance whenever needed. I would also like to thank him for his dedication to the graduate program and for providing me with years of training, mentoring and friendship. Thanks to my committee members, Dr. Kotzan, Dr. Wade, Dr. Dipiro and Dr. Dorfman for agreeing to participate in this project and extending help whenever needed. Heartfelt thanks goes to my loving family that supported and encouraged me throughout my entire graduate experience. Also, special thanks to my fellow departmental classmates that shared the experience with me some of whom became close friends in the process.

TABLE OF CONTENTS

LIST OF TABLES

Page

CHAPTER 1

INTRODUCTION

Over the past decade, prescription drug expenditures have risen dramatically. As health plans pay a growing share of these rising pharmacy costs, it is important that they manage their prescription drug benefits effectively and efficiently. Health plans have adopted a variety of techniques over the past decade to manage these costs and continue to look for new innovative methods. Recently, there has been a growing trend towards rewarding physicians financially for providing quality care. This trend has increasingly targeted physicians for cost containment. Physician profiling has become a mainstay for ensuring efficient prescribing. More recently, a few health plans have shifted a portion of prescription costs to the physician and some have considered capitating physicians based on prescription expenditures. To be meaningful, equitable, and to prevent physicians from cherry picking the healthiest patients, any effort to profile or capitate physicians should account for differences in the clinical characteristics of patient populations. Risk assessment techniques can be utilized to account for differences in patient case-mix before making inferences regarding the effectiveness of care. Currently, however, there is no published literature on risk assessment models

specifically for prescription expenditures. As prescription expenditures continue to be a focus of attention, the need for prescription cost risk assessment models becomes more profound.

The objective of this research was to empirically develop claims-based risk assessment models to predict pharmacy expenditures for both commercial and Medicaid populations. More specifically, risk assessment indices were developed based on

1) demographic information
2) demographic information and Medicaid Eligibility information (Medicaid population only)
3) diagnostic codes
4) combined information from diagnostic codes and prescription drugs.

The indices were adapted from previous work on risk assessment (Diagnostic Cost Group Hierarchical Condition Category, Chronic Illness and Disability Payment System, and Chronic Disease Score) and utilized the Clinical Classifications Software developed by AHRQ, a diagnosis categorization scheme, to develop risk assessment models specific for prescription expenditures. The models were developed in a prospective fashion where information in year 1 was utilized to predict year 2 expenditures.

Prescription cost risk assessment models were developed separately for two large segments of the adult non-elderly health care market; the private commercially insured population and the Medicaid population. These populations represent the vast majority of all people in the U.S. who have health insurance that also covers

prescription drugs. Data from the MEDSTAT MarketScan database and the California Medicaid administrative claims database were used to develop and validate the risk assessment models for these populations.

The specific aims of this research were to:

1. Develop a new code-based ICD-9-CM prescription specific risk assessment model based on AHRQ's Clinical Classification Software.

2. Develop and validate a new prescription specific risk assessment model based on combined information from ICD-9-CM codes and drug markers (mixed model) based on an updated Chronic Disease Score (CDS) developed by Von Korff.

Additional aims of this research were to:

- Validate the new indices on a random holdout sample of each population.
- Assess and compare the performance of the risk assessment models separately using the MEDSTAT MarketScan data and the California Medicaid data.
- Compare the performance of the new empirically derived models to the performance of Ash's Diagnostic Cost Group Hierarchical Condition Category (DCG-HCC) for the commercial population and to Kronick's Chronic Illness and Disability Payment System (CDPS) for the Medicaid population.

CHAPTER 2

BACKGROUND AND SIGNIFICANCE

PRESCRIPTION EXPENDITURES

As health plans strive to control their medical costs while improving the quality of care provided to their patients, they will increasingly focus on the most rapidly growing component of their cost structure – drugs. (Litton 2000) National attention will increasingly focus on prescription drugs as pharmacy costs are rising in excess of general and medical cost inflation. Although outpatient prescription drug spending represented only a small portion (11% or $140.6 billion) of personal drug spending, it was the fastest growing component of health care. (The Kaiser Family Foundation 2003) Spending on prescription drugs in the U.S. rose at double-digit rates throughout the past decade. (Employee Benefit Research Institute 1999) Between 1995 and 1998, prescription expenditures grew nearly 50% while expenditures for physician services and hospital care grew only 14% and 10% respectively. (The Kaiser Family Foundation 2000) From 2000 to 2001, the U.S. prescription drug expenditure continued to outgrow the expenditure for hospital care and physician services, however the rates were not quite as dramatic; 16% for prescription drugs vs. 8% and 9% for hospital care and physician services. (The Kaiser Family Foundation 2003)

From 1997 to 2001, three primary factors are attributed to the growth in prescription drug expenditure. The increasing number of prescriptions dispensed (utilization) was responsible for most of the increase in drug expenditure (47%). (The Kaiser Family Foundation 2003) The number of drugs dispensed per capita in the U.S. rose from 8.4 in 1995 to 10.9 in 2001. (The Kaiser Family Foundation 2003) Several factors may account for the increase in prescription utilization and spending including an aging U.S. populace and an accelerated rate of new drug discoveries and approvals offering new and better treatments. (IMS Health 1999; The Kaiser Family Foundation 2000) Increases in the average life expectancy and the aging baby boomers will further propel prescription expenditures higher. The second factor contributing 27% to rising drug expenditures was the types of drugs used with newer higher priced drugs replacing older, less-expensive drugs. Price increases for existing drugs, although the least of the three factors, contributed for 26% of the rising drug expenditure. From 2000 to 2001, the average retail price of a brand name drug rose 9% and the average retail price of a generic prescription drug rose 14%.

PRESCRIPTION BENEFIT MANAGEMENT

As a result of the rising cost of drugs and the increased utilization, especially of newer, more expensive drugs, managed care organizations (MCOs) and pharmacy benefit managers (PBMs) have explored various strategies for managing pharmacy benefits. Alternatives for controlling drug costs and utilization have mainly consisted of: restricting patient access, shifting costs to enrollees, sharing cost increases through higher premiums, and shifting risk to providers. (Kleinke 2001) Formularies, prior authorizations, step therapy, limits, and higher premiums and copays have all been

utilized to restrict patient access and shift costs to enrollees. These measures have all been shown to be effective in reducing pharmaceutical costs. (Galt 2001; Motheral 1999)

Although the techniques described above remain an important component in managing prescription expenditures, another more innovative way to control drug costs and utilization has been to shift the risk to providers. There is a growing trend toward paying more for higher quality. Two such movements of late are the Pay for Performance (P4P) initiative in California and the Bridges to Excellence (BTE) program for diabetes care. (National Committee for Quality Assurance 2003, Bridges to Excellence 2004)

The P4P program is a statewide initiative in California that uses standard measures (six HEDIS measures) to evaluate the performance of physician organizations. Physicians will be given bonuses based on their HEDIS scores in six areas: childhood immunizations, breast cancer screening, cervical cancer screening, appropriate medications for asthma, cholesterol management after acute cardiovascular events and comprehensive diabetes management. There are currently six health plans representing over 7 million commercial enrollees participating in the program. The BTE is a program that rewards good diabetes care. The program requires physicians to be recognized as providing good quality diabetes care. To do this, a physician must be recognized through the Diabetes Physician Recognition Program; a program that was a joint effort between NCQA and the American Diabetes Association. (NCQA 2004) Once recognized as a quality provider, a physician receives a bonus of $100 per diabetic

patient. This growing trend toward incentive for quality has already been tried (on a limited basis) as a means to reward physicians for judicious prescribing behaviors.

These types of provider incentives can potentially be very influential in terms of promoting judicious prescribing behaviors. Here, the incentives are positioned at the point of care where prescribing decisions are initially made. Other incentive programs where the financial burden is shifted to patients, such as differential copays or higher coinsurance, places the financial incentive beyond the initial prescribing decision. Here, patients are not in a position to evaluate alternative substitute drugs, however they are held accountable financially for the decisions their physicians make. Physicians are the key decision makers in selecting prescription products but bear none or very little of the prescription costs.

Physician profiling and physician capitation have both been used in the past as a means of shifting risk to providers. (Burton 2001; Carroll 2000; United 2003) Physician profiling is the comparison of physician practice patterns to determine the existence and effects of significant differences in outcomes. (Tucker 2000) It seeks to isolate variances in outcomes that are attributable to provider choices or treatment modalities. To optimize the quality of care, these comparisons can be used to influence provider behavior through awareness, education and even financial consequences. Physician profiling has been used extensively and consistently yields better patient outcomes and lower costs without reducing patient access. (Berlowitz 1998; DeLiberty 1998; DeLong 1997; Peterson 1998;) Recently, physician profiling was used to introduce financial incentives for physicians to prescribe effectively and efficiently. (AFSCME 2003; United 2003; Graden 1998) Bonus structures reward physicians who prescribe judiciously and

hold accountable those that do not. In addition, capitation for prescription drug costs passes financial incentives to control drug spending on to physicians, who are paid a set allowance for each patient that must cover the patient's yearly prescription drug costs, with any excess coming out of the physician's own pocket. (Burton 2001) Capitation based on prescription drugs is a relatively new technique and has not been used extensively. The advantage of using these methods to shift costs to providers is that it gives physicians an incentive to resist excessive consumer demand for expensive or overprescribed drugs. (Burton 2001) Disadvantages can include an increased emphasis on cost rather than medical necessity in drug choice and the creation of adverse selection incentives where more healthy patients are chosen over less healthy patients by physician providers. To combat adverse selection, physician profiling and capitation should incorporate risk assessment in order to adjust for differences in patient case-mix. Risk assessment techniques attempt to control for clinical differences in patient risk and to isolate quality differences. Risk adjustment techniques can then be utilized to allow physicians who treat more severe patients' higher prescription drug expenditures. When shifting risk to providers, risk assessment/adjustment results in a system that doesn't penalize physicians who treat high-risk patients. In addition to financial incentives, there is no other way to begin a productive dialogue with physicians about using outcomes information to motivate quality improvement without risk assessment. (Iezzoni 1997) Without risk assessment/adjustment physicians will argue that their patients are sicker.

The vast majority of physician profiling and capitation techniques used in the past have concentrated on general health care costs and utilization. However, with the

recent focus on increasing drug expenditures, MCOs and PBMs will likely begin to utilize these techniques more prolifically for prescription costs and utilization.

RISK ASSESSMENT

Measuring and monitoring outcomes of care requires a way to assess and adjust for patients' risks for various outcomes. Risk assessment is essential in most outcome studies. It is a way to remove or reduce the effects of confounding factors in studies where the cases are not randomly assigned to different treatments (Blumberg 1986). The goal of risk assessment is to account for pertinent patient characteristics before making inferences about the effectiveness or quality of care based on patient outcomes (Iezzoni 1997). Simply put, it is a means of leveling the playing field to allow like comparisons.

To date, most risk assessment techniques have emphasized severity of illness and have focused on outcomes of acute care hospitalization (Iezzoni 1997). Many risk assessment techniques, especially the earlier ones, such as the Computerized Severity Index (Horn 1991) and the MedisGroups (Brewster 1985), were formulated to use clinical data elements abstracted from the medical record or some other primary source. While these clinically based techniques have been validated and shown to be predictive, they are both expensive and cumbersome to use because they require data from medical charts that are not readily accessible. Administrative data (data used for purposes such as billing) has become an alternative to timely, expensive clinical data. Although administrative databases were created for purposes other than research, they have increasingly become a mainstay for exploring medical care and outcomes. (Connell 1987; Wennberg 1987; Anderson 1990; Sullivan 1991; Mitchell 1994; Lave

1994) Risk assessment models have been utilized for various outcomes and populations as specific as surgical complications for individuals undergoing lumbar spinal surgery (Deyo 1992) and as general as overall healthcare costs for ambulatory patients. (Fowles 1996) Most risk assessment models have been utilized in a prospective manner where data in one period is used to predict outcomes for a subsequent period. Prospective models used to predict costs have become particularly useful over the last decade for setting capitation rates. (Ash 1989; Ellis 1995) Risk assessment models can also be utilized in a concurrent fashion where data in one period is analyzed to predict outcomes for the same period. These models have been particularly useful over the last decade to profile physicians. (Berlowitz 1998; Roblin 1998; Chang 1996; Parente 1996)

Administrative databases have become a means of exploring a vast bank of patient information both inexpensively and with ease. Some models were designed to be used over time in longitudinal follow-up studies. The Charlson Index, (Charlson 1987) which was later adapted for use with an administrative database (Deyo 1992; Romano 1993), was designed for inpatient populations. Other longitudinal claims-based models were designed for ambulatory patient populations such as the Ambulatory Care Groups (ACG) (Starfield 1991; Fowles 1996) and Ambulatory Patient Groups (APG) (Goldfield 1997). Other code based longitudinal measures such as, Principle Inpatient Diagnostic Cost Group Model for Medicare Risk Adjustment (PIP-DCG) (Ellis 1995), and the Chronic Illness and Disability Payment System (CDPS) (Kronick 2000), were initially designed for certain types of populations; Medicare enrollees and Medicaid enrollees respectively. The Charlson-based models use

comorbidities (secondary diagnoses) to adjust for case-mix while other indices (Ambulatory Care Groups) use all diagnoses. Still, other models, Diagnostic Cost Group Hierarchical Condition Category (DCG-HCC) (Ash 2000), Clinical Risk Groups (CRGs) (Hughes 2004) and CDPS, use a hierarchical approach where diagnostic clusters are utilized to adjust for only the most severe diagnosis coded in each cluster.

While most early risk adjustment techniques focused on the Medicare population, many state Medicaid systems are utilizing risk assessment models to predict reimbursement rates. (Tollen 1998; Weiner 1998). As of 2001, ten states (CO, DE, MD, MI, MN, NJ, OR, TN, UT and WA) utilized risk adjustment models mostly for payment purposes. (Kaelin 2002) Eight states use CDPS/DPS and two states, MD and MN, use ACGs. Numerous other states have evaluated risk adjustment models in an effort to promote more efficient payment. Four states, CO, MD, MI and WA, utilized risk adjustment models for purposes other than payment such as profiling and disease management. With increasing cost controls and the development of new Medicaid specific models such as the CDPS, state Medicaid programs may utilize risk assessment models even more frequently in the future to manage their health care costs.

The primary statistic used to determine how well a model performs is the R-square statistic or the proportion of the variability explained by the model. For models that predict costs, such as the ACG, DCG-HCC, CDPS, and the PIP-DCG models, the R-square value is the proportion of the variability in actual costs that are explained by the model. Health risk can be divided into three categories from a health economist's perspective: fixed effects, time-varying effects, and random effects. (Newhouse 1989;

Newhouse 1998) Fixed effects alter the costs of health for an individual indefinitely as would a chronic disease and can explain approximately 15 to 20 percent of variance in spending. Time-varying effects alter the cost of health for a time but not indefinitely and explain approximately 3 to 5 percent. Random effects explain the rest of the variance and are by definition unpredictable. Based on this premise, the upper bound estimate of the actual spending that can be explained by risk assessment models is approximately 20 percent or that of the fixed effects. This premise and estimate applies to the overall population. However, when assessing a specific population such as the disabled Medicaid population or a group of individuals diagnosed with a specific condition, the amount of spending that can be explained increases. The R-square values for the cost models mentioned above are between .08 and .21 depending on the model and the population, or sub-population, used for the model.

For the most part, risk assessment indices have relied solely on standard demographic data and diagnostic information to adjust for case-mix. As an alternative, population-based automated pharmacy data serves as another source of information to adjust for case-mix. Like administrative databases housing ICD-9-CM codes, these databases are readily available and inexpensive to use. In 1992, Group Health Cooperative (GHC) of Pudget Sound, Seattle, WA developed a drug-based index to evaluate chronic disease status: the chronic disease score (CDS). (Von Korff 1992) Von Korff found that the CDS was associated with both physician-rated disease severity and patient-rated health status and that it predicts subsequent mortality and hospitalization rates. Weights for the CDS were later derived empirically, in a revised version, and shown to perform better than the original clinical weights for predicting costs. (Clark

1995) Clark's CDS has been revised and replicated for both pediatric and adult populations to predict overall healthcare costs. (Gilmer 2001; Fishman 1999; Lamers 1999; Fishman 2003) Each of these found that drug-based models perform much better than demographic models and similarly to a diagnosis-based model in terms of R-square value. The following R-square values were reported for each of the drug-based models: Clark et al. (0.10 for commercial population), Gilmer et al. (0.11 for TANF and 0.15 for disabled Medicaid population), Fishman et al. 2000 (0.14 for pediatric population), Lamers et al. (0.09 for a Netherlands health plan) Fishman et al. 2003 (.094 for commercial population). A few models have been developed and published using both diagnoses and drug information. (Gilmer 2001; Ricci 2001; Clark 1995; Parker 2003) This mixed approach incorporates the most information into the model and has the potential to account for severity by using drug information. This type of model also has the potential to offset some of the disadvantages of a diagnostic model. Some of the problems with diagnostic models include the quality and completeness of diagnostic data, and the absence of diagnosis codes which reflect secondary chronic conditions that may have been coded in a prior visit. A mixed model incorporating drug information would potentially overcome these problems. Mixed models have shown a slight increase over diagnostic models in terms of R-square: Gilmer et al. (from 0.12 to 0.15 for TANF and from 0.24 to 0.26 disabled Medicaid population), Ricci et al. (from 0.13 to 0.15 for a Medicaid population) and Clark et al. (from 0.08 to 0.12), Parker et al. (from 0.26 to 0.263). Because these mixed models have shown a slight improvement over diagnosis-based models for overall health care cost, the researchers believe adding

drug information to diagnostic models may provide even further gains in predictability for prescription expenditures.

Another problem that exists in both diagnostic and drug-based models is reimbursement incentives. When using diagnostic models for payment purposes, the inclusion of prior utilization rewards physicians with higher utilization. Physicians have an incentive to code diagnoses more prolifically and in a manner that optimizes reimbursement. Similarly, drug-based models incorporating prior drug utilization, rewards physicians for prescribing excess medications or more costly medications and encourages poor prescribing habits. For this reason, most risk assessment indices do not take into account how frequently a condition is coded, how many prescriptions were prescribed or outright utilization measures such as total prior health care costs.

Risk assessment indices are confined to outcomes available from the data source being used. In general, most indices, which use an administrative database, predict provider costs (or charges) such as the DRG, DCG-HCC, CDPS, and ACG. Numerous models have been developed and utilized to predict medical costs to capitate and profile physicians. To date, however, there have been no published reports of risk assessment models developed to predict pharmacy expenditures. MCOs and PBMs already evaluate physician prescribing patterns based on their overall prescription expenditures. This very basic form of profiling, based on overall prescription expenditures, cannot be utilized effectively because no risk assessment technique is used to adjust for the patient case-mix of each physician. A physician (or practice) with a patient population consisting primarily of geriatric patients with numerous chronic conditions will naturally consume more drug costs than a physician (or practice) with a

patient population consisting primarily of younger more healthy adults. Escalating drug prices have raised concern for managed care organizations and they have begun to focus more heavily on managing their pharmacy benefits. Risk assessment indices developed specifically for prescription expenditures would be an important tool for MCOs to efficiently evaluate, control and monitor drug costs and utilization.

RATIONALE AND SIGNIFICANCE

This research sought to develop and validate risk assessment models to predict prescription drug expenditures for both commercial and Medicaid populations. These models will be the first risk assessment models publicly reported that were developed specifically for prescription expenditures. The models will be developed in a prospective manner and can be utilized to provide physicians with a prescription expenditure goal for the upcoming year. These prospective models may or may not be used for financial incentives or to capitate physicians based on prescription expenditures. The models will also be utilized in a concurrent manner to profile physicians based on prescription expenditures. Here, they can be utilized by a health plan to reinforce physicians who are prescribing judiciously and inform and educate physicians who are not. Profiling can also be used to build in financial incentives to reward physicians based on prescribing practices.

REFERENCES

American Federation of State, County and M Employees. Prescription Drugs and State Budgets. Health Focus 2003; www.afscme.org/publications/healht_focus/focus302.htm.

Anderson G, Steinberg EP, Whittle J, Powe NR, Antebi S, Herbert R. Development of clinical and economic prognoses from medicare claims data. JAMA 1990; 263(7):967-972.

Ash A, Porell F, Gruenberg L, Sawitz E, Beiser A. Adjusting Medicare capitation payments using prior hospitalization data. Health Care Financ Rev 1989; 10(4):17-29.

Ash AS, Ellis RP, Pope GC, Ayanian JZ, Bates DW, Burstin H et al. Using diagnoses to describe populations and predict costs. Health Care Financ Rev 2000; 21(3):7-28.

Berlowitz DR, Ash A, Hidkey EC, Kader B, Friedman R, Moskowitz MA. Profiling outcomes of ambulatory care: casemix affects perceived performance. Medical Care 1998; 6:928-933.

Blumberg MS. Risk adjustment for Medicare. N Engl J Med 1999; 340(19):1514-1515.

Brewster AC, Karlin BG, Hyde LA, Jacobs CM, Bradbury RC, Chae YM. MEDISGRPS: A Clinically Based Approach to Classifying Hospital patients at Admission. Inquiry 1985; 22(4):377-387.

Bridges to Excellence. Diabetes Care Link for Physicians and Providers. 2004; www.bridgestoexcellence.com/bte/diabetescarelink/gty_physicians.htm.

Burton SL, Randel L, Titlow K, Emanuel EJ. The ethics of Pharmaceutical benefit management. Health Aff (Millwood) 2001; 20(5):150-163.

Carroll J. Physicians Reconsider Taking on Pharmacy Risk. Managed Care July 2000.

Chang W, McCracken SB. Applying case mix adjustment in profiling primary care physician performances. Journal of Health Care Financing Review 1996; 22(4):1-9.

Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. J Chronic Dis 1987; 40(5):373-383.

Clark DO, Von Korff M, Saunders K, Baluch WM, Simon GE. A chronic disease score with empirically derived weights. Med Care 1995; 33(8):783-795.

Connell FA, Diehr P, Hart LG. The use of large data bases in health care studies. Annual Review of Public Health 1987; 8:51-74.

Deyo RA, Cherkin DC, Ciol MA. Adapting a clinical comorbidity index for use with ICD-9-CM administrative databases. J Clin Epidemiol 1992; 45(6):613-619.

Ellis R.P., Ash A. Refinements to the Diagnostic Cost Group model. Inquiry 1995; 32:1-12.

Employee Benefit Research Institute. Prescription Drug Costs Up Sharply - but Still Small Overall. Press Release 470 Washington EBRI . 1999.

Fishman PA, Shay DK. Development and estimation of a pediatric chronic disease score using automated pharmacy data. Medical Care 1999; 37:874-883.

Fishman PA, Goodman MJ, Hornbrook MC, Meenan RT, Bachman DJ, O'Keefe Rosetti MC. Risk Adjustment Usng Automated Ambulatory Pharmacy Data: The Risk Model. Medical Care 2003; 41(1): 84-99.

Fowles JB, Weiner JP, Knutson D, Fowler E, Tucker AM, Ireland M. Taking health status into account when setting capitation rates: a comparison of risk-adjustment methods. JAMA 1996; 276(16):1316-1321.

Galt KA, Rich C, Dralewski JE, Turner PD, Bernhardt TS, Dowd B, Feldman R, de Vries A. Group practice strategies to manage pharmaceutical cost in an HMO network. American Journal of Managed Care 2001; 7(11):1081-1090.

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

Goldfield N, Averill RF, Grant T, Gregg LW. The clinical development of an ambulatory classification system: version 2.0 Ambulatory Patient Groups. J Ambulatory Care Manage 1997; 20(3):49-56.

Graden SE, Schafermeyer KW. Performance Reporting for Managed Care Prescription Programs. Journal of Managed Care Pharmacy 1998; 4(2): 160-166.

Horn SD, Horn RA. The Computerized Severity Index. A new tool for case-mix management. J Med Syst 1986; 10(1):73-78.

Hughes JS, Averill RF, Eisenhandler J, Goldfield NI, Muldoon J, Neff JM, Gay JC. Clinical Risk Groups (CRGs): A Classification System for Risk-Adjusted Capitation-Based Payment and Health Care Management. Medical Care 2004; 42(1): 81-90

Iezzoni L, Ash AS, Daley J, Hughes JS, Shwartz M. Risk adjustment for measuring healthcare outcomes. 2nd ed. Chicago: Health Administration Press, 1997.

IMS Health I. Integrated Promotional Services (Office Promotion Reports, Hospital Promotion Reports, National Journal Audit) and Integrated Share of Voice Report (with data from Competitive Media Reporting). 1999.

Kleinke JD. The Price of Progress: Prescription Drugs in the Health Care Market. Health Affairs 2001; 20(5):43-60.

Kaelin JJ. Risk Adjustment for State Medicaid Programs: Lessons Learned and Prospects for the Future. Center for Health Program Development and Management April 2002; wwwacg.jhsph.edu/library/conference_2002/d2.plenary%203.%20kaelin.ppt

The Kaiser Family Foundation. Prescription Drug Trends. May 2003; www.kff.org/rxdrugs/3057-03-index.cfm

The Kaiser Family Foundation. Prescription Drug Trends. September 2000; www.kff.org

The Kaiser Family Foundation. Prescription Drugs: Facts at a Glance. 2003; www.kaisernetwork.org/static/spotlight_rxdrugs_facts.cfm.

The Kaiser Family Foundation. Medicaid and the Prescription Drug Benefit: Cost Containment Strategies and State Experiences. September 2002; www.kff.org/medicaid/4063.index.cfm

Kronick R, Gilmer T, Dreyfus T, Lee L. Improving health-based payment for medicaid beneficiaries: CDPS. Health Care Financ Rev 2000; 21(3):29-64.

Lamers LM. Pharmacy cost groups: a risk-adjuster for capitation payments based on the use of prescribed drugs. Medical Care 1999; 37(8):824-830.

Lave JR, Pashos CL, Anderson GF, Brailer D, Bubolz T, Conrad D et al. Costing medical care: using Medicare adminstrative data. Medical Care 1994; 32(7):JS77-JS89.

Litton LM, Sisk FA, Akins ME. Managing drug costs: the perception of managed care pharmacy directors. Am J Manag Care 2000; 6(7):805-814.

Mitchell JB, Bubolz T, Paul JE, Pashos CL, Escarce JJ, Muhlbaier LH et al. Using Medicare claims for outcomes research. Medical Care 1994; 32(7):JS38-JS51.

Motheral BR, Henderson R. The effects of a closed formulary on prescription drug use and costs. Inquiry 1999-00; 36(4):481-491.

National Committee for Quality Assurance. Diabetes Physician Recognition Program (DPRP) Frequently Asked Questions. 2004; www.ncqa.org/dprp/dprpfaq.htm

National Committee for Quality Assurance. Integrated Healthcare Association Pay for Performance Program: 2004 Clinical Measure Specifications and Audit Review. December 2003; www.iha.org

Newhouse JP, Buntin MB, Chapman JD. Risk adjustment and Medicare: taking a closer look. Health Affairs 1997; 16(5) 26-43.

Newhouse JP, Manning WG, Keeler EB, Sloss EM. Adjusting Capitation Rates Using Objective health measures and prior utilization. Health Care Financing Review 1989; 10(3):41-54.

Parente ST, Weiner JP, Garnickk DW, Fowles J, Lawthers AG, Palmer RH. Profiling resource use by primary-care practices: managed Medicare implications. Health Care Financing Review 1996; 17:23-42.

Ricci JF, Dorfman JH, Martin BC. Development and Validation of Prospective Cost Risk Adjustment Indices Using Administrative Medical and Drug Information for Medicaid Populations. Poster presentation at the Academy for Health Services Research and Health Policy, 2001 Annual Meeting, Atlanta, GA, June 2001.

Roblin DW. Physician profiling using outpatient pharmacy data as a source for case mix measurement and risk adjustment. Journal of Ambulatory Care Management 1998; 21(4):68-84.

Romano PS, Roos LL, Jollis JG. Adapting a clinical comorbidity index for use with ICD-9-CM administrative data: differing perspectives. J Clin Epidemiol 1993; 46(10):1075-1079.

Starfield B, Weiner J, Mumford L, Steinwachs D. Ambulatory Care Groups: A Categorization of Diagnoses for Research and Management. Health Services Research 1991; 26(1):53-74.

Sullivan LW, Wilensky GR. Medicare Hospital Mortality Information. 1987, 1988, 1989. 1991. Washington, D.C., U.S. Department of Health and Human Services, Health Care Financing Adminstration.

Tollen L, Rothman M. Case study: Colorado Medicaid HMO risk adjustment. Inquiry 1998; 35:154-170.

Tucker JL, III. The theory and methodology of provider profiling. Int J Health Care Qual Assur Inc Leadersh Health Serv 2000; 13(6-7):316-321.

United physicians incentive program reduces pharmacy costs. PR Newswire. April 17, 2003.

Von Korff M, Wagner EH, Saunders K. A chronic disease score from automated pharmacy data. J Clin Epidemiol 1992; 45(2):197-203.

Wahls TL, Barnett MJ, Rosenthal GE. Predicting Resource Utilization in a Veterans Health Administration Primary Care Population: Comparison of Methods Based on Diagnoses and Medications. Medical Care 2004; 42(2): 123-128.

Weiner JP, Tucker AM, Collins AM, Fakhraei H, Lieberman R, Abrams C, Trapnell GR, Folkemer JG. The development of a risk-adjusted capitation payment system: the Maryland Medicaid model. Journal of Ambulatory Care Management 1998; 21(4):29-52.

Wennberg JE, Roos N, Sola L, Lchori A, Jaffe R. Use of claims data systems to evaluate health care outcomes. Mortality and reoperation following prostatectomy. JAMA 1987; 257(7):933-936.

CHAPTER 3

METHODS

OVERVIEW

A retrospective longitudinal review was employed to empirically develop risk assessment models to predict pharmacy expenditures for both a managed care commercial and Medicaid populations. Three new models were developed separately for each population utilizing claims data from the MEDSTAT MarketScan database (commercial) and the California Medicaid administrative claims database.

» Demographic based model

» ICD-9-CM diagnosis-based model (RxCost Model)

» Model incorporating both diagnostic and drug information (MRxCost Model)

An additional model was developed for the Medicaid population that added Medicaid eligibility criteria to the demographic model.

» Demographic and Medicaid Eligibility Model

The risk assessment models were developed in a prospective fashion and utilized information from one plan year to predict pharmacy expenditures for the following year. The demographic models and the Medicaid demographic and eligibility model were used to establish a baseline to observe the predictive power of a model that can be derived with minimal effort. The RxCost models (diagnostic models) were the primary models for this research. The MRxCost models (mixed models) were used to

explore how much predictive power would be obtained by adding drug information. A random sample of each population was utilized to validate the models.

DATA

MEDSTAT MarketScan data was utilized to develop and validate risk assessment indices for the commercial population and California Medicaid data was utilized for the Medicaid population. Both administrative databases consist of enrollment information, prescription claims, inpatient and outpatient hospital claims, and ambulatory claims in a recipient level linkable format. These databases provide the level of detail on costs and utilization needed to develop and validate risk assessment models. Both populations use data collected from the same time period, 1998 through 2000, to develop the models. All methods employed to mine the data and generate the model samples were performed using SAS Version 8.2 (SAS Inc. 2004).

MarketScan data

The MEDSTAT Group Inc. (MEDSTAT Inc 2001) has built and maintained the healthcare databases (The MarketScan Research databases) since 1989. The data consists of paid medical encounters, including prescription medications, for over 2 million persons annually nationwide who are under age 65 and who are enrolled in commercial insurance plans. The annual database includes private sector health data from approximately 100 payers including large employers, health plan, and government and public organizations. The data represents medical experience of insured employees and their dependents for active employees, their dependents, and early retirees of companies who participate in the database. The MarketScan database captures person-specific clinical utilization, expenditures and enrollment across

inpatient, outpatient, prescription drug, and carveout services. These data have been found to be valid in previous epidemiological and health policy studies. (Gandhi 2001, Ozminkowski 2000, Zhao 1999)

Three years of MarketScan data, 1998 through 2000, was utilized. The data was converted to SAS data files using the SAS xport function and proc copy. Each year of data was loaded onto a password protected server and all counts were verified using the documentation provided. Each year was merged to create one dataset spanning three years and counts were verified; the data is linkable to allow members to span years. The total unduplicated count for the new MarketScan dataset was just over 9 million lives (Table 3.1).

California Medicaid Data

The California Department of Human services (CDHS 2001) is the state agency responsible for the Medicaid program in California. A 20% sample of the California Medicaid data (MediCal) was purchased through the CDHS. The CDHS randomly selected 20% of all Medicaid enrollees based on the last two or three digits of their Social Security Number (SSN). Their SSN was then encrypted to allow for an anonymous patient identifier to link claims. The data consists of all paid medical encounters delivered to Medicaid eligible beneficiaries enrolled in Fee-for-Service plans, including prescription medications, inpatient claims, and outpatient claims. These data have been found to be valid in previous epidemiological and health policy studies and have been used in developing other risk assessment indices. (Adams 2001; Perkins 2001; Croghan 1999)

Three years of MediCal data, 1998 through 2000, was utilized. The data was converted from SAS xport files to SAS data files using the SAS xport function. Each year of data was loaded onto a password protected server and all counts were verified using the documentation provided. The 20% sample included over 1.2 million Medicaid beneficiaries annually. Each year was merged to create one dataset spanning three years and the counts were verified; the data is linkable to allow members to span years. The total unduplicated count for the new MediCal dataset was over 1.6 million lives (Table 3.2).

SUBJECTS

MarketScan

Persons meeting the following criteria were included in the MarketScan sample to develop and validate the commercial risk assessment indices.

- Continuously eligible for a minimum of thirteen months from January 1998 through December 2000
- Have prescription drug coverage
- Age 18 to 62 years at the beginning of the observational period
- Not admitted to institutions or nursing home facilities and who do not have periods of inpatient care in excess of 30 consecutive days at any time during the entire period
- Not dually eligible for Medicare

The member count for the MarketScan sample dropped from approximately 9 million to 1.6 million after excluding members who did not meet the above criteria (Table 3.3). After the other criteria were enforced there were no members remaining who were

dually eligible for Medicare. From the 1.6 million members who met the criteria for inclusion, a random 80% sample was selected (referred to as the training sample) and used to develop the risk assessment models. Of the remaining 20%, two-thirds of the sample was randomly chosen to validate the models. The remaining one-third was set aside as spare data in case problems were encountered and the models need to be redeveloped / re-calibrated and validated (Table 3.4). The percentages for each sample were chosen to maximize the number of members available for development of the models while maintaining an adequate number in which to validate the models. This study was reviewed and approved by the University of Georgia Institutional Review Board; IRB # H2002-10473-2.

MediCal

Persons meeting the following criteria were included in the MediCal sample to develop and validate the Medicaid risk assessment indices.

- Not dually eligible for Medicare
- Continuously eligible for a minimum of thirteen months from January 1998 through December 2000
- Age 18 to 62 years at the beginning of the observational period
- Eligible for Aid to Families with Dependent Children (AFDC), Temporary Assistance for Needy Children (TANF) or Supplemental Security Income (SSI)
- Not admitted to institutions or nursing home facilities and who do not have periods of inpatient care in excess of 30 consecutive days.

On a national level, TANF replaced AFDC when The Personal Responsibility and Work Opportunity Reconciliation Act of 1996 became law. (US Dept. of Health and Human Services 2001) However, states are given a large amount of latitude in using various aid categories as long as the minimum standard for TANF exists. For this reason, AFDC categories are still used in many states including California. Members were excluded if they were eligible for California Medicaid due to state funded programs in which the state receives no national funds (i.e.)

The member count for the MediCal sample dropped from approximately 1.68 million to 276 thousand after excluding members who did not meet the above criteria (Table 3.5). From the 276 thousand members who met the criteria for inclusion, a random 50% sample was selected (training sample) and used to develop the risk assessment models. Of the remaining 50%, two-thirds of the sample was randomly chosen to validate the models. The remaining one-third was set aside as spare data in case problems were encountered (Table 3.6). Once again, the percentages for each sample were chosen to maximize the number of members available for development of the models while maintaining an adequate number in which to validate the models.

RISK ASSESSMENT

The same methodology was employed for both the commercial and Medicaid populations to develop the risk assessment models. Three models, a demographics-based model, a diagnostic-based model (RxCost Model) and a mixed model (MRxCost Model) incorporating both diagnostic and drug information, were developed for each population. Medicaid eligibility information was also utilized in the Medicaid population

to develop a fourth model. This model adds Medicaid eligibility information to the demographics model.

Demographic Models

A risk assessment model based solely on age and gender demographic information was empirically developed to predict prescription expenditures. Although not always statistically significant, increasing age is fairly consistent with increased expenditures and is easily abstracted from most administrative databases. (Iezzoni 1997; Wray 1997) Gender is another variable easily abstracted, however the actual relationship of cost to age and sex is complex. (Diehr 1999) The curves are nonlinear and differ by sex. Males and females have similar utilization until puberty, at which time women increase their utilization because of childbearing. Men's utilization is low until about age 40. For these reasons, the following mix of age-sex dummy variables was utilized where 1 indicates presence and 0 indicates absence:

» (Male 18-22years), (Male 23-30years), (Male 31-40years)
(Male 41-50years), (Male 51-60years), (Male 61-64)

» (Female 18-22years), (Female 23-30years), (Female 31-40years),
(Female 41-50years), (Female 51-60years), (Female 61-64)

Race is another variable that could have been utilized in the Medicaid models as it is a variable reported in the MediCal data. However, race would not be a variable in which state Medicaid programs or MCOs would use so it is not utilized to develop the models.

Demographic and Medicaid Eligibility Models

Medicaid eligibility status is a contributing factor in prescription drug utilization. (Baugh 1999) Members were included in the Medicaid sample if they were eligible for

TANF, AFDC or SSI. The Medicaid Demographic and Eligibility Model adds Medicaid eligibility information to the Demographic Model defined above. Due to TANF replacing AFDC, the aid categories are very similar or even identical. TANF incorporates a time limit and only provides assistance for up to 60 months. After 60 months a state can continue to assist these individuals through state funds. In this situation, the TANF category may be replaced by the former AFDC category. Here, they can be identical; TANF would expire and AFDC would begin. The Medicaid Demographic and Eligibility Model treats TANF and AFDC as one category and SSI as a separate category. The model includes a dummy (0 / 1) variable to indicate Medicaid eligibility status for SSI. When an individual is eligible for both TANF/AFDC and SSI he/she is treated as being eligible for SSI.

Diagnostic-Based Models

The RxCost Model, an ICD-9-CM code-based diagnostic risk assessment model, was developed for both populations. This model uses diagnostic information in addition to demographic information (and Medicaid Eligibility for the Medicaid population) to prospectively predict prescription expenditures. Both populations utilize the identical classification schemes.

With more than 12,000 diagnostic codes making up the ICD-9-CM classification system, the first challenge is to aggregate these diagnostic codes into clinically meaningful categories that reflect similar prescription expenditures. Techniques used by Ash and Kronick in developing the Diagnostic Cost Groups-Hierarchical Coexisting Conditions (DCG-HCC) (Ash 2000) and Chronic Illness and Disability Payment System (CDPS) (Kronick 2000) were used as a guide to identify diagnostic groupings. The

classification system groups all relevant diagnoses into large major categories. Each of these large clusters of diagnosis groups were subdivided into smaller homogenous groups according to low and higher cost groups. The initial classification approach was based on AHRQ's Clinical Classification Software (CCS). (Clinical Classification Software 2001) The CCS is a multi-level classification scheme that aggregates individual ICD-9-CM codes into clinically meaningful categories that group similar conditions. (Agency for Health Care Policy and Research) The CCS Software is available for download at www.ahrq.org. The multi-level scheme used by CCS aggregates ICD-9-CM codes into 17 broad categories (e.g., Infectious Diseases, Neoplasms, and Mental Disorders) excluding the residual E codes (Appendix A). These 17 major categories are further split into more refined subcategories as in the following example:

1 Infections and parasitic diseases

1.1 Bacterial infection

1.1.1 Tuberculosis

1.1.2 Septicemia (except in labor)

1.1.2.1 Streptococcal septicemia

1.1.2.2 Staphylococcal septicemia

The determining factor in creating these categories was the extent to which conditions could be grouped into relatively homogeneous clusters of interest to public policy researchers. These CCS categories were not created specifically for prescription expenditures and, therefore, do not necessarily reflect conditions with similar prescription expenditures.

Based on the prior use of diagnostic models to predict total health care costs, (Ash 2000, Kronick 2000) six of the major CCS categories that were thought to be too broad to explain prescription drug costs were further subdivided. The revised classification scheme consists of 34 major categories. All subcategories remained under their appropriate classification. For example, “diabetes mellitus” was taken out of the original CCS category 3 (endocrine and metabolic diseases) and placed as its own major category. The subcategories “diabetes mellitus without complications” and “diabetes mellitus with other complications” were also removed in order to remain a subcategory of diabetes mellitus. The revised major classifications are presented in Table 3.7.

A clinical panel, consisting of two practicing clinical pharmacists that were part of the doctoral dissertation committee, was assembled to review the CCS-based categories and organize the categories into a meaningful classification that will have the potential to further stratify expected pharmacy costs. Utilizing the multi-level CCS-based classification the most refined subcategory deemed appropriate by the clinical panel will be considered a diagnostic category. Therefore, within each of the 34 major categories several diagnostic categories could potentially exist. Each clinician was instructed to review the CCS-based multi-level categories independently and charged with the following tasks:

- Review each of the 34 major categories and further divide any category that they feel is too broad to explain outpatient prescription expenditures.
- Group conditions within each category (main categories or subcategories) that are likely to result in similar outpatient prescription expenditures.

» Identify all conditions that are likely to result in zero or negligible prescription expenditures.

» Identify all conditions that are not well defined clinically. Here, a condition is well defined if the diagnosis for that condition has a clear, shared meaning among clinicians.

Diagnostic coding in general is often highly variable because of the absence of operational clinical definitions. (Iezzoni 1997) In addition, the diagnostic coding of certain conditions is even more variable because of the uncertainty of these conditions. For this reason, even if the clinical panel recognizes that two or more diagnoses will significantly impact prescription expenditures differently, they were asked to group these conditions together if they deem that distinctions between these diagnoses cannot be easily made. For example, the CDPS classification groups paranoid schizophrenia and catatonic schizophrenia together even though paranoid schizophrenia is associated with more elevated future cost because the clinical panel was not convinced that these subtypes could be diagnosed consistently. (Kronick 2000) When discrepancies occurred between the two panel members, the broadest classification was utilized. The discrepancies were perceived as a case where there is variability in diagnostic coding as explained above. The low-cost and ill-defined conditions identified by the clinical panel were excluded from the classification in order to make the system more reliable. (Ash 1998; Kronick 2000) (Table 3.8)

The clinical panel felt that the 34 major categories proposed were appropriate for aggregating diagnostic codes and therefore were not further subdivided. Numerous subcategories were chosen to be collapsed into broader categories. For example, all

subcategories under Neoplasms were collapsed into one category (the major category) “Neoplasms”. All subcategories under Septicemia (except in labor) were collapsed into one category called “Septicemia (except in labor)”. The final CCS-based diagnostic classification containing 90 diagnostic classes is presented in Table 3.9. All categories that are missing in the final models were either discarded (identified in Table 3.8) or collapsed into the broader categories.

One additional concern remained with the MarketScan population. Approximately 34% of members in this population were enrolled in capitated plans. Data for these members consist of encounter data and not claims data used for billing purposes. The potential exists for encounter data to be much less complete because there is no incentive driving physicians to code thoroughly. Realizing this concern, the new diagnostic classification system was applied separately to the Fee For Service (FFS) members in the MarketScan training sample and all members in the MarketScan training sample (Table 3.10). Looking at the frequencies of each category there was no real concern with the FFS patients differing from the Encounter patients and therefore all patients were utilized. The frequencies for the MediCal population are presented in Table 3.11. The prevalence of the diagnostic categories were checked to ensure that an adequate number of cases exist in each. When the prevalence was low in any category, it was either dropped or merged with another category. The only category in the diagnostic classification for Medicaid that differs from the Commercial classification is female infertility. This category was excluded from the MediCal model due to low prevalence presumably because MediCal does not cover this condition.

The RxCost Model was empirically derived for each population. Dummy variables (1 / 0) were utilized to indicate presence or absence of a diagnostic category. Both inpatient and outpatient claims (ICD-9CM codes) were utilized for diagnostic information. Two separate methods were initially utilized to capture diagnostic information within the RxCost models. One method, the full method, identifies and enters into the model all diagnostic categories for which there was a corresponding ICD-9-CM code present regardless if there were other similar diagnostic categories coded for each individual. The other method utilizes a hierarchical approach for certain variables where some categories are grouped together in clusters and only the most costly category within each cluster is entered into the model. In order to use the hierarchical method, relevant categories must first be grouped into clusters. The MarketScan training sample was utilized to identify variables that would be appropriate to cluster in a hierarchical fashion. To define the clusters, correlations between all possible diagnostic category variables were generated to identify those with a Pearson Correlation ≥ 0.2. After these variables were identified a 2x2 contingency table for each of the correlated variable pairs was output and the phi coefficient was estimated (Table 3.12). Based on the contingency results, 16 variables were considered appropriate for a total 6 hierarchical clusters. This method was utilized because it identifies variables that are often coded together on the same claim. These 6 clusters each contain variables that are often coded for the same condition. When this occurs only the most costly variable should be captured by the model not multiple variables for the same condition. For example, if a patient is coded for diabetes mellitus without complications and diabetes mellitus with other complications only the most costly of these variables,

diabetes with complications, should be captured by the model. To determine the most costly variable for each cluster a random 20% sample of each training sample (MarketScan for commercial population and MediCal for Medicaid population) was utilized to calculate the average annual prescription cost for each of the 16 variables in each population. The cost results for the Commercial and Medicaid populations are presented in Table 3.13 and Table 3.14 respectively. Based on these results the variables within each cluster were assigned a hierarchical classification. Utilizing the hierarchical approach, only the single-highest category within each hierarchical cluster will be captured by the model.

This hierarchical approach was initially adopted to prevent code proliferation and "gaming" of the risk assessment system where clinicians may have incentive to code more conditions which may be justified but do not substantially add to the prescription costs of those persons. This hierarchical method of counting has been shown to simplify the model, strengthen its resistance to additional coding, and produce only small decreases in the accuracy. (Kronick 2000)

Because only 16 variables were affected and only 6 clusters were formed using the hierarchical counting scheme, these 6 hierarchical clusters were incorporated into the full method. The 6 clusters chosen are potential areas for gaming an incentive-based system. Additionally, the minor changes should result in very little loss of predictive power. However, to ensure this was the case, the non-hierarchical full method was later run in addition to the method incorporating the hierarchical clusters for comparison. In both populations the R-square values remained the same.

Combined Code-Based and Drug-Based Measures

Diagnosis-based models have some shortcomings that are evident regardless of counting method. These models rely on administrative claims-based data that are not always complete. Diagnosis codes often suffer from left censoring where an individual may be treated for a chronic condition but only diagnosed once for the condition or diagnosed sporadically over long periods of time. Also, diagnostic coding is often an uncertain practice where many ill-defined conditions may be coded in numerous ways. If reimbursement relies on or is supplemented by diagnostic codes the data will be much more complete but incentives exists for physicians to code more prolifically or to code uncertain conditions in a manner that increases reimbursement rates. Adding drug information to a diagnosis-based model can address some of the problems with these diagnosis-based models. One advantage to utilizing drug information is that the data is much more complete than diagnostic information. Secondly, drug information can help identify individuals who are being treated for a chronic condition but who are not diagnosed at each physician-patient encounter. Another advantage is that drug information can potentially help explain severity. The number of classes of medication prescribed to treat a condition can sometimes be used to indicate severity. The disadvantages to utilizing drug information to predict drug expenditures is that a certain degree of endogeneity exists because prescription fills lead directly to greater prescription cost. However, endogeneity is less of a concern using models prospectively because drug costs in one period are only indirectly related to drug costs in a previous period. (Clark 1995) Another disadvantage of using drug information occurs when the risk assessment models are used for shifting some of the risk to

providers (i.e. financial incentives). Here, there will be some incentive for providers to prescribe medications more liberally, particularly for medications that can potentially be prescribed for conditions with varying levels of severity or even prevention.

A new risk assessment model, the MRxCost Model, based on both diagnostic information and drug information was developed for both populations. This “Mixed Model” will be utilized to ascertain how much predictive ability can be gained by adding drug information to the RxCost Model explained above. This model is an exploratory model with both advantages and disadvantages as compared to the primary RxCost Model. The RxCost Model was supplemented with drug information based on Gilmer's updated version of Clark's revised Chronic Disease Score to develop the MRxCost Model. (Gilmer 2001; Clark 1995)

Virtually all of the published literature utilizing prescription drugs to predict costs has been based on the Von Korff's Chronic Disease Score. (Gilmer 2001; Fishman 1999; Fishman 2003; Lamers 1999; Clark 1995; Johnson 1994; Von Korff 1992) Gilmer's Medicaid Rx model provides the latest approach to updating the CDS and utilizing prescription drugs to model overall healthcare expenditures. (Gilmer 2001) Gilmer et al. first updated the categories and drugs used in Clark's revised CDS by adding some drugs and excluding others based on a pharmacological review by clinicians. Next, an empirical review was used to suggest additional categories. Here, they utilized the Multum Lexicon (Multum 2001) to add drugs not already in the model and regressed the entire set on subsequent year Medicaid expenditure data. Several new categories were added based on the empirical analysis. Gilmer’s Medicaid Rx model differs from the CDS because it includes a few categories that are predictive of

future expenditures but are not necessarily related to a specific chronic condition. The Medicaid Rx model categories are presented in Table 3.14. These categories were utilized as a starting point; however, they were not created specifically for prescription expenditures and, therefore, do not necessarily reflect conditions with similar prescription expenditures.

As with Gilmer et al., the Multum Classification system was utilized both to refine Gilmer's Medicaid Rx categories and to classify drug claims (Appendix B). Both the MEDSTAT MarketScan data and the California Medicaid data utilize a drug classification scheme, however these classifications differ from each other and have not been updated with newer therapeutic classes. For these reasons, Multum's therapeutic classification system was used to classify National Drug Codes (NDCs) into therapeutic classifications. Multum Information Systems Inc. developed their therapeutic classification system and database by obtaining information from the pharmaceutical industry, wholesalers, the Federal government, drug catalogs, etc. Multum updates their database monthly and as of February 2002 there were 284 therapeutic classes. However, the Multum class numbers are not entirely sequential due to former classes being dropped or split into multiple classes.

The Clinical panel reviewed Gilmer's Medicaid Rx model categories and medication classes to ensure that they reflect a proxy of severity and comorbidity that would consistently reflect an individual's expected outpatient prescription expenditure. The clinicians were instructed to review the Medicaid Rx model categories and medication classes independently and were charged with the following tasks:

- » Review Gilmer's Medicaid Rx categories and further subdivide or delete any category they feel is too broad or not appropriate to reflect a proxy of severity and comorbidity consistent with outpatient prescription expenditures.
- » Review the Multum classes and add any condition that is not already present in the Medicaid Rx model categories that could reflect a proxy of severity and comorbidity for outpatient prescription expenditures.
- » Identify conditions where the number of classes prescribed will consistently indicate severity.
- » Review the summary drug descriptions and add or delete medications to ensure they reflect a proxy of severity and comorbidity consistent with outpatient prescription expenditures.
- » Identify conditions where prescribing is highly susceptible to practice patterns. For example, Reflux disorders are highly susceptible to practice patterns because PPIs may very well be used to treat patients suffering from more severe conditions; however, they may just as easily be used to treat less severe patients. Here, a physician essentially faces no consequences for prescribing a PPI medication (instead of an H2 blocker) to a patient who would benefit equally from either one.

Conditions that were identified by the clinical panel as conditions where prescribing is highly susceptible to practice patterns were not further subdivided even if the clinical panel felt that the condition further subdivided could better reflect a proxy of severity and comorbidity for outpatient prescription. Next, the prevalence of the drug categories in each population were checked to ensure that an adequate number of cases exist in

each. When the prevalence was low in any category, it was either dropped or merged with another category. The new drug classification system deleted 6 original categories, combined Tuberculosis with PCP pneumonia and added numerous drugs to the drug list of some categories (Table 3.15). There were 2 categories that differ in the Medicaid population from the Commercial population. The categories of Alzheimer's disease and Iron deficiency were dropped from the Medicaid MRxCost Model due to MediCal formulary restrictions. The classification used and the prevalence of each category for the commercial MRxCost Model and the Medicaid MRxCost Model are shown in Tables 3.16 and 3.19 respectively.

The MRxCost Model was empirically derived for each population. Dummy variables (1 / 0) were utilized to indicate presence or absence of a diagnostic category or drug category. Both inpatient and outpatient claims (ICD-9CM codes) were utilized for diagnostic information and pharmacy claims (NDC codes) were utilized for drug information. Here, the presence of a drug in each category will be counted rather than number of drugs in each category. However, three drug categories were identified by the clinical panel as categories that could potentially help explain severity based on the number of drug classes prescribed. These categories are asthma, cardiovascular, and seizure disorders. Here, the dummy variables for each class were not included in the model but were summed in a single additive variable. Both asthma and seizure disorders categories have 6 classes of medication that could potentially be prescribed. A variable was created for each and was coded as 0 through 6 depending upon how many drug classes are prescribed. The Cardiovascular category has 8 classes that could be potentially prescribed. Here a variable was created in which a value of 0 to 8

was assigned. For example, if a patient filled medication that is classified into Multum categories 131, 180 and 243 then he/she would be assigned a 3 for their asthma additive variable.

The new drug classification system, that was used to complement the RxCost Model, contains some categories that are very similar or identical to the RxCost Model. Although similar (and collinear), these variables will be helpful to identify patients with conditions who did not receive a diagnosis within the timeframe the model is run but filled a prescription for the condition. These variables were combined so that either a diagnosis or a drug would indicate the presence of a condition. Some of these variables, such as asthma, are straightforward. Here both the diagnostic portion and the drug portion of the mixed model contain a category for asthma. However, other categories are more complicated. The diagnostic classification system uses two categories for diabetes while the drug classification system uses one. Other categories are concise in one classification and broad in the other. To help combine categories correlations were again calculated to identify drug categories that were correlated with diagnostic categories. All variables that were correlated ≥ 0.2 were chosen and 2x2 contingency tables were created (Table 3.18). Fifteen drug categories were chosen and combined with their diagnostic counterparts. These categories and the logic used are presented in Table 3.19. Because dummy variables are used in both the diagnostic portion and the drug portion of the MRxCost Model, the presence of a drug in a drug category that was combined with a diagnostic category will only add to the model when there is no presence of a diagnosis. For example, if a patient is diagnosed with asthma, the presence of a drug in the asthma category (patient filled an asthma drug) will not

affect the model because these categories were combined and the diagnosis is already present.

ANALYSIS

A minimum of 13 months of data and up to 24 months was analyzed for each enrollee where the first twelve months served as the index year and was used to collect ICD-9CM and drug class information. Information gathered from the first 12 months was used to predict pharmacy expenditures for months 13 through 24. Where possible, the most recent twenty-four months of continuously eligible data was used for each enrollee. When twenty-four months of continuously eligible data was not available, the longest span of continuously eligible data was utilized. Here, the first 12 months of data was used as the index year and the remaining months were used for the cost year. Prescription costs for individuals with partial second-year data was annualized using the method described in Ash et al described below (Ash 1989)

Total Rx Expenditure = Rx Cost $\times$ (12 $\div$ Months Eligible)

Ordinary Least Squares regression (OLS) was initially considered to model prescription expenditures. OLS is the most common method for analyzing health utilization data; however, this procedure is not without problems. Health care utilization data are typically consists of numerous persons with zero expenditure and a small number of individuals with very high expenditures. While OLS models may not reflect this distribution, they are simple to use and have been shown to work just as well as other more complicated models for predicting future costs. (Diehr 1999) Another advantage of using OLS is that almost all risk assessment models developed in the past

have used OLS and the individuals who utilize such models are familiar with OLS techniques and prefer their simplicity. For these reasons, OLS was utilized.

» $C_i = \Sigma \delta_j x_{ij} + \Psi$ where C is cost, δ is a vector of coefficients and Ψ is the residual error

Variables included in the OLS model are as follows:

The *demographic models* were estimated with:

» Intercept

» dummy variables for age-sex classification (1 = year; 0 = absence)

The *demographic and Medicaid eligibility model* was estimated with:

» same as demographic model

» dummy variable for Medicaid SSI eligibility (1 = yes; 0 = no)

The *RxCost model* was estimated with:

» same as demographic model (demographic and eligibility model for Medicaid)

» dummy variables for the presence or absence of a given condition in the revised diagnostic classification system

(1 = diagnosis present in a clinical classification; 0 = absence)

The *MRxCost model* (both diagnostic and drug information) were estimated with:

» same as RxCost Model

» dummy variable for the presence of absence of a drug from a given therapeutic class in the revised Drug classification system

(1 = drug present in therapeutic class; 0 = absence)

» additive variable for the 3 conditions identified as indicating severity based on the number of medication classes prescribed

To assess the distribution of both data sources and to check the proportion of non-spenders, the deciles were viewed in terms of cost. As expected the data was not normally distributed in either source (commercial or Medicaid). There were a large proportion of non-spenders in both populations. Due to the non-normal distribution, a two-part regression model was also estimated.

A two-part regression model (Duan 1983) was chosen because of the problem faced when using medical data including a high percentage of non-spenders and a small number of very heavy spenders. This type of model attempts to correct the problems by first using a logit equation for the dichotomous event of having zero or positive pharmacy expenses. The next part of the model is a regression equation that is conditional on having positive pharmacy expenses and is used to model the level of positive expenses. A log-linear regression model was utilized for the second part of the model. Health utilization data are often transformed to the log scale to shorten the long right tail, lessen heteroscedasticity, and decrease the influence of outliers. (Diehr 1999) Because the second part of the model is log transformed, the regression equation predicts log dollars rather than dollars. Exponentiation of a person's estimated log cost provides an estimate of the median cost, rather than the arithmetic mean cost. Because the mean and median costs are usually different in health care data due to the long right tail, a factor is needed to correct for the retransformation bias. A "smearing" estimator was used to retransform the predicted values back to the mean of the original distribution. (Duan 1983) The two-part model is composed of two equations, the "hurdle component" (that assesses the likelihood of incurring any cost) and the "levels component" (that deals with the size of the cost) respectively (Mullahy 1999):

1. Logit: $[\Pr(c>0|x) = \frac{e^{x\alpha}}{1+e^{x\alpha}}$ where c is cost and α is a vector of coefficients

2. $\ln(c_i) = \Sigma\, \delta_j x_{ij} + \Psi$ for c>0 where δ is a vector of coefficients and Ψ is the residual error

For the two-part model, a person's estimated cost is his probability of having any use multiplied by the expected cost conditional on being a user. The probability of having any use is estimated from the logistic regression equation. The expected cost, given some use, is estimated by exponentiating the second part estimate and multiplying it by the smearing estimate. (Diehr 1999)

The predicted annual cost per person was estimated by:

$E(C_j|x) = \Pr(c_j>0|x) \times e^{x\delta} \times \lambda_j$ where λ_j is the smearing factor for the subgroup j

MODEL DERIVATION

All model development was carried out using STATA Intercooled Version 6.0. (STATACORP 1999) The RxCost Model was estimated for both populations (commercial and Medicaid) using OLS and the two-part model (2PM). The models estimated using OLS were compared to the models estimated using the two-part modeling technique. Based on this comparison either OLS or the two-part modeling technique was chosen to estimate all models.

Even though the two-part model seems to fit the distribution better, OLS performed slightly better when predicting future prescription costs. The Root Mean Squared Error (RMSE) of each technique was compared to assess model performance (RMSE Commercial: OLS = 1,312, 2PM = 1,465; Medicaid: OLS = 2,003 2PM = 2,260). This is consistent with other work using models to predict cost. (Diehr 1999) Another advantage of OLS is that an R-square value is calculated whereas the two-part

model would not yield an interpretable R-square value (negative value). Both the typical R-square calculation method and a synthetic R-square calculation method (Ettner 1998) were used to attempt to calculate the two-part R-square value. Neither method was able to calculate an interpretable R-square value. The synthetic R-square calculation method is as follows:

$$R^2 = 1 - [\Sigma(C_j - \text{expected } C_j)^2 / \Sigma(C_j - (1/N)\ \Sigma C_j)^2]$$

Other researchers have found this to be the case as well. (Khandker 1998) Additionally, because the two-part model estimates log costs rather than actual costs, the coefficients for each category are not easily interpretable. One advantage of the two-part modeling technique is that OLS estimated a negative intercept. With a negative intercept, the reference case (male 18 to 22 in the commercial population) with no diagnostic information present would have a negative cost prediction for the following year which would be unacceptable. The intercept was then constrained to zero and the OLS models were re-estimated. The negative cost was then shifted to the age-sex variables. To circumvent this problem the age-sex variables were constrained to their mean prediction value when diagnostic information was absent. More specifically, members with no diagnostic information classified (all diagnostic variables were coded 0 or absent) were used to calculate the mean prediction value of each age-sex variable on following year's prescription cost. These values were then used to constrain the age-sex variables for everyone. This allowed OLS to predict only positive costs and resulted in no loss in predictive power in terms of R-square. OLS was chosen to estimate the RxCost Model and MRxCost Model for both populations using the

constraining method described above. The Demographic Models required no constraining as they did not predict a negative intercept.

Each model was then estimated using OLS. A stepwise selection process was utilized for all potential covariates. The stepwise variable selection combines both backward and forward selection and allows for a variable to be added or deleted from the model before the final model is attained. A variable, which might initially have appeared insignificant in the presence of some variables, might become very significant in the presence of others (and vice versa). Stepwise selection was utilized that allowed a variable to enter the model at a significance level of 0.20 and stay in the model at a significance level of 0.10. (Mantel 1970) The age-sex variables and the Medicaid eligibility variable (in the Medicaid models) were considered essential variables and were included in all models regardless of their significance to minimize model mispecification errors and to ensure every member will have a predicted cost for the following year. The hierarchical variables in the RxCost Models and MRxCost Models were grouped into clusters for the stepwise procedure so that they either all remained in the model or all fell out of the model.

A major concern in model development is model "overfitting" - including variables that may be useful predictors in the development database, but do not have the same relationship to the outcome in other databases. (Iezzoni 1997) Two main tactics were used to guard against model overfitting. First, the stepwise variable selection (discussed above) was employed to either keep or drop variables from the model based on their significance to the model. Secondly, the number of candidate variables was limited to a ratio of at least 10 cases to each predictor variable. (Harrell 1996) The use

of large databases kept data reduction techniques from being employed except in a three cases. Female infertility was dropped from the Medicaid RxCost Model due to low prevalence presumably because MediCal did not cover female infertility. Alzheimer's and iron deficiency categories were dropped from the Medicaid MRxCost Model because of formulary restrictions within MediCal.

PREDICTIVE ACCURACY

There are two terms for describing the components of predictive accuracy: calibration and discrimination. (Harrell 1996) Discrimination measures a predictor's ability to separate patients with different responses. Calibration refers to the extent of bias. There are at least three uses of measures of predictive accuracy (Harrell 1996):

1) To quantify the utility of a model to be used for prediction

2) To check the model for overfitting

3) To rank competing models

Discrimination for OLS is related to the expected squared error and to the correlation between predicted and observed responses. It can be measured by the squared multiple correlation coefficient R^2. If a predictor model has poor discrimination, no assessment or calibration can correct the model. However, if discrimination is good, the predictor can be calibrated, based on the population sample it will be applied, to reduce bias without sacrificing the discrimination.

The discrimination (R^2) was adjusted for shrinkage. Shrinkage is the flattening of the plot of predicted versus observed away from the line, caused by overfitting. Shrinkage closely relates to the concept of regression to the mean. The heuristic

shrinkage estimator of van Howelingen was utilized to adjust the R^2 value as follows (van Houwelingen 1990):

$$R^2_{adj} = 1 - (1- R^2) \times (n-1) / (n-p-1)$$ where n = the number of subjects and p = the number of candidate variables

The R^2_{adj} values are reported for each model.

MODEL VALIDATION

All of the model for both populations were frozen and tested on the validation sample. Model discrimination, the amount of variance explained by each model, was evaluated by calculating the model's R^2. The heuristic shrinkage estimator was used to adjust for shrinkage and calculate the R^2_{adj} values for each model submitted to validation.

All models have been trained using all costs with no removal or trimming techniques utilized. The frozen model estimates were also applied to the validation sample after Year 2 prescription costs were trimmed at $20K. Here, any individual's prescription drug cost of over $20K was set equal to $20K. The value of $20K was chosen because it trims only the most severe outliers that could have a substantial effect on the models. The trimming procedure only affects 0.03% (67 observations) of the MarketScan validation sample and 0.11% (100 observations) of the MediCal validation sample.

Additionally, the RxCost Model for both populations was also run in a concurrent fashion to see how well the frozen coefficients predict year 1 prescription expenditures. The MRxCost Model was not run using the concurrent approach because endogeniety was deemed to be too great. Here, the models would draw upon prescription drug information to predict prescription drug expenditures.

MODEL PERFORMANCE

The one universally reported, single-number summary performance measure for risk-assessment payment models is the R-square value; the proportion of variance in costs that the model explains. (Ash 2000) Model R-square values and R^2_{adj} values are reported for each of the models; training sample and validation sample. Training sample R-square values can be compared to the values obtained on the validation sample. If a model's R-square value in the training sample drops substantially in the validation sample then the model may be overspecified.

The R-square value described above assesses the amount of variance explained by the model for each individual member applied to the model. These models are designed to be utilized in a physician or physician group setting where a group of patients (e.g. 10, 50 or 100 patients) could be modeled to predict a single prescribing expenditure for the entire groups of patients. When predicting a prescribing expenditure for a physician or physician group based on their patients the amount of variance explained for individual patients is less important than the amount of variance explained for the entire group. Here a pooled R-square value will help assess the amount of variance the model explains for a group (or pool) of patients. Using these models in a physician or physician group setting would require the model to perform well for varying numbers of patients. To assess how well the model performs for varying numbers of patients, pooled R-square values were calculated based on groups of 10, 20, 50, 100, 200, 300, 400, and 500 patients utilizing the validation sample. To calculate the pooled R-square value for groups of 10 patients the groups were randomly selected without replacement and each of the patients actual year 2 prescription costs are summed into

one group value. Additionally, each of their predicted cost was summed into one value. Next, each group was then counted as 1 observation and an R-square value was calculated for the entire sample based on the groups of 10 patients. This procedure was replicated for groups of 10, 20, 50, 100, 200, 300, 400 and 500 patients. The ultimate goal here was to assess how many patients a physician or physician group needed to have before the model could be used to predict prescription costs effectively.

Another measure of model performance is a predictive ratio. (Ash 2000) Here, the model is applied to a subgroup of people and the predictive ratio is calculated by dividing the model-predicted costs for the group by their actual costs. Each model applied to the validation sample was used to predict costs of specific subgroups including, asthma, depression, diabetes, HIV infection and hypertension. These subgroups were identified by ICD-9-CM codes (and NDC codes in the Mixed Model) during year 1. The predicted costs were then utilized to calculate a predictive ratio. The predictive ratio was used to evaluate how well the models perform for the chosen specific subgroups. An ideal predictive ratio of 1.0 would indicate that the predicted costs and actual costs were exactly the same. As with R-square values, pooled predictive ratios were also calculated on groups of patients. Here, the random groups identified above in the validation sample were used to calculate a predictive ratio for each group and an average predictive ratio was calculated for groups of 10, 20, 50, 100, 200, 300, 400 and 500 patients. Once again, this was utilized to assess how well the models predict a single prescription expenditure for varying groups of patients.

Although there are no other risk assessment indices publicly reported for predicting pharmacy expenditures there are publicly available approaches for predicting

overall health care costs. Two such indices are Kronick's Chronic Illness and Disability Payment System (CDPS) developed for Medicaid populations and Ash's Diagnostic Cost Group Hierarchical Condition Category (DCG-HCC). Each of these models was applied prospectively to the validation samples to predict pharmacy expenditures for both the Medicaid and commercial populations. The R-square values for these models were then compared to the RxCost Model and the MRxCost Model for each population. The CDPS and the DCG-HCC models are two of the more sophisticated models available. Comparing these models allowed further evaluation of model performance. The CDPS model and the software to implement the model are free of charge and publicly available online at www.medicine.ucsd.edu/fpm/cdps. The model was downloaded and applied to the validation sample of the Medicaid population. The DCG-HCC model, and the software to implement the model, are not free of charge but are publicly available from the DxGROUP online at www.dxcg.com. The model was obtained from the DxGROUP and applied to the validation sample of the Commercial population. Because neither model was developed specifically for predicting prescription expenditures, both were recalibrated to estimate prescription costs. Without recalibration both models would substantially over predict costs. The benchmark weights were used for each model; however the models were recalibrated using a proportional calibration method explained in the DxCG Analytic Manual (DxCG Inc. 2001). This method results in the mean of the individual predicted expense to be equal to the mean observed year 1 expense. So, the models are predicting costs that are in line with prescription costs instead of overall health costs. Even after employing the recalibration technique, the RxCost Model and the MRxCost Model is still expected

to perform better (higher R^2 values) because they were estimated on the basis of prescription drug expenditures. The R^2 values of these models must be at least as large as the CDPS model for the Medicaid population and DCG-HCC model for the commercial population to conclude that they potentially perform better for predicting prescription expenditures.

REFERENCES

Adams EK, Bronstein JM, Becker ER, Hood CR. Payment levels, resource use, and insurance risk of Medicaid versus private insured in three states. Journal of Health Care Finance 2001; 28(1):72-91.

Agency for Health Care Policy and Research. Department of Health and Human Services. www.ahrq.gov

Ash A, Porell F, Gruenberg L, Sawitz E, Beiser A. Adjusting Medicare capitation payments using prior hospitalization data. Health Care Financ Rev 1989; 10(4):17-29.

Ash AS, Ellis RP, Pope GC, Ayanian JZ, Bates DW, Burstin H et al. Using diagnoses to describe populations and predict costs. Health Care Financ Rev 2000; 21(3):7-28.

Baugh DK, Pine PL, Blackwell S. Trends in Medicaid prescription drug utilization and payments, 1990-97. Health Care Financ Rev 1999; 20(3):79-105.

The California Department of Health Services. 2004; www.dhs.ca.gov.

Clark DO, Von Korff M, Saunders K, Baluch WM, Simon GE. A chronic disease score with empirically derived weights. Med Care 1995; 33(8):783-795.

Clinical Classifications Software (ICD-9-CM) Summary and Download. Summary and Downloading Information. 2001 Agency for Health Care Policy and Research, Rockville, MD. http://www.ahrq.gov/data/hcup/ccs.htm

Croghan TW, Johnstone BM, Buesching DP, Kessler RC. Information needs for medication coverage decisions in a state Medicaid program. Medical Care 1999; 37(4 Suppl):AS24-31.

Diehr P, Yanez D, Ash A, Hornbrook M, Lin DY. Methods for Analyzing Health Care Utilization and Costs. Annual Review of Public Health 1999; 20:25-144.

Duan N, Manning WG, Morris CN, Newhouse JP. A Comparison of Alternative Models for the Demand of Medical Care. Journal of Business and Economic Statistics 1983; 1(2):115-126.

DxCG Inc. DxCG Risk Adjustment Software: Analytic Guide Release 6.0. 2001; Boston MA. support@dxcg.com

Ettner SL, Frank RG, McGuire TG, Newhouse JP, Notman EH. Risk Adjustment of Mental Health and Substance Abuse Payment. Inquiry 1998; 35: 223-239.

Fishman PA, Shay DK. Development and estimation of a pediatric chronic disease score using automated pharmacy data. Medical Care 1999; 37:874-883.

Fishman PA, Goodman MJ, Hornbrook MC, Meenan RT, Bachman DJ, O'Keefe Rosetti MC. Risk Adjustment Usng Automated Ambulatory Pharmacy Data: The Risk Model. Medical Care 2003; 41(1): 84-99.

Gandhi SK, Arguelles L, Boyer JG. Economic impact of neutropenia and febrile neutropenia in breast cancer: estimates from two national databases. Pharmacotherapy 2001; 21(6):684-690.

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

Harrell FEJ, Lee KL, Mark DB. Tutorial in Biostatistics Multivariable Prognostic Models: Issues in Developing Models, Evaluating Assumptions and Adequacy, and Measuring and Reducing Errors. Statistics in Medicine 1996; 15:361-387.

Iezzoni L, Ash AS, Daley J, Hughes JS, Shwartz M. Risk adjustment for measuring healthcare outcomes. 2nd ed. Chicago: Health Administration Press, 1997.

Johnson RE, Hornbrok MC, Nichols GA. Replicating the chronic disease score (CDS) from automated pharmacy data. Journal of Clinical Epidemiology 1994; 47(10):1191-1199.

Khandker RK, Simoni-Wastila LJ. Differences in Prescription Drug Utilization and Expenditures Between Blacks and Whites in the Georgia Medicaid Population. Inquiry 1998; 35: 78-87.

Kronick R, Gilmer T, Dreyfus T, Lee L. Improving health-based payment for medicaid beneficiaries: CDPS. Health Care Financ Rev 2000; 21(3):29-64.

Lamers LM. Pharmacy cost groups: a risk-adjuster for capitation payments based on the use of prescribed drugs. Medical Care 1999; 37(8):824-830.

Mantel N. Why Stepdown procedures in variable Selection. Technometrics 1970; 12:621-625.

The MEDSTAT Group Incorporated. Ann Arbor, Michigan. 2001; www.medstat.com.

Mullahy J. Much Ado About Two: Reconsidering Retransformation and the Two-Part Model in Health Econometrics. Journal of Health Economics 1998; 17:247-281.

The Multum Lexicon. Multum Information Systems Inc., editor. 2001; www.multum.com.

Ozminkowski RJ, Wang S, Marder WD, Azzolini J, Schutt D. Cost implications for the use of inhaled anti-inflammatory medications in the treatment of asthma. Pharmacoeconomics 2000; 18(3):253-264.

SAS Institute Inc. 2004; www.sas.com

StataCorp. Stata Statistical Software: Release 6.0. 1999 College Station Texas

U.S. Department of Health and Human Services. Administration for Children and Familes News: Fact Sheets. September 2001 www.acf.dhhs.gov/news/facts/tanf.html.

van Houwelingen JC, le Cessie S. Predictive value of statistical models. Statistics in Medicine 1990; 8:1303-1325.

Von Korff M, Wagner EH, Saunders K. A chronic disease score from automated pharmacy data. J Clin Epidemiol 1992; 45(2):197-203.

Wray NP, Hollingsworth JC, Peterson NJ, Asthon CM. Case-mix adjustment using administrative databases: a paradigm to guide future research. Medical Care Res Rev 1997; 54(3): 326-356

Zhao SZ, Arguelles LM, Dedhiya SD, Morgan DG. Healthcare utilization associated with dyspepsia in patients with arthritis. Am J Manag Care 1999; 5(10):1285-1295.

Table 3.1

MarketScan Data Counts	
Year	**Members**
1998	7,034,519
1999	8,086,185
2000	9,039,018
Total Unduplicated Count	**9,043,605**

Table 3.2

MediCal Data Counts	
Year	**Members**
1998	1,238,483
1999	1,250,809
2000	1,300,871
Total Unduplicated Count	**1,689,019**

Table 3.3

Data Counts Resulting from MarketScan Eligibility Requirements	
Total unduplicated count	9,043,605
Count after continuous eligibility requirements	3,016,354
Count after deleting people without drug coverage reported	2,320,553
Count after deleting members not 18 to 62 years of age	1,644,988
Count after deleting members admitted to institutions or with inpatient stays > 30 days	1,634,427

Table 3.4

MarketScan Eligibility Counts - Training, Validation and Spare Samples	
Training sample	1,308,705
Validation sample	218,383
Spare sample	107,339
Total	**1,634,427**

Table 3.5

Data Counts Resulting from MediCal Eligibility Requirements	
Total unduplicated count	1,689,019
Count after deleting members with Medicare eligibility	1,474,660
Count after continuous eligibility requirements	874,102
Count after deleting members not 18 to 62 years of age	305,877
Count after deleting members not eligible for AFDC/TANF or SSI/Disability	280,474
Count after deleting members admitted to institutions or with inpatient stays > 30 days	276,518

Table 3.6

MediCal Eligibility Counts - Training, Validation and Spare Samples	
Training sample	138,454
Validation sample	92,621
Spare sample	45,443
Total	**276,518**

Table 3.7

Revised CCS-Based Classification Based on Prior Risk Assessment Model Use

1. Infectious and parasitic diseases
2. HIV Infection (previously CCS 1.1.3)
3. Neoplasms
4. Diabetes Mellitus (previously CCS 3)
5. Other endocrine, nutritional, and metabolic diseases and immunity disorders (previously CCS 3)
6. Diseases of the blood and blood-forming organs
7. Mental retardation (previously CCS 5)
8. Senility and organic mental disorders (previously CCS 5)
9. Other mental disorders (previously CCS 5)
10. Alcohol and substance-related mental disorders
11. Eye Disorders (previously CCS 6)
12. Ear conditions (previously CCS 6)
13. Diseases of the Central nervous system and other sense organs (previously CCS 6)
14. Hypertension (previously CCS 7)
15. Heart valve disorders (previously CCS 7)
16. Acute myocardial infarction (previously CCS 7)
17. Coronary atherosclerosis and other heart disease (previously CCS 7)
18. Congestive heart failure (nonhypertensive) and pulmonary heart disease (previously CCS 7)
19. Cardiac dysrhythmias (previously CCS 7)
20. Cerebrovascular disease (previously CCS 7)
21. Diseases of arteries, arterioles, capillaries, veins and lymphatics (previously CCS 7)
22. Other diseases of the circulatory system (previously CCS 7)
23. Diseases of the respiratory system
24. Diseases of the digestive system (previously CCS 9)
25. Liver disease (previously CCS 9)
26. Diseases of the urinary system (previously CCS 10)
27. Diseases of the genitourinary system (previously CCS 10)
28. Complications of pregnancy, childbirth, and the puerperium
29. Diseases of the skin and subcutaneous tissue
30. Diseases of the musculoskeletal system and connective tissue
31. Congenital anomalies
32. Certain conditions originating in the perinatal period
33. Injury and poisoning
34. Symptoms, signs, and ill-defined conditions and factors influencing health status

Table 3.8

CCS-Based Categories Discarded: Categories that Represent Ill-Defined Conditions or Low Prescription Drug Cost	
1.2.1 Candidiasis of the mouth (thrush)	27.3.1 Ectopic pregnancy
1.3.3 Other viral infections	27.3.2 Hemorrhage during pregnancy, abruptio placenta, placenta previa
1.5 Immunizations and screening for infectious disease	27.3.4 Early or threatened labor
4.2 Other endocrine disorders	27.3.5 Prolonged pregnancy
4.3 Nutritional deficiencies	27.3.7 Other complications of pregnancy
4.6 Fluid and electrolyte disorders	27.4.1 Malposition, malpresentation
8.5 Preadult disorders	27.4.2 Fetopelvic disproportion, obstruction
8.7 Personal history of mental disorder, screening for mental condition	27.4.3 Previous cesarean section
12.2.3 Other hereditary and degenerative nervous system conditions	27.4.4 Fetal distress and abnormal forces of labor
12.5.2 Other headache	27.4.5 Polyhydramnios and other problems of amniotic cavity
12.6 Coma, stupor, and brain damage	27.5 Complications during labor
12.7 Other nervous system disorders	27.5.2 Trauma to perineum and vulva
20.2 Aortic, peripheral, and visceral artery aneurysms	27.5.3 Forceps delivery [194.]
20.3 Aortic and peripheral arterial embolism or thrombosis	27.6 Other complications of birth, puerperium affecting management of mother
20.7 Hemorrhoids	27.7 Normal pregnancy and/or delivery
20.8 Other diseases of veins and lymphatics	28.4 Other skin disorders
21.5 Cardiac arrest and ventricular fibrillation	29.6 Acquired deformities
22.4 Aspiration pneumonitis, food/vomitus	30.2 Digestive congenital anomalies
22.5 Pleurisy, pneumothorax, pulmonary collapse	30.3 Genitourinary congenital anomalies
22.6 Respiratory failure, insufficiency, arrest (adult)	30.5 Other congenital anomalies
23.2 Disorders of teeth and jaw	32.1 Joint disorders and dislocations, trauma-related
23.3 Diseases of mouth, excluding dental	32.2 Fractures
23.5 Abdominal hernia	32.7 Sprains and strains
25.2 Acute and unspecified renal failure	32.8 Superficial injury, contusion
25.7 Other diseases of bladder and urethra	32.10.1 Complication of device, implant or graft
25.8 Genitourinary symptoms and ill-defined conditions	32.11 Poisoning
26.2.6 Ovarian cyst	32.12 Other injuries and conditions due to external causes
26.2.9 Other female genital disorders	33.2 Factors influencing health care
27.2 Abortion-related disorders	

* All subcategories below these categories were discarded as well

Table 3.9

New CCS-Based Classification
1. Infectious and parasitic diseases
1.1 Bacterial infection
1.1.1 Tuberculosis
1.1.2 Septicemia (except in labor)
1.1.3 Sexually transmitted infections (not HIV or hepatitis)
1.1.4 Other bacterial infections
1.2 Mycoses other than candidiasis
1.3 Hepatitis
2. HIV Infection
3. Neoplasms
4. Diabetes Mellitus
4.1 Diabetes mellitus without complication
4.2 Diabetes mellitus with other complications
5. Other endocrine, nutritional, and metabolic diseases and immunity disorders
5.1 Thyroid disorders
5.2 Disorders of lipid metabolism
5.3 Gout and other crystal arthropathies
5.4 Cystic fibrosis
5.5 Immunity disorders
5.6 Other nutritional, endocrine, and metabolic disorders
6. Diseases of the blood and blood-forming organs
6.1 Anemia
6.2 Coagulation and hemorrhagic disorders
6.3 Diseases of white blood cells AND other hematologic conditions
7. Mental retardation
8. Senility and organic mental disorders
9. Other mental disorders
9.1 Affective disorders
9.2 Schizophrenia and related disorders and other psychoses
9.3 Anxiety, somatoform, dissociative, and personality disorders
9.4 Other mental conditions
10. Alcohol and substance-related mental disorders
11. Eye Disorders
11.1 Glaucoma
11.2 Other eye disorders
12. Ear conditions
12.1 Otitis media and related conditions
12.2 Conditions associated with dizziness or vertigo AND other sense organ disorders

13. Diseases of the Central nervous system and other sense organs
13.1 Central nervous system infection
13.2 Hereditary and degenerative nervous system conditions
13.2.1 Parkinson's disease
13.2.2 Multiple sclerosis
13.3 Paralysis
13.4 Epilepsy, convulsions
13.5 Migraine
14. Hypertension
15. Heart valve disorders
16. Acute myocardial infarction
17. Coronary atherosclerosis and other heart disease
18. Congestive heart failure (nonhypertensive) and pulmonary heart disease
19. Cardiac dysrhythmias
20. Cerebrovascular disease
21. Diseases of arteries, arterioles, capillaries, veins and lymphatics
21.1 Peripheral and visceral atherosclerosis
21.2 Hypotention and Other unspecified circulatory disease
21.3 Phlebitis, thrombophlebitis and thromboembolism AND Varicose veins of lower extremity
22. Other diseases of the circulatory system
22.1 Peri-, endo-, and myocarditis, cardiomyopathy (except that caused by TB or STD)
22.2 Conduction disorders
23. Diseases of the respiratory system
23.1 Respiratory infections
23.1.1 Pneumonia (except that caused by TB or STD)
23.1.2 Influenza, Acute and chronic tonsillitis, Acute bronchitis AND other URIs
23.2 Chronic obstructive pulmonary disease and bronchiectasis
23.3 Asthma
23.4 Lung disease due to external agents
23.5 Other lower respiratory disease
23.6 Other upper respiratory disease
24. Diseases of the digestive system
24.1 Intestinal infection
24.2 Upper gastrointestinal disorders
24.2.1 Esophageal disorders
24.2.2 Gastroduodenal ulcer (except hemorrhage) AND Gastritis and duodenitis
24.2.3 Other disorders of stomach and duodenum
24.3 Lower gastrointestinal disorders AND Biliary tract disease
24.4 Pancreatic disorders (not diabetes)
24.5 Constipation, Dysphagia, and Other unspecified GI disorders
25. Liver disease

26. Diseases of the urinary system
26.1 Nephritis, nephrosis, renal sclerosis
26.2 Chronic renal failure
26.3 Urinary tract infections, Calculus of the urinary tract AND Other diseases of the kidney and ureters
27. Diseases of the genitourinary system
27.1 Diseases of male genital organs
27.2 Diseases of female genital organs
27.2.1 Nonmalignant breast conditions, Inflammatory diseases of female pelvic organs, Endometriosis, AND Prolapse of female genital organs
27.2.2 Menstrual disorders AND Menopausal disorders
27.2.3 Female infertility
28. Complications of pregnancy, childbirth, and the puerperium
28.1 Contraceptive and procreative management
28.2 Complications mainly related to pregnancy
28.2.1 Hypertension complicating pregnancy, childbirth and the puerperium
28.2.2 Diabetes or abn. glucose tolerance complicating pregn., childbirth, or the puerperium
29. Diseases of the skin and subcutaneous tissue
29.1 Skin and subcutaneous tissue infections AND Other inflammatory conditions of skin
29.2 Chronic ulcer of skin
30. Diseases of the musculoskeletal system and connective tissue
30.1 Infective arthritis and osteomyelitis (except that caused by TB or STD)
30.2 Non-traumatic joint disorders
30.3 Spondylosis, intervertebral disc disorders, other back problems
30.4 Osteoporosis AND Pathological fracture
30.5 Systemic lupus erythematosus and connective tissue disorders, Other connective tissure disease AND Other bone disease and musculoskeletal deformities
31. Congenital anomalies
31.1 Cardiac and circulatory congenital anomalies
31.2 Nervous system congenital anomalies
32. Certain conditions originating in the perinatal period
33. Injury and poisoning
33.1 Spinal cord injury
33.2 Intracranial injury
33.3 Crushing injury or internal injury
33.4 Open wounds
33.5 Burns
33.6 Complications of surgical procedures or medical care
34. Symptoms, signs, and ill-defined conditions and factors influencing health status

Table 3.10

New CCS-Based Classification: Frequencies Training Sample (n=1,308,705)	Freqency in MarketScan (All Claims)*			Freqency in MarketScan (FFS Only)		
Classification	Freq. of all Diagnoses	Freq. Of Unique Patients	Prevelance of Unique Patients	Freq. of all Diagnoses	Freq. Of Unique Patients	Prevelance of Diagnosis
1. Infectious and parasitic diseases						
1.1 Bacterial infection						
1.1.1 Tuberculosis	1,694	601	0.05%	1,027	404	0.05%
1.1.2 Septicemia (except in labor)	4,743	1,071	0.08%	3,087	731	0.09%
1.1.3 Sexually transmitted infections (not HIV or hepatitis)	7,051	2,114	0.16%	4,535	1,410	0.16%
1.1.4 Other bacterial infections	10,523	3,478	0.27%	7,326	2,442	0.28%
1.2 Mycoses other than candidiasis	88,934	32,698	2.50%	58,739	21,947	2.56%
1.3 Hepatitis	61,336	5,588	0.43%	35,945	3,490	0.41%
2. HIV Infection	30,294	1,087	0.08%	18,314	630	0.07%
3. Neoplasms	1,479,204	162,447	12.41%	1,001,955	111,754	13.02%
4. Diabetes Mellitus						
4.1 Diabetes mellitus without complication	475,142	47,941	3.66%	314,921	33,653	3.92%
4.2 Diabetes mellitus with other complications	141,771	17,438	1.33%	97,553	12,061	1.41%
5. Other endocrine, nutritional, and metabolic diseases and immunity disorders						
5.1 Thyroid disorders	339,801	55,201	4.22%	225,152	39,519	4.61%
5.2 Disorders of lipid metabolism	804,982	131,872	10.08%	476,124	88,556	10.32%
5.3 Gout and other crystal arthropathies	28,568	6,141	0.47%	18,903	4,174	0.49%
5.4 Cystic fibrosis	3,978	295	0.02%	3,046	212	0.02%
5.5 Immunity disorders	8,311	956	0.07%	6,384	675	0.08%
5.6 Other nutritional, endocrine, and metabolic disorders	147,158	31,673	2.42%	84,692	19,123	2.23%

6. Diseases of the blood and blood-forming organs						
6.1 Anemia	210,780	29,734	2.27%	137,243	21,511	2.51%
6.2 Coagulation and hemorrhagic disorders	40,904	4,296	0.33%	28,408	3,048	0.36%
6.3 Diseases of white blood cells AND other hematologic conditions	39,252	5,741	0.44%	25,370	3,972	0.46%
7. Mental retardation	794	151	0.01%	349	88	0.01%
8. Senility and organic mental disorders	9,409	1,775	0.14%	6,464	1,202	0.14%
9. Other mental disorders						
9.1 Affective disorders	385,647	37,536	2.87%	269,253	24,885	2.90%
9.2 Schizophrenia and related disorders and other psychoses	12,828	1,686	0.13%	8,680	1,113	0.13%
9.3 Anxiety, somatoform, dissociative, and personality disorders	204,458	32,551	2.49%	136,974	20,303	2.37%
9.4 Other mental conditions	297,788	45,039	3.44%	193,880	27,927	3.25%
10. Alcohol and substance-related mental disorders	44,094	9,719	0.74%	22,652	4,706	0.55%
11. Eye Disorders						
11.1 Glaucoma	56,695	16,542	1.26%	38,648	11,480	1.34%
11.2 Other eye disorders	311,322	98,924	7.56%	216,205	69,868	8.14%
12. Ear conditions						
12.1 Otitis media and related conditions	95,445	39,709	3.03%	63,133	26,321	3.07%
12.2 Conditions associated with dizziness or vertigo AND other sense organ disorders	215,858	59,475	4.54%	137,849	39,220	4.57%
13. Diseases of the Central nervous system and other sense organs						
13.1 Central nervous system infection	6,954	1,113	0.09%	4,511	784	0.09%
13.2 Hereditary and degenerative nervous system conditions						
13.2.1 Parkinson's disease	3,211	502	0.04%	2,085	352	0.04%
13.2.2 Multiple sclerosis	31,269	2,426	0.19%	19,286	1,653	0.19%

13.3 Paralysis	13,751	1,462	0.11%	8,443	999	0.12%
13.4 Epilepsy, convulsions	49,682	6,565	0.50%	33,070	4,490	0.52%
13.5 Migraine	93,162	19,978	1.53%	66,933	13,677	1.59%
14. Hypertension	886,487	147,513	11.27%	573,066	101,589	11.84%
15. Heart valve disorders	97,083	18,136	1.39%	64,207	12,708	1.48%
16. Acute myocardial infarction	19,861	2,962	0.23%	11,916	1,957	0.23%
17. Coronary atherosclerosis and other heart disorders	275,823	30,390	2.32%	183,331	21,505	2.51%
18. Congestive heart failure (nonhypertensive) and pulmonary heart disease	43,304	5,848	0.45%	29,705	4,194	0.49%
19. Cardiac dysrhythmias	214,163	33,949	2.59%	143,913	23,582	2.75%
20. Cerebrovascular disease	60,224	8,634	0.66%	41,050	6,083	0.71%
21. Diseases of arteries, arterioles, capillaries, veins and lymphatics						
21.1 Peripheral and visceral atherosclerosis	22,465	4,424	0.34%	14,927	3,075	0.36%
21.2 Hypotention and Other unspecified circulatory disease	98,035	25,230	1.93%	65,085	16,742	1.95%
21.3 Phlebitis, thrombophlebitis and thromboembolism AND Varicose veins of lower extremity	50,309	9,179	0.70%	33,915	6,401	0.75%
22. Other diseases of the circulatory system						
22.1 Peri-, endo-, and myocarditis, cardiomyopathy (except that caused by TB or STD)	24,556	3,754	0.29%	16,668	2,617	0.30%
22.2 Conduction disorders	9,565	2,009	0.15%	6,354	1,419	0.17%
23. Diseases of the respiratory system						
23.1 Respiratory infections						
23.1.1 Pneumonia (except that caused by TB or STD)	63,748	14,109	1.08%	44,239	9,643	1.12%
23.1.2 Influenza, Acute and chronic tonsillitis, Acute bronchitis AND other URIs	944,248	295,324	22.57%	660,034	201,052	23.43%

23.2 Chronic obstructive pulmonary disease and bronchiectasis	142,084	39,352	3.01%	94,526	26,687	3.11%
23.3 Asthma	170,726	33,690	2.57%	108,372	21,946	2.56%
23.4 Lung disease due to external agents	2,616	904	0.07%	1,831	644	0.08%
23.5 Other lower respiratory disease	293,076	77,242	5.90%	196,081	51,471	6.00%
23.6 Other upper respiratory disease	698,355	104,715	8.00%	499,565	71,508	8.33%
24. Diseases of the digestive system						
24.1 Intestinal infection	17,555	6,264	0.48%	12,220	4,191	0.49%
24.2 Upper gastrointestinal disorders						
24.2.1 Esophageal disorders	153,233	37,722	2.88%	102,745	24,990	2.91%
24.2.2 Gastroduodenal ulcer (except hemorrhage) AND Gastritis and duodenitis	85,627	20,868	1.59%	61,351	14,380	1.68%
24.2.3 Other disorders of stomach and duodenum	30,043	9,255	0.71%	20,581	6,283	0.73%
24.3 Lower gastrointestinal disorders AND Biliary tract disease	213,051	36,321	2.78%	148,646	24,723	2.88%
24.4 Pancreatic disorders (not diabetes)	13,762	1,620	0.12%	8,852	1,096	0.13%
24.5 Constipation, Dysphagia, and Other unspecified GI disorders	176,818	48,401	3.70%	118,631	32,449	3.78%
25. Liver disease	85,030	13,241	1.01%	53,132	8,797	1.03%
26. Diseases of the urinary system						
26.1 Nephritis, nephrosis, renal sclerosis	12,426	1,357	0.10%	7,838	894	0.10%
26.2 Chronic renal failure	82,155	1,966	0.15%	57,417	1,399	0.16%
26.3 Urinary tract infections, Calculus of the urinary tract AND Other diseases of the kidney and ureters	434,891	77,992	5.96%	307,241	54,823	6.39%
27. Diseases of the genitourinary system						
27.1 Diseases of male genital organs	220,894	46,051	3.52%	141,748	31,278	3.64%

27.2 Diseases of female genital organs						
27.2.1 Nonmalignant breast conditions, Inflammatory diseases of female pelvic organs, Endometriosis, AND Prolapse of female genital organs	513,502	128,445	9.81%	363,682	90,242	10.52%
27.2.2 Menstrual disorders AND Menopausal disorders	499,049	116,960	8.94%	356,725	83,557	9.74%
27.2.3 Female infertility	59,222	5,472	0.42%	35,342	3,430	0.40%
28. Complications of pregnancy, childbirth, and the puerperium						
28.1 Contraceptive and procreative management	64,725	22,241	1.70%	35,888	12,270	1.43%
28.2 Complications mainly related to pregnancy			0.00%			
28.2.1 Hypertension complicating pregnancy, childbirth and the puerperium	11,259	1,767	0.14%	6,937	1,152	0.13%
28.2.2 Diabetes or abn. glucose tolerance complicating pregn., childbirth, or the puerperium	15,041	2,217	0.17%	9,075	1,407	0.16%
29. Diseases of the skin and subcutaneous tissue						
29.1 Skin and subcutaneous tissue infections AND Other inflammatory conditions of skin	194,477	56,271	4.30%	130,075	38,174	4.45%
29.2 Chronic ulcer of skin	22,193	2,760	0.21%	15,425	2,036	0.24%
30. Diseases of the musculoskeletal system and connective tissue						
30.1 Infective arthritis and osteomyelitis (except that caused by TB or STD)	12,273	1,664	0.13%	8,399	1,175	0.14%
30.2 Non-traumatic joint disorders	775,346	123,431	9.43%	517,505	28,087	3.27%
30.3 Spondylosis, intervertebral disc disorders, other back problems	1,482,637	127,599	9.75%	989,708	85,038	9.91%
30.4 Osteoporosis AND Pathological fracture	42,177	15,396	1.18%	30,354	11,394	1.33%
30.5 Systemic lupus erythematosus and connective tissue disorders, Other connective tissue disease AND Other bone disease and musculoskeletal deformities	1,377,727	170,314	13.01%	917,144	114,244	13.31%

31. Congenital anomalies						
31.1 Cardiac and circulatory congenital anomalies	12,184	2,632	0.20%	7,907	1,803	0.21%
31.2 Nervous system congenital anomalies	2,511	532	0.04%	1,676	358	0.04%
32. Certain conditions originating in the perinatal period	16,539	6,836	0.52%	6,154	2,837	0.33%
33. Injury and poisoning						
33.1 Spinal cord injury	2,685	683	0.05%	1,730	472	0.06%
33.2 Intracranial injury	10,169	2,616	0.20%	6,626	1,676	0.20%
33.3 Crushing injury or internal injury	7,166	2,138	0.16%	4,679	1,445	0.17%
33.4 Open wounds	110,633	26,680	2.04%	73,145	17,264	2.01%
33.5 Burns	642	126	0.01%	433	88	0.01%
33.6 Complications of surgical procedures or medical care	28,127	6,620	0.51%	19,298	4,497	0.52%
34. Symptoms, signs, and ill-defined conditions and factors influencing health status	1,117,451	210,780	16.11%	735,621	141,424	16.48%
Total Diagnosis Claims	17,776,741	3,108,127		11,879,854	2,052,576	

*** Claims consist of both Fee For Service and Encounter Claims**
Members: (FFS = 858,166 (66%) & Encounter = 450,539 (34%))

Table 3.11

New CCS-Based Classification: Frequencies (MediCal) Training Sample n = 92,621	Freqency in MediCal		
Classification	Freq. Of all Diagnoses	Freq. of Unique Patients	Pevelance of Diagnosis
1. Infectious and parasitic diseases			
1.1 Bacterial infection			
1.1.1 Tuberculosis	3,543	305	0.3293%
1.1.2 Septicemia (except in labor)	2,976	442	0.4772%
1.1.3 Sexually transmitted infections (not HIV or hepatitis)	2,291	650	0.7018%
1.1.4 Other bacterial infections	2,723	612	0.6608%
1.2 Mycoses other than candidiasis	12,691	4,219	4.5551%
1.3 Hepatitis	24,529	2,502	2.7013%
2. HIV Infection	23,991	761	0.8216%
3. Neoplasms	88,343	5,346	5.7719%
4. Diabetes Mellitus			
4.1 Diabetes mellitus without complication	60,906	5,401	5.8313%
4.2 Diabetes mellitus with other complications	22,333	2,383	2.5729%
5. Other endocrine, nutritional, and metabolic diseases and immunity disorders			
5.1 Thyroid disorders	29,390	3,862	4.1697%
5.2 Disorders of lipid metabolism	34,696	5,365	5.7924%
5.3 Gout and other crystal arthropathies	2,093	346	0.3736%
5.4 Cystic fibrosis	1,224	31	0.0335%
5.5 Immunity disorders	1,484	143	0.1544%
5.6 Other nutritional, endocrine, and metabolic disorders	20,682	3,542	3.8242%

6. Diseases of the blood and blood-forming organs			
6.1 Anemia	107,435	5,658	6.1088%
6.2 Coagulation and hemorrhagic disorders	3,669	394	0.4254%
6.3 Diseases of white blood cells AND other hematologic conditions	4,144	494	0.5334%
7. Mental retardation	12,909	865	0.9339%
8. Senility and organic mental disorders	2,350	552	0.5960%
9. Other mental disorders			
9.1 Affective disorders	11,926	2,305	2.4886%
9.2 Schizophrenia and related disorders and other psychoses	16,802	1,947	2.1021%
9.3 Anxiety, somatoform, dissociative, and personality disorders	18,597	4,622	4.9902%
9.4 Other mental conditions	18,349	3,936	4.2496%
10. Alcohol and substance-related mental disorders	18,395	2,820	3.0447%
11. Eye Disorders			
11.1 Glaucoma	3,686	908	0.9803%
11.2 Other eye disorders	59,661	16,347	17.6493%
12. Ear conditions			
12.1 Otitis media and related conditions	10,127	3,335	3.6007%
12.2 Conditions associated with dizziness or vertigo AND other sense organ disorders	30,886	6,653	7.1830%
13. Diseases of the Central nervous system and other sense organs			
13.1 Central nervous system infection	1,828	217	0.2343%
13.2 Hereditary and degenerative nervous system conditions			
13.2.1 Parkinson's disease	534	65	0.0702%
13.2.2 Multiple sclerosis	2,592	161	0.1738%
13.3 Paralysis	14,707	870	0.9393%
13.4 Epilepsy, convulsions	28,209	2,428	2.6214%
13.5 Migraine	12,407	2,092	2.2587%
14. Hypertension	123,376	10,570	11.4121%
15. Heart valve disorders	22,210	3,162	3.4139%

16. Acute myocardial infarction	3,699	404	0.4362%
17. Coronary atherosclerosis and other heart disease	29,157	3,017	3.2574%
18. Congestive heart failure (nonhypertensive) and pulmonary heart diseases	18,964	1,602	1.7296%
19. Cardiac dysrhythmias	28,391	4,286	4.6275%
20. Cerebrovascular disease	10,390	1,786	1.9283%
21. Diseases of arteries, arterioles, capillaries, veins and lymphatics			
21.1 Peripheral and visceral atherosclerosis	8,569	1,551	1.6746%
21.2 Hypotention and Other unspecified circulatory disease	9,044	2,133	2.3029%
21.3 Phlebitis, thrombophlebitis and thromboembolism AND Varicose veins of lower extremity	14,921	2,696	2.9108%
22. Other diseases of the circulatory system			
22.1 Peri-, endo-, and myocarditis, cardiomyopathy (except that caused by TB or STD)	4,895	754	0.8141%
22.2 Conduction disorders	1,755	259	0.2796%
23. Diseases of the respiratory system			
23.1 Respiratory infections			
23.1.1 Pneumonia (except that caused by TB or STD)	13,680	1,928	2.0816%
23.1.2 Influenza, Acute and chronic tonsillitis, Acute bronchitis AND other URIs	100,892	19,509	21.0633%
23.2 Chronic obstructive pulmonary disease and bronchiectasis	51,299	7,030	7.5901%
23.3 Asthma	34,655	4,712	5.0874%
23.4 Lung disease due to external agents	249	91	0.0982%
23.5 Other lower respiratory disease	87,398	10,247	11.0634%
23.6 Other upper respiratory disease	23,573	6,372	6.8796%
24. Diseases of the digestive system			
24.1 Intestinal infection	2,181	755	0.8151%
24.2 Upper gastrointestinal disorders			
24.2.1 Esophageal disorders	10,814	2,595	2.8017%
24.2.2 Gastroduodenal ulcer (except hemorrhage) AND Gastritis and duodenitis	27,779	4,576	4.9406%
24.2.3 Other disorders of stomach and duodenum	4,697	1,507	1.6271%

24.3 Lower gastrointestinal disorders AND Biliary tract disease	22,575	2,809	3.0328%
24.4 Pancreatic disorders (not diabetes)	5,347	435	0.4697%
24.5 Constipation, Dysphagia, and Other unspecified GI disorders	21,150	4,817	5.2008%
25. Liver disease	20,067	2,074	2.2392%
26. Diseases of the urinary system			
26.1 Nephritis, nephrosis, renal sclerosis	2,187	204	0.2203%
26.2 Chronic renal failure	50,693	413	0.4459%
26.3 Urinary tract infections, Calculus of the urinary tract AND Other diseases of the kidney and ureters	65,883	8,686	9.3780%
27. Diseases of the genitourinary system			
27.1 Diseases of male genital organs	20,934	1,375	1.4845%
27.2 Diseases of female genital organs			
27.2.1 Nonmalignant breast conditions, Inflammatory diseases of female pelvic organs, organs Endometriosis, AND Prolapse of female genital organs	49,854	8,009	8.6471%
27.2.2 Menstrual disorders AND Menopausal disorders	31,814	7,732	8.3480%
28. Complications of pregnancy, childbirth, and the puerperium			
28.1 Contraceptive and procreative management	25,908	5,999	6.4769%
28.2 Complications mainly related to pregnancy			
28.2.1 Hypertension complicating pregnancy, childbirth and the puerperium	1,808	394	0.4254%
28.2.2 Diabetes or abn. glucose tolerance complicating pregnancy, childbirth, or the puerperium	5,098	570	0.6154%
29. Diseases of the skin and subcutaneous tissue			
29.1 Skin and subcutaneous tissue infections AND Other inflammatory conditions of the skin	21,539	4,628	4.9967%
29.2 Chronic ulcer of skin	3,528	461	0.4977%

30. Diseases of the musculoskeletal system and connective tissue			
30.1 Infective arthritis and osteomyelitis (except that caused by TB or STD)	1,233	216	0.2332%
30.2 Non-traumatic joint disorders	87,778	12,214	13.1871%
30.3 Spondylosis, intervertebral disc disorders, other back problems	145,788	13,394	14.4611%
30.4 Osteoporosis AND Pathological fracture	2,432	854	0.9220%
30.5 Systemic lupus erythematosus and connective tissue disorders, Other connective tissure disease AND Other bone disease and musculoskeletal deformities	70,751	11,718	12.6516%
31. Congenital anomalies			
31.1 Cardiac and circulatory congenital anomalies	4,689	833	0.8994%
31.2 Nervous system congenital anomalies	1,991	140	0.1512%
32. Certain conditions originating in the perinatal period	34,908	5,666	6.1174%
33. Injury and poisoning			
33.1 Spinal cord injury	1,295	179	0.1933%
33.2 Intracranial injury	3,149	506	0.5463%
33.3 Crushing injury or internal injury	1,110	289	0.3120%
33.4 Open wounds	14,948	2,868	3.0965%
33.5 Burns	251	30	0.0324%
33.6 Complications of surgical procedures or medical care	4,558	911	0.9836%
34. Symptoms, signs, and ill-defined conditions and factors influencing health status	220,195	20,893	22.5575%
Total	2,289,183	303,438	

Table 3.12

Phi Coefficients for Correlated RxCost Model Variables

Variable1	Variable 2	Phi Coefficient
CCS 4.1 (diabetes w/o complications)	CCS 4.2 (diabetes w/ complications)	0.4242
CCS 5.2 (disorders of lipid metabolism)	CCS 14 (hypertension)	0.2382
CCS 16 (acute myocardial infarction)	CCS17 (cornary atherosclerosis and other heart disease)	0.2110
CCS 23.1.2 (influenza, tonsilitis, acute bronchitis and other URIs)	CCS 23.6 (other upper respiratory disease)	0.2121
CCS 24.2.1 (Esophageoal disorders)	CCS24.2.2 (gastroduodenal ulcer AND gastritis and duodenitis)	0.2143
CCS 27.2.1 (nonmalignant breast conditions, inflammatory disease of female pelvic organs, endometriosis AND prolapse of female genital organs)	CCS 27.2.2 (menstrual disorders AND menopausal disorders)	0.2090
CCS 30.2 (non-traumatic joint disorders)	CCS 30.5 (systemic lupus erythematosus connective tissue disorders, other connective tissue disease AND other bone disease and musculoskeletal deformities)	0.2971
CCS 30.3 (spondylosis, intervetebral disc disorders, other back problems)	CCS 30.5 (systemic lupus erythematosus connective tissue disorders, other connective tissue disease AND other bone disease and musculoskeletal deformities)	0.2282

* Results based on MarketScan Training Sample

Table 3.13

Commercial Population: Hierarchical Categories and Prescription Drug Costs	
CCS-Based Category	***Avg Annual Rx Cost per Patient with Diagnosis**
4.2 Diabetes mellitus with other complications	$863
4.1 Diabetes mellitus without complication	$695
9.1 Affective disorders	$1,535
9.3 Anxiety, somatoform, dissociative, and personality disorders	$955
9.4 Other mental conditions	$901
23.6 Other upper respiratory disease	$942
23.1.2 Influenza, Acute and chronic tonsillitis, Acute bronchitis AND other URIs	$682
24.2.1 Esophageal disorders	$1,304
24.2.2 Gastroduodenal ulcer (except hemorrhage) AND Gastritis and duodenitis	$1,084
27.2.1 Nonmalignant breast conditions, Inflammatory diseases of female pelvic organs, Endometriosis, AND Prolapse of female genital organs	$769
27.2.2 Menstrual disorders AND Menopausal disorders	$758
30.1 Infective arthritis and osteomyelitis (except that caused by TB or STD)	$1,836
30.4 Osteoporosis AND Pathological fracture	$1,242
30.2 Non-traumatic joint disorders	$1,057
30.5 Systemic lupus erythematosus and connective tissue disorders, Other connective tissue disease AND Other bone disease and musculoskeletal deformities	$835
30.3 Spondylosis, intervertebral disc disorders, other back problems	$735

* Annual Rx cost associated with each condition:
based on 20% random sample of MarketScan training data

** Average cost for all members for all conditions using 20% sample = $497

Table 3.14

Medicaid Population: Hierarchical Categories and Prescription Drug Costs	
CCS-Based Category	***Avg Annual Rx Cost per Patient with Diagnosis**
4.2 Diabetes mellitus with other complications	$530
4.1 Diabetes mellitus without complication	$509
9.4 Other mental conditions	$2,533
9.1 Affective disorders	$2,366
9.3 Anxiety, somatoform, dissociative, and personality disorders	$1,829
23.6 Other upper respiratory disease	$1,568
23.1.2 Influenza, Acute and chronic tonsillitis, Acute bronchitis AND other URIs	$1,347
24.2.1 Esophageal disorders	$2,458
24.2.2 Gastroduodenal ulcer (except hemorrhage) AND Gastritis and duodenitis	$1,901
27.2.1 Nonmalignant breast conditions, Inflammatory diseases of female pelvic organs, Endometriosis, AND Prolapse of female genital organs	$1,260
27.2.2 Menstrual disorders AND Menopausal disorders	$1,012
30.4 Osteoporosis AND Pathological fracture	$3,102
30.5 Systemic lupus erythematosus and connective tissue disorders, Other connective tissue disease AND Other bone disease and musculoskeletal deformities	$2,172
30.2 Non-traumatic joint disorders	$1,864
30.1 Infective arthritis and osteomyelitis (except that caused by TB or STD)	$1,684
30.3 Spondylosis, intervertebral disc disorders, other back problems	$1,087

* Annual Rx cost associated with each condition:
based on 20% random sample of MediCal training data

** Average cost for all members for all conditions using 20% sample = $510

Table 3.15

Gilmer's Medicaid Prescription Model

Categories	Drug Description
Alcoholism	Disulfiram
Alzheimers	Tacrine
Anticoagulants	Heparins
Burns	Silver Sulfadiazine
Cardiovascular	ACE inhibitors, beta blockers, nitrates, digitalis, vasodilators
Cystic fibrosis	Pancrelipase
Depression/anxiety	Antidepressants, antianxiety
Diabetes	Insulin, sulfonylureas
ESRD/renal	Erythropoietin, Calcitriol
Folate deficiency	Folic acid
Gallstones	Ursodiol
Gastric acid disorders	Cimetidine
Gout	Colchicine
Hemophilia/von Willebrands	Factor IX concentrates
Hepatitis	Interferon beta
Herpes	Acyclovir
HIV/AIDS	Antiretrovirals
Hyperlipidemia	Antihyperlipidemics
Infections, high	Aminogycosides
Infections, medium	Vancomycin, Fluoroquinolones
Infections, low	Cephalosporins, Erythromycins
Inflammatory/autoimmune	Glucocorticosteroids
Insomnia	Sedatives, Hypnotics
Iron deficiency	Iron
Irrigating solutions	Sodium chloride
Liver disease	Lactulose
Malignancies	Antineoplastics
Multiple sclerosis/paralysis	Baclofen
Nausea	Antiemetics
Neurogenic bladder	Oxybutin
Osteoporosis/pagets	Etidronate/calcium regulators
Pain	Narcotics
Parkinsons/tremor	Benztropine, Trihexyphenidyl
PCP pneumonia	Pentamidine, Atovaquone
Psychotic illness/bipolar	Antipsychotics, lithium
Replacement solution	Potassium chloride
Seizure disorders	Anticonvulsants
Thyroid disorder	Thyroid hormones
Transplant	Immunosuppressive agents
Tuberculosis	Rifampin

Table 3.16

New Drug Classification		
Rx #	**Rx Category**	**Drug Description and (Multum Classes)**
1	Alzheimers	Tacrine, Donepezil, Rivastigmine, Galantamine (in 80)
2	Anticoagulants	Heparins (261), Warfarin (262)
3	Asthma	Antiasthmatic combinations (131), Adrenergic bronchodilators (180), Bronchodialators combinations (181), Methylxantihines (126), Leukotriene modifier (243), Respiratory Inhalants (130)
4	Autoimmune	Azathioprine (in 104 and 192)
5	Burns	Silver Sulfadiazine (in140)
6	Cardiovascular	ACE inhibitors (42), Beta blockers (274 & 275), Nitrates (45), Vasodilators (52 & 53), Calcium channel blockers (48), Digoxin (in 50)
7	Arrhythmias	Antiarrhythmic agents (46)
8	Cystic fibrosis	Pancrelipase (in 91)
9	Depression/anxiety	Antidepressants (76, 208 & 209), Antianxiety (in 69 - alprazolam, chlordiazepoxide, clonazepam, clorazepate, diazepam, halazepam, lorazepam & oxazepam) and (in 70 - buspirone, doxepin, ethchlorvynol, & meprobamate)
10	Diabetes	Insulin (215), Sulfonylureas (213), Alpha-glucosidase inhibitors(216), Thiazolidinediones (271), Metformin (214)
11	ESRD/renal	Epoietin Alfa (in 36), Calcitriol (in 119)
12	Gastric acid disorders	H2 Blockers (272), PPIs (94)
13	Gout	Colchicine (in 194), Allopurinol (in 194)
14	Hepatitis	Interferon beta (in 256), Peginterferon (in 256 & 177), Ribavirin (in 229)
15	Herpes	Acyclovir (in 229)
16	HIV/AIDS	Antiretrovirals (175, 176, & 227)
17	Hyperlipidemia	Antihyperlipidemics (173 & 174)

18	Infections	Quinolones (14), Cephalosporins (159 - 162), Penicillins (223, 224, & 226), Macrolides (11), Sulfonamides (15), Tetracyclines (16), Penicillinase resistant Penicillins (222), Beta-lactamase inhibitors (225), Urinary anti-infectives (17)
19	Insomnia	Sedatives, Hypnotics (in 69 estazolam, flurazepam, midazolam, quazepam, temazepam & triazolam) (in 70 acetylcarbromal, chloral hydrate, chlormezanone, dexmedetomidine, doxylamine, hydroxyzine, paraldehyde, propiomazine, pyrilamine, zaleplon, & zolpidem)
20	Iron deficiency	Iron (in 116) (B8"Carbonyl Iron", "Iron Dextran", "Iron Polysaccharide", & "Multivitamin with Iron", "Iron Sucrose")
21	Liver disease	Lactulose (in 95)
22	Malignancies	Antineoplastics (22 - 25)
23	Multiple sclerosis/paralysis	Baclofen (178)
24	Nausea	Antiemetics (195 - 198 minus diphenhydramine)
25	Neurogenic bladder	Oxybutynin (in 264)
26	Osteoporosis/pagets	Etidronate/calcium regulators (217)
27	Pain	Narcotics (60 & 191)
28	Parkinsons/tremor	Dopaminergic antiparkinson agents (276), Benztropine (in 205), Trihexyphenidyl (in 205),
29	Psychotic illness/bipolar	Antipsychotics, Lithium (77 & 79)
30	Seizure disorders	Anticonvulsants (199 - 204)
31	Thyroid disorder	Thyroid hormones (103)
32	Transplant	Immunosuppressive agents (104)
33	Tuberculosis AND PCP pneumonia	Rifampin (in 232), Isoniazid (in231)

Table 3.17

New Drug Classification - MarketScan (All variables utilized*)					
#	Rx Category	Drug Description and (Multum Classes)	Freq. Of All Rx Claims	Freq. Of Unique Patients	Prevalence of Rx Category
1	Alzheimers	Tacrine, Donepezil, Rivastigmine, Galantamine (in 80)	1,225	227	0.0173%
2	Anticoagulants	Heparins (261), Warfarin (262)	49,363	8,725	0.6667%
3	Asthma	Antiasthmatic combinations (131), Adrenergic bronchodilators (180), Bronchodialators combinations (181), Methylxantihines (126), Leukotriene modifier (243), Respiratory Inhalants (130)	379,118	83,806	6.4037%
4	Autoimmune	Azathioprine (in 104 and 192)	7,816	1,229	0.0939%
5	Burns	Silver Sulfadiazine (in140)	3,847	3,092	0.2363%
6	Cardiovascular	ACE inhibitors (42), Beta blockers (274 & 275), Nitrates (45), Vasodilators (52 & 53), Calcium channel blockers (48), Digoxin (in 50)	1,268,717	159,566	12.1927%
7	Arrhythmias	Antiarrhythmic agents (46)	13,420	2,322	0.1774%
8	Cystic fibrosis	Pancrelipase (in 91)	3,193	775	0.0592%
9	Depression/anxiety	Antidepressants (76, 208 & 209), Antianxiety (in 69 - alprazolam, chlordiazepoxide, clonazepam, clorazepate, diazepam, halazepam, lorazepam & oxazepam) and (in 70 - buspirone, doxepin, ethchlorvynol, & meprobamate)	1,194,499	187,234	14.3068%

10	Diabetes	Insulin (215), Sulfonylureas (213), Alpha-glucosidase inhibitors(216), Thiazolidinediones (271), Metformin (214)	438,562	42,430	3.2421%
11	ESRD/renal	Epoietin Alfa (in 36), Calcitriol (in 119)	3,970	865	0.0661%
12	Gastric acid disorders	H2 Blockers (272), PPIs (94)	472,163	108,649	8.3020%
13	Gout	Colchicine (in 194), Allopurinol (in 194)	50,392	9,750	0.7450%
14	Hepatitis	Interferon beta (in 256), Peginterferon (in 256 & 177), Ribavirin (in 229)	6,281	813	0.0621%
15	Herpes	Acyclovir (in 229)	33,162	13,927	1.0642%
16	HIV/AIDS	Antiretrovirals (175, 176, & 227)	18,009	907	0.0693%
17	Hyperlipidemia	Antihyperlipidemics (173 & 174)	483,321	81,748	6.2465%
18	Infections	Quinolones (14), Cephalosporins (159 - 162), Penicillins (223, 224, & 226), Macrolides (11), Sulfonamides (15), Tetracyclines (16), Penicillinase resistant Penicillins (222), Beta-lactamase inhibitors (225), Urinary anti-infectives (17)	1,241,568	552,558	42.2217%
19	Insomnia	Sedatives, Hypnotics (in 69 estazolam, flurazepam, midazolam, quazepam, temazepam & triazolam) (in 70 acetylcarbromal, chloral hydrate, chlormezanone, dexmedetomidine, doxylamine, hydroxyzine, paraldehyde, propiomazine, pyrilamine, zaleplon, & zolpidem)	122,002	42,911	3.2789%
20	Iron deficiency	Iron (in 116) (B8"Carbonyl Iron", "Iron Dextran", "Iron Polysaccharide", & "Multivitamin with Iron", "Iron Sucrose")	10,753	4,388	0.3353%
21	Liver disease	Lactulose (in 95)	4,358	1,551	0.1185%
22	Malignancies	Antineoplastics (22 - 25)	189,958	44,749	3.4193%

23	Multiple sclerosis/paralysis	Baclofen (178)	8,368	1,802	0.1377%
24	Nausea	Antiemetics (195 - 198 minus diphenhydramine)	102,964	59,156	4.5202%
25	Neurogenic bladder	Oxybutynin (in 264)	10,815	3,426	0.2618%
26	Osteoporosis/pagets	Etidronate/calcium regulators (217)	36,124	7,781	0.5946%
27	Pain	Narcotics (60 & 191)	610,657	238,067	18.1910%
28	Parkinsons/tremor	Dopaminergic antiparkinson agents (276), Benztropine (in 205), Trihexyphenidyl (in 205),	26,975	7,581	0.5793%
29	Psychotic illness/bipolar	Antipsychotics, Lithium (77 & 79)	44,856	6,921	0.5288%
30	Seizure disorders	Anticonvulsants (199 - 204)	137,661	20,848	1.5930%
31	Thyroid disorder	Thyroid hormones (103)	368,611	57,556	4.3979%
32	Transplant	Immunosuppressive agents (104)	12,327	1,029	0.0786%
33	Tuberculosis AND PCP pneumonia	Rifampin (in 232), Isoniazid (in231)	1,766	706	0.0539%
		Total	7,356,821	1,757,095	

*Some variables are added to the Diagnostic Model and some variables are combined with Diagnostic Model categories

Table 3.18

New Drug Classification - MediCal (All variables utilized*)					
#	Category	Drug Description and (Multum Classes)	Freq. Of all Rx Claims	Freq. of unique Patients	Prevalence of Rx Category
1	Anticoagulants	Heparins (261), Warfarin (262)	3,929	711	0.7676%
2	Asthma	Antiasthmatic combinations (131), Adrenergic bronchodilators (180), Bronchodialators combinations (181), Methylxantihines (126), Leukotriene modifier (243), Respiratory Inhalants (130)	55,164	8,918	9.6285%
3	Autoimmune	Azathioprine (in 104 and 192)	517	91	0.0982%
4	Burns	Silver Sulfadiazine (in140)	530	341	0.3682%
5	Cardiovascular	ACE inhibitors (42), Beta blockers (274 & 275), Nitrates (45), Vasodilators (52 & 53), Calcium channel blockers (48), Digoxin (in 50)	66,431	9,852	10.6369%
6	Arrhythmias	Antiarrhythmic agents (46)	777	150	0.1620%
7	Cystic fibrosis	Pancrelipase (in 91)	1,314	448	0.4837%
8	Depression/anxiety	Antidepressants (76, 208 & 209), Antianxiety (in 69 - alprazolam, chlordiazepoxide, clonazepam, clorazepate, diazepam, halazepam, lorazepam & oxazepam) and (in 70 - buspirone, doxepin, ethchlorvynol, & meprobamate)	104,715	14,277	15.4144%
9	Diabetes	Insulin (215), Sulfonylureas (213), Alpha-glucosidase inhibitors (216), Thiazolidinediones (271), Metformin (214)	41,354	4,464	4.8196%

10	ESRD/renal	Epoietin Alfa (in 36), Calcitriol (in 119)	783	177	0.1911%
11	Gastric acid disorders	H2 Blockers (272), PPIs (94)	31,695	8,949	9.6620%
12	Gout	Colchicine (in 194), Allopurinol (in 194)	1,704	419	0.4524%
13	Hepatitis	Interferon beta (in 256), Peginterferon (in 256 & 177), Ribavirin (in 229)	204	22	0.0238%
14	Herpes	Acyclovir (in 229)	2,312	949	1.0246%
15	HIV/AIDS	Antiretrovirals (175, 176, & 227)	13,554	654	0.7061%
16	Hyperlipidemia	Antihyperlipidemics (173 & 174)	16,937	3,869	4.1772%
17	Infections	Quinolones (14), Cephalosporins (159 - 162), Penicillins (223, 224, & 226), Macrolides (11), Sulfonamides (15), Penicillinase resistant Penicillins (222), Tetracyclines (16), Beta-lactamase inhibitors (225), Urinary anti-infectives (17)	90,352	33,561	36.2348%
18	Insomnia	Sedatives, Hypnotics (in 69 estazolam, flurazepam, midazolam, quazepam, temazepam & triazolam) (in 70 acetylcarbromal, chloral hydrate, chlormezanone, dexmedetomidine, doxylamine, hydroxyzine, paraldehyde, propiomazine, pyrilamine, zaleplon, & zolpidem)	22,847	5,968	6.4435%
19	Liver disease	Lactulose (in 95)	278	80	0.0864%
20	Malignancies	Antineoplastics (22 - 25)	7,585	2,380	2.5696%
21	Multiple sclerosis/paralysis	Baclofen (178)	1,515	324	0.3498%
22	Nausea	Antiemetics (195 - 198 minus diphenhydramine)	18,963	6,796	7.3374%
23	Neurogenic bladder	Oxybutynin (in 264)	1,311	365	0.3941%

24	Osteoporosis/pagets	Etidronate/calcium regulators (217)	1,030	297	0.3207%
25	Pain	Narcotics (60 & 191)	63,571	18,239	19.6921%
26	Parkinsons/tremor	Dopaminergic antiparkinson agents (276), Benztropine (in 205), Trihexyphenidyl (in 205),	20,753	3,224	3.4809%
27	Psychotic illness/bipolar	Antipsychotics, Lithium (77 & 79)	60,158	5,961	6.4359%
28	Seizure disorders	Anticonvulsants (199 - 204)	41,397	4,674	5.0464%
29	Thyroid disorder	Thyroid hormones (103)	9,245	2,270	2.4508%
30	Transplant	Immunosuppressive agents (104)	1,403	109	0.1177%
31	Tuberculosis AND PCP pneumonia	Rifampin (in 232), Isoniazid (in231)	680	240	0.2591%
		Total	683,008	138,779	

*Some variables are added to the Diagnostic Model and some variables are combined with Diagnostic Model categories

† The categories of Alzheimer's disease and iron deficiency were dropped due to formulary restrictions in the MediCal system

Table 3.19

Phi Coefficients for Correlated MRxCost Model Variables

Variable1	Variable 2	Phi Coefficient
CCS 23.3 (asthma)	RX 3 (asthma)	0.4317
CCS 4.1 (diabetes w/o complications)	RX 10 (diabetes)	0.6650
CCS 4.2 (diabetes w/ complications)	RX 10 (diabetes)	0.4916
CCS 5.3 (gout and other crystal athropathies)	RX 13 (gout)	0.3553
CCS 5.4 (Cystic fibrosis)	RX 7 (Cystic fibrosis)	0.2108
CCS 9.1 (affective disorders)	RX 9 (depression/anxiety)	0.2869
CCS 9.3 (anxiety, somatoform, dissociative and pesronality disorders)	RX 9 (depression/anxiety)	0.2285
CCS 9.4 (other mental conditions)	RX 9 (depression/anxiety)	0.2140
CCS 14 (hypertension)	RX 6 (cardiovascular)	0.4959
CCS17 (coronary atherosclerosis and other heart disease)	RX 6 (cardiovascular)	0.2472
CCS 13.2.2 (multiple sclerosis)	RX 15 (hepatitis)	0.5137
CCS 24.2.1 (Esophageoal disorders)	RX 12 (gastic acid disorders)	0.3907
CCS24.2.2 (gastroduodenal ulcer AND gastritis and duodenitis)	RX 12 (gastic acid disorders)	0.2557
CCS 2 (HIV/AIDS)	RX 17 (HIV/AIDS)	0.6896
CCS 5.2 (disorders of lipid metabolism)	RX 18 (hyperlipidemia)	0.4535
CCS17 (coronary atherosclerosis and other heart disease)	RX 18 (hyperlipidemia)	0.2794
CCS 23.1.2 (influenza, tonsilitis, acute bronchitis and other URIs)	RX 19 (infections)	0.4118
CCS 13.2.2 (multiple sclerosis)	RX 24 (multiple sclerosis/paralysis)	0.2068
CCS 9.2 (schizophrenia and related disorders and other psychosis)	RX 31 (psychotic illness/bipolar)	0.2018
CCS 5.1 (thyroid disorders)	RX 33 (thyroid disorder)	0.5868
CCS 13.4 (epilepsy, convulsions)	RX 32 (seizures disorders)	0.3769
CCS 27.2.2 (menstrual disorders AND menopausal disorders)	RX 23 (malignancies)	0.2010
CCS 30.4 (osteoporosis AND pathological fracture)	RX 27 (osteoporosis)	0.3063

* Results based on MarketScan Training Sample

Table 3.20

New Drug Classification - Combined with CCS Categories		
Rx Category	**Drug Description and (Multum Classes)**	**Combined with CCS Category***
Asthma (Rx 2)	Antiasthmatic combinations (131), Adrenergic bronchodilators (180), Bronchodialators combinations (181), Methylxantihines (126), Leukotriene modifier (243), Respiratory Inhalants (130)	CCS 23.3 - Asthma
Cardiovascular (Rx5)	ACE inhibitors (42), Beta blockers (274 & 275), Nitrates (45), Vasodilators (52 & 53), Calcium channel blockers (48), Digoxin (in 50)	CCS 14 - Hypertension CCS 17 - Coronary atherosclerosis and other heart disease **Logic: 1.** Multum class 42 (ACE inhibitors), 48 (Ca channel blockers), 52 & 53 (Vasodilators), 274 & 275 (Beta blockers) combined with CCS 14. **2.** Multum class 45 (Nitrates) and Digoxin combined with CCS 17
Cystic fibrosis (Rx 7)	Pancrelipase (in 91)	CCS 5.4 Cystic fibrosis
Depression/anxiety (Rx 8)	Antidepressants (76, 208 & 209), Antianxiety (in 69 - alprazolam, chlordiazepoxide, clonazepam, clorazepate, diazepam, halazepam, lorazepam & oxazepam) and (in 70 - buspirone, doxepin, ethchlorvynol, & meprobamate)	CCS 9.1 - Affective disorders CCS 9.3 - Anxiety, somatoform, dissociative, and personality disorders CCS 9.4 - Other mental conditions **Logic:** if no diagnosis is present within CCS 9.1, 9.3, or 9.4 then combined with CCS 9.1

Diabetes (Rx9)	Insulin (215), Sulfonylureas (213), Alpha-glucosidase inhibitors(216), Thiazolidinediones (271), Metformin (214)	CCS 4.1 - Diabetes w/o complications CCS 4.2 - Diabetes w/ complications **Logic:** if no diagnosis is present within CCS 4.1 or 4.2 then combined with CCS 4.1
Gastric acid disorders (Rx11)	H2 Blockers (272), PPIs (94)	CCS 24.2.1 - Esophageal disorder CCS 24.2.2 - Gastroduodenal ulcer and gastritis and duodenitis **Logic:** if no diagnosis is present within CCS 24.2.1 or 24.2.2 then combined with CCS 24.2.2
Gout (Rx12)	Colchicine (in 194), Allopurinol (in 194)	CCS 5.3 - Gout and other crystal arthropathies
Hepatitis (Rx13)	Interferon beta (in 256), Peginterferon (in 256 & 177), Ribavirin (in229)	CCS 1.3 - Hepatitis
HIV/AIDS (Rx15)	Antiretrovirals (175, 176, & 227)	CCS 2 - HIV Infection
Hyperlipidemia (Rx 16)	Antihyperlipidemics (173 & 174)	CCS 5.2 - Disorders of lipid metabolism
Infections (Rx17)	Quinolones (14), Cephalosporins (159 - 162), Penicillins (223, 224, & 226), Macrolides (11), Sulfonamides (15), Tetracyclines (16), Penicillinase resistant Penicillins (222), Beta-lactamase inhibitors (225), Urinary anti-infectives (17)	CCS 1.1.1, 1.1.2, 1.1.3, 1.1.4, CCS 1.2, 1.3 CCS 2, CCS 12.1, 12.2 CCS 13.1, CCS 23.1.1, 23.1.2, CCS 24.1 CCS 26.3, CCS 29.1 - Infections **Logic:** If no diagnosis within CCS 1.1.1, 1.1.2, 1.1.3, 1.2, 1.3, 2, 12.1, 12.2, 13.1, 23.1.1, 23.1.2, 24.1, 26.3, or 29.1 then combined with CCS 1.1.4 - other bacterial infections

Malignancies (Rx 22)	Antineoplastics (22 - 25)	CCS 3 Neoplasms
Multiple sclerosis/paralysis (Rx21)	Baclofen (178)	CCS 13.2.2 - Multiple sclerosis
Osteoporosis/pagets (Rx 24)	Etidronate/calcium regulators (217)	CCS 30.4 - Osteoporosis and pathological fracture
Psychotic illness/bipolar (Rx 27)	Antipsychotics, Lithium (77 & 79)	CCS 9.2 - Schizophrenia and related disorders and other psychoses
Seizure disorders (Rx 28)	Anticonvulsants (199 - 204)	CCS 13.4 - Epilepsy, convulsions
Thyroid disorder (Rx29)	Thyroid hormones (103)	CCS 5.1 - Thyroid disorders

* Model variables are binomial 1/0 (1 if present). If diagnosis is present within combined CCS groups (variable = 1) then the presence of a Rx group will not affect the model.

CHAPTER 4

DEVELOPMENT AND VALIDATION OF A CLAIMS-BASED RISK ASSESSMENT MODEL TO PREDICT PHARMACY EXPENDITURES IN A COMMERCIAL POPULATION[1]

[1]CR Cantrell, BC Martin. To be submitted to *The American Journal of Managed Care*

ABSTRACT

Objectives: Empirically develop and validate the RxCost Model a prospective risk assessment model that uses claims-based diagnostic information to predict future pharmacy expenditures for a U.S. commercial population. Additionally, another model, the Mixed RxCost (MRxCost) Model, was empirically developed and validated as well to explore the gain in predictive power associated with adding drug information to the RxCost Model. The performance of these models were compared to each other as well as a Demographic-only model and the DCG-HCC model.

Study Design: Retrospective longitudinal cohort study

Data Sources: Three years, 1998 through 2000, of MEDSTAT MarketScan U.S. commercial claims data.

Subjects: All persons enrolled in the health plans collected in MarketScan who were continuously enrolled for at least 13 months, who were 18 to 64 years of age, not eligible for Medicare and were not admitted for a hospital or nursing home stay > 30 days.

Methods: A training sample consisting of over 1.3 million lives was utilized to develop the models. Initially, both OLS and a two-part model were evaluated but OLS was chosen to estimate the model coefficients. A random holdout sample of 218,383 was utilized to validate the models and to compare the performance of each model. The discrimination of the model was compared to the commercially available Diagnostic Cost Group Hierarchical Condition Category (DCG-HCC) model.

Results: The R-square value for the RxCost Model, the MRxCost Model and the DCG-HCC using the validation sample was 0.22, 0.34 and 0.16 respectively.

Conclusions: The RxCost Model was successfully developed and it outperformed the DCG-HCC model in terms of R-square even after re-calibrating the DCG-HCC model. The MRxCost Model also proved that supplementing drug information can improve discriminatory power although discretion should be utilized.

Key Words: Risk assessment, risk adjustment, prescription cost, risk models, drug cost

INTRODUCTION

As health plans strive to control their medical costs while preserving the quality of care provided to their patients, they will increasingly focus on the most rapidly growing component of their cost structure – drugs. (Litton 2000) Although outpatient prescription drug spending represented only a small portion (11% or $140.6 billion) of personal health spending in 2001, it was one of the fastest growing components (The Kaiser Family Foundation 2003). Spending on prescription drugs in the U.S. rose at double-digit rates throughout the past decade. (Employee Benefit Research Institute 1999) From 1997 to 2001, three primary factors are attributed to the growth in prescription drug expenditure; increased utilization (47%), use of newer, more expensive medications (27%) and prices increases (26%). (The Kaiser Family Foundation 2003)

As a result of the rising cost of drugs and the increased utilization, especially of newer, more expensive drugs, managed care organizations (MCOs) and pharmacy benefit managers (PBMs) have explored various strategies for managing pharmacy benefits. Alternatives for controlling drug costs and utilization have mainly consisted of: restricting patient access, shifting costs to enrollees, sharing cost increases through higher premiums, and shifting risk to providers. (Kleinke 2001) Formularies, prior

authorizations, step therapy, limits, and higher premiums and copays have all been utilized to restrict patient access and shift costs to enrollees. (Galt 2001; Motheral 1999)

Although the techniques described above remain an important component in managing prescription expenditures, another more innovative way to control drug costs and utilization has been to shift the risk or otherwise incentivize providers. There is a growing trend toward paying more for higher quality. Two such movements of late are the Pay for Performance (P4P) initiative in California and the Bridges to Excellence (BTE) program for diabetes care (National Committee for Quality Assurance 2003, Bridges to Excellence 2004). Both programs provide physicians with financial incentives (bonuses) to provide high quality care. This approach may well be used in the future to promote judicious prescribing behaviors since it is the physician that has the best knowledge of the patient and acts as the patient's agent to select prescription drugs. Creating structures that financially affect the patient through copays, formularies, and other patient incentives only indirectly influences physician prescribing decisions since the physicians themselves are not directly rewarded for more judicious prescribing. Physician profiling and physician capitation have both been used in the past as a means of shifting risk to providers. (Burton 2001; Carroll 2000; United 2003) Recently, physician profiling was used to introduce financial incentives for physicians to prescribe effectively and efficiently. (AFSCME 2003; United 2003; Graden 1998) Bonus structures reward physicians who prescribe judiciously and hold accountable those that do not. In addition, capitation for prescription drug costs passes financial incentives to control drug spending on to physicians, who are paid a set allowance for each patient that must cover the patient's yearly prescription drug costs, with any excess coming out

of the physician's own pocket. (Burton 2001) These new techniques have not been used extensively but may become more commonplace in the future as health plans increasingly focus on prescription expenditures. The advantage of using these methods to shift costs to providers is that it gives physicians an incentive to resist excessive consumer demand for expensive or overprescribed drugs. (Burton 2001) Disadvantages can include an increased emphasis on cost rather than medical necessity in drug choice and the creation of adverse selection incentives where more healthy patients are chosen over less healthy patients by physicians. To combat adverse selection, these profiling techniques should incorporate risk assessment in order to adjust for differences in patient case-mix. Risk assessment techniques attempt to control for clinical differences in patient risk and to isolate quality differences. Risk adjustment techniques can then be utilized to allow physicians who treat more a severe patient-mix higher prescription drug expenditures. When shifting risk to providers, the goal of risk assessment/adjustment is to develop a system that doesn't penalize physicians who treat high-risk patients. Without risk assessment/adjustment physicians will argue that their patients are sicker than plan averages.

To date there are no published literature on risk assessment models specifically designed to predict prescription expenditures. The objective of this research is to empirically develop a diagnostic-based risk assessment model to predict prescription expenditures in a commercial population. Additionally, a model using both diagnostic and drug information will be developed to explore the advantages of supplementing a diagnostic model with drug information. The models could potentially be used to set prescribing goals for physicians or physician groups based on their patient population.

These goals could then be used to build in financial incentives or bonuses for those physicians that achieve their goals.

METHODS

A retrospective longitudinal review was employed to empirically develop risk assessment models to predict pharmacy expenditures for a commercial managed care population. Three new models, a demographic-only model, a diagnostic model (RxCost Model) and a model incorporating both diagnostic and drug information (MRxCost Model), were developed separately utilizing claims data. The risk assessment models were developed in a prospective fashion and utilized information from one plan year to predict pharmacy expenditures for the following year. The demographic model was used to establish a baseline to observe the predictive power of a model that can be derived with minimal effort. The RxCost Model was the primary model for this research. The MRxCost Model was used to explore how much predictive power would be obtained by adding drug information. A 13.4% random sample of the population was utilized to validate both the RxCost and MRxCost Models (two-thirds of a 20% sample; one-third was set aside as a spare dataset).

Data

Three years, 1998 through 2000, of MEDSTAT MarketScan data was utilized to develop and validate the risk assessment indices. (MEDSTAT Inc. 2001) This administrative database consists of enrollment information, prescription claims, inpatient and outpatient hospital claims, and ambulatory claims in a recipient level linkable format. This database provides the level of detail on costs and utilization needed to develop and validate risk assessment models. The data has been found to be valid in

previous epidemiological and health policy studies. (Gandhi 2001, Ozminkowski 2000, Zhao 1999) The total unduplicated count for the dataset utilized was 9,043,605 lives. All methods employed to mine the data and generate the model samples were performed using SAS Version 8.2 (SAS Inc. 2004).

Subjects

Persons meeting the following criteria were included in the sample to develop and validate the risk assessment indices.

- Continuously eligible for a minimum of thirteen months from January 1998 through December 2000
- Have prescription drug coverage
- Age 18 to 62 years at the beginning of the observational period
- Not admitted to institutions or nursing home facilities and who do not have periods of inpatient care in excess of 30 consecutive days at any time during the entire period
- Not dually eligible for Medicare

The member count for the MarketScan sample dropped from approximately 9 million to 1,634,427 after excluding members who did not meet the above criteria Table 4.1. Most members, just over 6 million, were lost due to the continuous eligibility requirement. From the members who met the criteria for inclusion, a random 80% sample was selected, referred to as the training sample, and used to develop the risk assessment models. Of the remaining 20%, two-thirds of the sample was randomly chosen to validate the models. The remaining one-third was set aside as spare data in case problems were encountered and the models need to be redeveloped / re-calibrated and

validated. The percentages for each sample were chosen to maximize the number of members available for development of the models while maintaining an adequate number in which to validate the models. This study was reviewed and approved by the University of Georgia Institutional Review Board; IRB # H2002-10473-2.

Risk Assessment

For each model the following mix of age-sex dummy variables was utilized where 1 indicates presence and 0 indicates absence:

- (Male 23-30 years), (Male 31-40 years) (Male 41-50 years), (Male 51-60 years), (Male 61-64 years)
- (Female 18-22 years), (Female 23-30 years), (Female 31-40 years), (Female 41-50 years), (Female 51-60 years), (Female 61-64 years)

The RxCost Model, the primary model for this research, is an ICD-9-CM code-based diagnostic model that prospectively predicts prescription expenditures. With more than 12,000 diagnostic codes making up the ICD-9-CM classification system, the first challenge is to aggregate these diagnostic codes into clinically meaningful categories that reflect similar prescription expenditures. Some of the techniques used by Ash and Kronick in developing the Diagnostic Cost Groups-Hierarchical Coexisting Conditions (DCG-HCC) (Ash 2000) and Chronic Illness and Disability Payment System (CDPS) (Kronick 2000) were used to guide the development of this model. The initial classification system to organize diagnostic codes was based on AHRQ's Clinical Classification Software (CCS). (Clinical Classification Software 2001) The CCS is a multi-level classification scheme that aggregates individual ICD-9-CM codes into clinically meaningful categories that group similar conditions. (Agency for Health Care

Policy and Research) The CCS Software is available for download at www.ahrq.org. The multi-level scheme used by CCS aggregates ICD-9-CM codes into 17 broad categories (e.g., Infectious Diseases, Neoplasms, and Mental Disorders) excluding the residual E codes. Within each of the 17 broad categories of the CCS, a multilevel categorization scheme divides each of the broad categories into more specific refined categories. The determining factor in creating these categories was the extent to which conditions could be grouped into relatively homogeneous clusters of interest to public policy researchers. These CCS categories were not created specifically for prescription expenditures and, therefore, do not necessarily reflect conditions with similar prescription expenditures.

Based on the prior use of diagnostic models to predict total health care costs, (Ash 2000, Kronick 2000) and a clinical panel review (consisting of two practicing clinical pharmacists) 34 major categories with 90 subcategories were proposed for aggregating diagnostic codes. The low-cost and ill-defined conditions were excluded and numerous CCS subcategories were collapsed into broader categories. For example, all subcategories under Neoplasms were collapsed into 1 major category “Neoplasms”. The final CCS-based diagnostic classification is presented in Table 4.2. All categories that are missing in the final models were either discarded or collapsed into the broader categories. Table 4.3 presents the 90 subcategories (diagnostic variables) that were used to estimate the RxCost Model.

The RxCost Model was empirically derived using dummy variables (1 / 0) to indicate presence or absence of a diagnostic category. Both inpatient and outpatient claims (ICD-9CM codes) were utilized for diagnostic information. Two separate

methods were initially utilized to capture diagnostic information within the RxCost models. One method, the full method, identifies and enters into the model all diagnostic categories for which there was a corresponding ICD-9-CM code present regardless if there were other similar diagnostic categories coded for each individual. The other method utilizes a hierarchical approach for certain variables where some categories are grouped together in clusters and only the most costly category within each cluster is entered into the model. In order to use the hierarchical method, relevant categories must first be grouped into clusters. The MarketScan training sample was utilized to identify variables with a Pearson Correlation ≥ 0.2. After these variables were identified a 2x2 contingency table for each of the correlated variable pairs was output and the phi coefficient was estimated. Based on the contingency results, 16 variables were considered appropriate for a total 6 hierarchical clusters. This method was utilized because it identifies variables that are often coded for the same person. These 6 clusters each contain variables that are often coded for the same condition. When a recipient has more than one variable coded within one of these hierarchical clusters, only the most costly variable should be captured by the model and not multiple variables for the same condition to avoid unnecessary code proliferation. For example, if a patient is coded for diabetes mellitus without complications and diabetes mellitus with other complications only the most costly of these variables, diabetes with complications, should be captured by the model. To determine the most costly variable for each cluster a random 20% sample of the MarketScan training data was utilized to calculate the average annual prescription cost associated with each of the 16 variables. Based on these results the variables within each cluster were assigned a hierarchical

classification. Utilizing the hierarchical approach, only the single-highest category within each hierarchical cluster will be captured by the model.

This hierarchical approach was initially adopted to prevent code proliferation and "gaming" of the risk assessment system where clinicians may have incentive to code more conditions which may be justified but do not substantially add to the prescription costs of those persons. This hierarchical method of counting has been shown to simplify the model, strengthen its resistance to additional coding, and produce only small decreases in the accuracy. (Kronick 2000)

Because only 16 variables were affected and only 6 clusters were formed using the hierarchical counting scheme, these 6 hierarchical clusters were incorporated into the full method. The 6 clusters chosen are potential areas for gaming an incentive-based system. Additionally, the minor changes should result in very little loss of predictive power. However, to ensure this was the case, the non-hierarchical full method was later run in addition to the method incorporating the hierarchical clusters for comparison. In both instances the R-square values remained the same.

Diagnosis-based models have some shortcomings that are evident regardless of counting method. These models rely on administrative claims-based data that are not always complete. Diagnosis codes often suffer from left censoring where an individual may be treated for a chronic condition but only diagnosed once for the condition or diagnosed sporadically over long periods of time. Also, diagnostic coding is often an uncertain practice where many ill-defined conditions may be coded in numerous ways. Adding drug information to a diagnosis-based model can attempt to alleviate some of

the problems with these diagnosis-based models. One advantage to utilizing drug information is that the data is much more complete than diagnostic information. Secondly, drug information can help identify individuals who are being treated for a chronic condition but who are not diagnosed at each physician-patient encounter. Another advantage is that drug information can potentially help explain severity. The number of classes of medication prescribed to treat a condition can sometimes be used to indicate severity. The disadvantages to utilizing drug information to predict drug expenditures is that a certain degree of endogeneity exists because prescription fills lead directly to greater prescription cost. However, endogeneity is less of a concern using models prospectively because drug costs in one period are only indirectly related to drug costs in a previous period. (Clark 1995) Another disadvantage of using drug information occurs when the risk assessment models are used for shifting some of the risk to providers (i.e. financial incentives). Here, there will be some incentive for providers to prescribe medications more liberally, particularly for medications that can potentially be prescribed for conditions with varying levels of severity or even prevention.

An alternative risk assessment model, the MRxCost Model, was developed and is based on both diagnostic information and drug information. This “Mixed Model” is utilized to ascertain how much predictive ability can be gained by adding drug information to the RxCost Model explained above. The RxCost Model was supplemented with drug information based on Gilmer's updated version of Clark's revised Chronic Disease Score to develop the MRxCost Model. (Gilmer 2001; Clark 1995)

Virtually all of the published literature utilizing prescription drugs to predict costs has been based on the Von Korff's Chronic Disease Score. (Gilmer 2001; Fishman 1999; Fishman 2003; Lamers 1999; Clark 1995; Johnson 1994; Von Korff 1992) Gilmer's Medicaid Rx model provides one of the latest approaches to updating the CDS and utilizing prescription drugs to model overall healthcare expenditures. (Gilmer 2001) The Medicaid Rx model categories were utilized as a starting point; however, they were not created specifically for prescription expenditures and, therefore, do not necessarily reflect conditions with similar prescription expenditures.

The Multum Classification system (Multum 2001), publicly available at www.multum.com, was utilized both to refine Gilmer's Medicaid Rx categories and to classify drug claims. Multum updates their database monthly and as of February 2002 there were 284 therapeutic classes. However, the Multum class numbers are not entirely sequential due to former classes being dropped or split out into multiple classes.

The Clinical panel reviewed Gilmer's Medicaid Rx model categories and revised the medication classes to reflect a proxy of severity and comorbidity that would consistently reflect an individual's expected outpatient prescription expenditure. Next, the prevalence of the drug categories in each population were checked to ensure that an adequate number of cases exist in each. When the prevalence was low in any category, it was either dropped or merged with another category. The new drug classification system deleted 6 original categories, combined Tuberculosis with PCP pneumonia and added numerous drugs to the drug list of some categories (Table 4.4).

The MRxCost Model was empirically derived using dummy variables (1 / 0) to indicate presence or absence of a diagnostic category or drug category. Here, the

presence of a drug in each category will be counted rather than number of drugs in each category. However, three drug categories were identified by the clinical panel as categories that could potentially help explain severity based on the number of drug classes prescribed. These categories are asthma, cardiovascular, and seizure disorders. Here, the dummy variables for each class were not included in the model but were summed in a single additive variable. Both asthma and seizure disorders categories have 6 classes of medication that could potentially be prescribed. A variable was created for each and was coded as 0 through 6 depending upon how many drug classes are prescribed. The Cardiovascular category has 8 classes that could potentially be prescribed. Here a variable was created in which a value of 0 to 8 was assigned. For example, if a patient filled medication that is classified into Multum categories 131, 180 and 243 then he/she would be assigned a 3 for their asthma additive variable.

The new drug classification system, that was used to complement the RxCost Model, contains some categories that are very similar or identical to those in the RxCost Model. Although similar (and collinear), these variables will be helpful to identify patients with conditions who did not receive a diagnosis within the timeframe the model is run but filled a prescription for the condition. These variables were combined so that either a diagnosis or a drug would indicate the presence of a condition. Some of these variables, such as asthma, are straightforward. Here both the diagnostic portion and the drug portion of the MRxCost Model contain a category for asthma. However, other categories are more complicated. The diagnostic classification system uses two categories for diabetes while the drug classification system uses one. Other categories

are concise in one classification and broad in the other. Drug categories that were empirically correlated were inspected and combined in the manner previously described. Seventeen drug categories were chosen and combined with their diagnostic counterparts. These categories and the logic used are presented in Table 4.5. The final diagnostic and Rx drug classification system used 109 variables to estimate the MRxCost Model. (Table 4.6) Because dummy variables are used in both the diagnostic portion and the drug portion of the MRxCost Model, the presence of a drug in a drug category that was combined with a diagnostic category will only add to the model when there is no presence of a diagnosis. For example, if a patient is diagnosed with asthma the presence of a drug in the asthma category (patient filled an asthma drug) will not affect the model because these categories were combined and the diagnosis is already present.

Analysis

A minimum of 13 months of data and up to 24 months was analyzed for each enrollee where the first twelve months served as the index year and was used to collect ICD-9CM and drug class information. Information gathered from the first 12 months was used to predict pharmacy expenditures for months 13 through 24. Where possible, the most recent twenty-four months of continuously eligible data was used for each enrollee. When twenty-four months of continuously eligible data was not available, the longest span of continuously eligible data was utilized. Here, the first 12 months of data was used as the index year and the remaining months were used for the cost year. Prescription costs for individuals with partial second-year data was annualized using the method described in Ash et al. (Ash 1989)

Total Rx Expenditure = Rx Cost × (12 ÷ Months Eligible)

Both inpatient and outpatient claims (ICD-9CM codes) were utilized for diagnostic information and pharmacy claims (NDC codes) were utilized for drug information.

The models were initially estimated using Ordinary Least Squares regression (OLS) and a Two-Part Model (2PM). OLS is the most common method for analyzing health utilization data and have been shown to work just as well as other more complicated models for predicting future costs (Diehr 1999). However, health care utilization data typically consists of a high proportion of persons with zero expenditure and a small number of individuals with very high expenditures. Because of this, a 2PM based on Duan et al. (Duan 1983) was used as an alternative to OLS. A two-part regression model was chosen because this type of model attempts to correct the problems associated with non-spenders by first using a logit equation for the dichotomous event of having zero or positive pharmacy expenses. The next part of the model is a regression equation that is conditional on having positive pharmacy expenses and is used to model the level of positive expenses. A log-linear regression model was utilized for the second part of the model. A "smearing" estimator was used to retransform the predicted values back to the mean of the original distribution. (Duan 1983)

Both models utilized a stepwise selection process where a variable was allowed to enter at a significance level of 0.20 and remain in the model at 0.10 was utilized. (Mantel 1970) The age-sex variables were considered essential variables and were included in all models regardless of their significance to minimize model mispecification errors and to ensure every member will have a predicted cost for the following year.

The hierarchical variables in the RxCost Model and MRxCost Model were grouped into clusters for the stepwise procedure so that they either all remained in the model or all fell out of the model.

Variables included in the models are as follows:

The *Demographic Model* was estimated with:

» Intercept

» dummy variables for age-sex classification (1 = year; 0 = absence)

The *RxCost Model was* estimated with:

» same as demographic model

» dummy variables for the presence or absence of a given condition in the revised diagnostic classification system
(1 = diagnosis present in a clinical classification; 0 = absence)

The *MRxCost Model* (both diagnostic and drug information) was estimated with:

» same as RxCost Model

» dummy variable for the presence of absence of a drug from a given therapeutic class in the revised drug classification system
(1 = drug present in therapeutic class; 0 = absence)

» additive variable for the 3 conditions identified as indicating severity based on the number of medication classes prescribed

Model Derivation

All model development was carried out using STATA Intercooled Version 6.0. (STATACORP 1999) The models estimated using OLS were compared to the models

estimated using the two-part modeling technique. Based on this comparison the OLS model was chosen to estimate all models.

Even though the 2PM seems to fit the distribution better, OLS performed slightly better when predicting future prescription costs. The Root Mean Squared Error (RMSE) of each technique was compared to assess model performance (RMSE RxCost Model: OLS = 1,312, 2PM = 1,465). This is consistent with other work using models to predict cost. (Diehr 1999) One problem encountered with OLS was that it estimated a negative intercept for both the RxCost Model and the MRxCost Model. With a negative intercept, the reference case (male 18 to 22) with no diagnostic information present would have a negative cost prediction for the following year which would be unacceptable. To circumvent this problem the intercept was constrained to zero and the age-sex variables were constrained to their mean prediction value when diagnostic information was absent. More specifically, members with no diagnostic information classified (all diagnostic variables were coded 0) were used to calculate the mean prediction value of each age-sex variable on following year's prescription cost. These values were then used to constrain the age-sex variables for everyone. This allowed OLS to predict only positive costs and resulted in no loss in predictive power in terms of R-square. OLS was chosen to estimate all models.

Two main tactics were used to guard against overfitting the models. First, the stepwise variable selection (discussed above) was employed to either keep or drop variables from the model based on their significance to the model. Secondly, the number of candidate variables was limited to a ratio of at least 10 cases to each predictor variable. (Harrell 1996)

Model Validation

All models were frozen and tested on the validation sample. Model discrimination, the amount of variance explained by each model, was evaluated by calculating the model's R-square value.

All models have been trained using all costs with no removal or trimming techniques utilized. The frozen model estimates were also applied to the validation sample after Year 2 prescription costs were trimmed at $20K. Here, any individual's prescription drug cost of over $20K was set equal to $20K. The value of $20K was chosen because it trims only the most severe outliers that could have a substantial effect on the models. The trimming procedure only affects 0.03% (67 observations) of the MarketScan validation sample.

Additionally, the RxCost Model was also run in a concurrent fashion to see how well the frozen coefficients predict year 1 prescription expenditures. The MRxCost Model was not run using the concurrent approach because endogeneity was deemed to be too great. Here, the model would draw upon prescription drug information to predict prescription drug expenditures in the same year.

Model Performance

The one universally reported, single-number summary performance measure for risk-assessment payment models is the R-square value; the proportion of variance in costs that the model explains. (Ash 2000) Model R-square values are reported for each of the models. Additionally, R-square values were also adjusted for shrinkage, the flattening of the plot of predicted versus observed away from the line, caused by

overfitting, using the heuristic shrinkage estimator of van Howelingen (van Houwelingen 1990)

$$R^2_{adj} = 1 - (1- R^2) \times (n-1) / (n-p-1)$$ where n = the number of subjects and p = the number of candidate variables

Both R^2 and R^2_{adj} values are reported for each of the training sample models and each of the validation sample models.

The R-square value described above assesses the amount of variance explained by the model for each individual member applied to the model. These models are designed to be utilized for a physician or physician group setting where a group of patients (e.g. 10, 50 or 100 patients) could be used to predict a mean prescribing expenditure across physician groups. When predicting a prescribing expenditure for a physician or physician group based on their patients, the amount of variance explained for individual patients is less important than the amount of variance explained for the groups of patients in each physician practice. Here a pooled R-square value will help assess the amount of variance the model explains for a group (or pool) of patients. Using these models in a physician or physician group setting would require the model to perform well for varying numbers of patients in each physician group. To assess how well the model performs for varying numbers of patients, pooled R-square values were calculated based on hypothetical physician groups of 10, 20, 50, 100, 200, 300, 400, and 500 patients randomly chosen from the validation sample. To calculate the pooled R-square value for groups of 10 patients the groups were randomly selected without replacement and each of the patients actual year 2 prescription costs are summed into one group value. Additionally, each of their predicted cost was summed into one value. Next, each group was then counted as 1 observation and an R-square value was

calculated for the entire sample based on the groups of 10 patients. This procedure was replicated for groups of 10, 20, 50, 100, 200, 300, 400 and 500 patients. The ultimate goal here was to assess how many patients are needed within a physician or physician group population before the model could be used to predict prescription costs effectively.

Another measure of model performance is a predictive ratio. (Ash 2000) Here, the model is applied to a subgroup of people and the predictive ratio is calculated by dividing the model-predicted costs for the group by their actual costs. Each model applied to the validation sample was used to predict costs of specific subgroups including, asthma, depression, diabetes, HIV infection and hypertension. These subgroups were identified by ICD-9-CM codes (and NDC codes in the MRxCost Model) during year 1. The predicted costs were then utilized to calculate a predictive ratio. The predictive ratio was used to evaluate how well the models perform for the chosen specific subgroups. An ideal predictive ratio of 1.0 would indicate that the predicted costs and actual costs were exactly the same. As with R-square values, pooled predictive ratios were also calculated using the hypothetical physician groups of patients. Here, the random groups identified above in the validation sample were used to calculate a predictive ratio for each group and an average predictive ratio was calculated for groups of 10, 20, 50, 100, 200, 300, 400 and 500 patients. Additionally, 95% confidence intervals were calculated based on the mean predicted ratio values and the standard deviation. Once again, this was utilized to assess how well the models predict a single prescription expenditure for varying groups of patients.

Although there are no other risk assessment indices publicly reported for predicting pharmacy expenditures there are publicly available approaches for predicting overall health care costs. One such index, Ash's Diagnostic Cost Group Hierarchical Condition Category (DCG-HCC), is one of the more sophisticated models available. This model was also applied prospectively to the validation sample to predict pharmacy expenditures. The R-square value for this model was then compared to the RxCost Model and the MRxCost Model. Comparing the models allowed further evaluation of model performance. The DCG-HCC model, and the software to implement the model, is commercially available from the DxGROUP online at www.dxcg.com. Because this model was not developed specifically for predicting prescription expenditures, it was recalibrated to estimate prescription costs. Without recalibration this model would substantially over predict costs. The benchmark weights were used for each model; however the models were recalibrated using a proportional calibration method explained in the DxCG Analytic Manual (DxCG Inc 2001). This method results in the mean of the individual predicted expense to be equal to the mean observed year 1 expense. Employing this method, the model generates costs that are in line with prescription costs instead of overall health costs.

RESULTS

Population Sample

The population description for both the MarketScan training and validation samples is presented in Table 4.7. The samples were randomly chosen and both are almost identical in terms of age, gender, eligibility, FFS and prescription cost. Just over

half the population of both samples is made up of females and the average age is approximately 42 years.

Models

The results of the RxCost Model including the coefficients, their p-values and the 95% confidence intervals are presented in Table 4.8. The R-square value was equal to 0.17 in the training sample and 87 of the possible, 101 variables remained in the model after the stepwise selection procedure. The reference group was male 18 to 22. Variable costs ranged from $0.25 for Male 23 to 30 up to $7,025.12 for CCS 2 (HIV Infection). Costs for the age-sex variables increased with age and females age 61 to 64 had the highest cost estimates. Seven diagnostic categories had cost estimates over one thousand dollars; septicemia, HIV infection, cystic fibrosis, immunity disorders, Parkinson's disease, multiple sclerosis and chronic renal failure. The average predicted cost for all members was 474.67 (95% CI: 472.21, 477.14). The average actual cost was 490.87 (95% CI: 485.66, 496.09). The predictive ratios, the ratio of predicted costs versus actual costs, of the RxCost Model for each of the subgroup populations were as follows: hypertension = 0.99 (95% CI: 0.98, 1.00), diabetes = 0.98 (95% CI: 0.96, 1.00), asthma = 0.97 (95% CI: 0.94, 1.00), HIV = 1.05 (95% CI: 0.90, 1.20), and depression = 0.93 (95% CI: 0.90, 0.96). The overall predicted ratio for all members = 0.97 (95% CI: 0.96, 0.98).

The MRxCost Model results from the training sample are presented in Table 4.9. The R-square value is 0.26 and 83 of the possible 120 variables remained in the model after the stepwise selection procedure. All of the 19 prescription variables remained in the model after selection. The reference group was male 18 to 22. Variable costs

ranged from $3.79 for Male 23 to 30 up to $6,574.96 for CCS 2 (HIV Infection). Once again, costs for the age-sex variables increased with age and females age 61 to 64 had the highest cost estimates. Five diagnostic categories and three prescription categories had cost estimates over one thousand dollars; HIV infection, cystic fibrosis, immunity disorders, Parkinson's disease, multiple sclerosis, ESRD/renal disorders and tuberculosis and PCP pneumonia. The average predicted cost for all members was 470.73 (95% CI: 467.94 and 473.52). The average actual cost was 490.87 (95% CI: 485.66 and 496.09). The predictive ratios of the MRxCost Model for each of the subgroup populations were as follows: depression = 1.01 (95% CI: 0.99, 1.03), diabetes = 1.03 (95% CI: 1.01, 105), HIV = 0.93 (95% CI: 0.81, 1.05), hypertension = 1.18 (95% CI: 1.16, 1.20) and asthma = 2.15 (95% CI: 2.00, 2.30). The overall predicted ratio for all members = 0.96 (95% CI: 0.95, 0.97).

The R-square values adjusted for shrinkage did not differ from the original R-square values in any of the models due to the ratio of variables to observations being extremely small. All of the model R-square values are presented in Table 4.10. The Demographics Model had an R-square = 0.03 in the training sample and 0.03 in the validation sample and shows what can be achieved with minimal effort.

The R-square value for the RxCost Model increased from 0.17, as estimated using the training sample, to 0.22 when applied to the validation sample. The R-square value was approximately 37% higher for the RxCost Model as compared to the DxCG model. When the weights for the RxCost Model were run concurrently the R-square value was over 50% higher as compared to the prospective model. The MRxCost

Model yielded a higher R-square value for both the training sample and the validation sample as compared to the RxCost Model.

The pooled R-square values and average predicted ratios for each model applied to various groups of patients are presented in Table 4.11. The R-square values are fairly high even with groups of 10 members (0.24) but increase as the group size increases.

Trimming the outliers in the validation sample proved to be especially useful. Here, only 67 (0.03%) observations were affected by the trimming technique; however the R-square value increased almost 20% for both prospective models. The highest R-square value for any model was achieved by the concurrent RxCost Model when applied to the trimmed validation sample with an r-square equal to 053

DISCUSSION

The objective of this study was to empirically develop a claims-based risk assessment model to predict prescription expenditures for a commercial population. Coupling the recent trend towards using incentives to drive physician behavior with the ever-growing attention to prescription drug expenditures, this model could potentially be used to set prescribing goals for physicians based on their patient-mix. Currently, there are no other published models developed specifically for predicting prescription expenditures. The primary model of interest is the RxCost Model which uses diagnostic claims-based information to predict prescription expenditures. Additionally, a MRxCost Model was developed that supplements the RxCost Model with prescription drug information.

The RxCost Model seemed to perform well when estimated on the training sample (in terms of R-square) as compared to other risk assessment models used to predict overall health costs. However, the model performed even better when applied to the validation sample and yielded an R-square value over 20% higher. The discriminatory power increased due to the validation sample having fewer extreme values than the training sample. The training and validation samples were chosen at random, however the validation sample was much smaller and slightly more uniform than the training sample. The training sample was almost 6 times larger than the validation sample and even after you account for the size difference the training sample had 52 more extreme values in terms of cost greater then $20K. The maximum cost in the training sample was $424,414 while the maximum in the validation sample was only $132,722. Additionally, the standard deviation for the training sample was $1424 vs. $1244 for the validation sample.

Compared to the DCG-HCC model the RxCost Model performed better even after the DCG-HCC model was re-calibrated to predict prescription costs. However, The DCG-HCC model was not developed specifically to predict prescription expenditures and even after re-calibration the RxCost Model is expected to have higher discriminatory power. While no model is available to directly compare in terms of prescription cost, this comparison provides a benchmark of how much predictive power can be gained over one of the more utilized models to date in regards to predicting prescription expenditures.

The predictive power of the RxCost Model was also tested for subgroups of the population with select disease states including asthma, depression, diabetes, HIV, and

hypertension. Here, the predictive ratio served as a marker for predictive power. The RxCost Model performed quite well for all subgroups only deviating at most 0.07 from the ideal ratio of 1.0.

To explore the potential use of the RxCost Model as a method for profiling physicians it was also run concurrently on the validation sample. Here year 1 diagnostic information was utilized to predict year 1 costs. The model performed well in comparison to the prospective model as R-square value increased over 50%. With an R-square value of 0.34, the model could potentially be very useful in profiling physicians to examine current prescribing behaviors.

In addition to a diagnostic model, an alternative model was developed, the MRxCost Model, to explore how much predictive power could be gained by supplementing the RxCost Model with drug information. Drug information proved to be quite powerful as the R-square values increased substantially for both the training and validation samples as compared to the RxCost Model. The predictive ratios for depression, diabetes, HIV infection and hypertension were very similar to the RxCost Model. However, the ratio for asthma was quite a bit worse predicting twice as much cost as actually incurred.

Viewing the pooled R-square value and predicted ratio for varying group sizes allows one to evaluate the performance of the models when utilized with different physician group sizes. Here, the models performed quite well in terms of both R-square and predicted ratio for a group size as small as 10 members. The R-square and predicted ratio did not improve until 200 members or more were utilized. Both the RxCost Model and the MRxCost Model increased uniformly with each other in terms of

R-square as the member size increased however the MRxCost Model did outperforming the RxCost Model. These pooled values indicate that the models could potentially be utilized to predict prescription expenditures for a physician or physician group with a patient base as small as 10 patients. In fact, the pooled values do not indicate a gain in prediction power until 200 patients or more are utilized and this gain in power is nominal until 300 or more patients are utilized.

The advantages gained by supplementing the diagnostic model with drug information must be tempered with some potential disadvantages. One advantage of drug information is that it can be more timely, reliable and complete than diagnostic information. This model identifies numerous individuals with conditions that were not diagnosed within the timeframe of the study period. Additionally, drug information helped explain severity in three conditions, asthma, cardiovascular, and seizure disorders. When supplemented with drug information, the diagnostic model performed much better in terms of discrimination. However, this model is associated with a certain amount of endogeneity as it uses drug information in a prior period which may be influenced by relative over or under prescribing to predict subsequent drug costs. To minimize the problem of endogeneity, the model only used binary variables to indicate whether a drug class was prescribed and did not count the number of prescriptions prescribed within a predefined class. Also, if this type model was used to influence prescribing behavior, it could potentially reward physicians for prescribing medication that would lead to the appearance of a more expensive patient-mix. This type of "gaming" happens with diagnostic model as well but there is no real utilization being consumed as there is with prescribing a medication. We did, however, attempt to

attend to this problem by utilizing hierarchies and combining drug categories with diagnostic categories when plausible. Regardless of the situation, the advantages and disadvantages of this type of mixed model should be carefully considered if this model were to be used to influence physician compensation.

The effect of outliers on the models was also of interest. We wanted to see the effect of trimming (capping) year 2 prescriptions costs at $20K. Here, the extreme values were not dropped from the model completely; only capped at a high level. This technique only affected 67 (0.03%) observations but had a substantial effect on R-square values as they increased almost 20% for both prospective models. This may be particularly useful if some plans employ re-insurance or stop-loss protections. These plans would be able to utilize this approach to even further increase the predictive power.

Overall, the R-square values obtained with the RxCost Model and the MRxCost Model were relatively high. Most general health risk assessment models produce R-square values at or below 0.18. However, prescription drug costs are, on average, less costly and have less variance than do general health care costs. For these reasons, the obtainable R-square values will be larger especially for models specifically derived to predict prescription drug costs. Despite this, the RxCost model did outperform the DCG-HCC model even after recalibrating to predict prescription drug costs.

Limitations

No risk assessment model will ever fully adjust for all differences in patient populations, nor will any model perfectly predict costs. The problem with trying to predict future costs is that random variation is so prevalent, however, the prescription

cost risk assessment models performed much better than the traditional age-and-sex method. The use of ICD-9-CM diagnosis and NDC drug codes presents another limitation. The use of this information has been found to be reliable but does not compare to the accuracy of clinical patient-specific data. Overall, there has been a lot of concern with accuracy, unreliability and clinical specificity of diagnostic information. (Iezzoni 1997; Romano 1994; Hannan 1992) Diagnosis-based indices rely solely on administrative claims-based coding of physician-patient encounters. Although ICD-9-CM diagnostic coding is not complete, the use of claims-based diagnostic coding in the past by both government and commercial programs for billing has resulted in substantial gains in completeness of data. A disadvantage, however, is that the completeness of data depends on the manner in which physicians are paid. If physicians are paid based on a discounted fee-for-service system, the data will be more complete than if physicians are paid on a sub-capitated system or fixed annual salary. One negative effect of reimbursement based on diagnosis is that a physician now has incentives to code more prolifically and to code ill-defined conditions in a manner to increase reimbursement. However, the hierarchical method utilized attempted to account for some of the potential up-coding. Finally, another disadvantage to using diagnosis-based models is that there can be a substantial lag time between when a patient receives services and when the diagnostic information is ready for use. Often a 6-month lag time is necessary to have reasonably complete diagnostic information available from health care encounters.

Another limitation is that the risk assessment weights proposed in this research may not be appropriate for each health plan. Due to varying formularies and benefit

design structures the weights may need to be re-estimated or at least re-calibrated to perform well in a plan setting.

Because risk assessment is based on averages and future cost uncertainty will always be high. A goal must be set to allow for some deviation from the predicted value. This is particularly important for any risk adjustment model used to adjust or otherwise profile small groups of patients or physician groups with relatively small practices.

CONCLUSION

In summary, the RxCost Model seems to do a nice job of predicting prescription expenditures both prospectively and concurrently. This model may serve well both setting prescribing goals for the coming year as well as profiling physician habits the previous year. The MRxCost Model also performed well when utilized prospectively, however this model should be weighed against the greater potential for gaming and rewarding practices for previously high prescription use.

REFERENCES

American Federation of State, County and M Employees. Prescription Drugs and State Budgets. Health Focus 2003; www.afscme.org/publications/healht_focus/focus302.htm.

Agency for Health Care Policy and Research. Department of Health and Human Services. www.ahrq.gov

Ash A, Porell F, Gruenberg L, Sawitz E, Beiser A. Adjusting Medicare capitation payments using prior hospitalization data. Health Care Financ Rev 1989; 10(4):17-29.

Ash AS, Ellis RP, Pope GC, Ayanian JZ, Bates DW, Burstin H et al. Using diagnoses to describe populations and predict costs. Health Care Financ Rev 2000; 21(3):7-28.

Bridges to Excellence. Diabetes Care Link for Physicians and Providers. 2004; www.bridgestoexcellence.com/bte/diabetescarelink/gty_physicians.htm.

Burton SL, Randel L, Titlow K, Emanuel EJ. The ethics of Pharmaceutical benefit management. Health Aff (Millwood) 2001; 20(5):150-163.

Carroll J. Physicians Reconsider Taking on Pharmacy Risk. Managed Care July 2000.

Clark DO, Von Korff M, Saunders K, Baluch WM, Simon GE. A chronic disease score with empirically derived weights. Med Care 1995; 33(8):783-795.

Clinical Classifications Software (ICD-9-CM) Summary and Download. Summary and Downloading Information. 2001 Agency for Health Care Policy and Research, Rockville, MD. http://www.ahrq.gov/data/hcup/ccs.htm

Diehr P, Yanez D, Ash A, Hornbrook M, Lin DY. Methods for Analyzing Health Care Utilization and Costs. Annual Review of Public Health 1999; 20:25-144.

Duan N, Manning WG, Morris CN, Newhouse JP. A Comparison of Alternative Models for the Demand of Medical Care. Journal of Business and Economic Statistics 1983; 1(2):115-126.

DxCG Inc. DxCG Risk Adjustment Software: Analytic Guide Release 6.0. 2001; Boston MA. support@dxcg.com

Employee Benefit Research Institute. Prescription Drug Costs Up Sharply - but Still Small Overall. Press Release 470 Washington EBRI . 1999.

Fishman PA, Shay DK. Development and estimation of a pediatric chronic disease score using automated pharmacy data. Medical Care 1999; 37:874-883.

Fishman PA, Goodman MJ, Hornbrook MC, Meenan RT, Bachman DJ, O'Keefe Rosetti MC. Risk Adjustment Usng Automated Ambulatory Pharmacy Data: The Risk Model. Medical Care 2003; 41(1): 84-99.

Galt KA, Rich C, Dralewski JE, Turner PD, Bernhardt TS, Dowd B, Feldman R, de Vries A. Group practice strategies to manage pharmaceutical cost in an HMO network. American Journal of Managed Care 2001; 7(11):1081-1090.

Gandhi SK, Arguelles L, Boyer JG. Economic impact of neutropenia and febrile neutropenia in breast cancer: estimates from two national databases. Pharmacotherapy 2001; 21(6):684-690.

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

Graden SE, Schafermeyer KW. Performance Reporting for Managed Care Prescription Programs. Journal of Managed Care Pharmacy 1998; 4(2): 160-166.

Hannan EL, Kilburn H, Lindsey M, et al. Clinical versus administrative databases for CABG surgery: does it matter? Medical Care 1992; 30:892-907.

Iezzoni L, Ash AS, Daley J, Hughes JS, Shwartz M. Risk adjustment for measuring healthcare outcomes. 2nd ed. Chicago: Health Administration Press, 1997.

Johnson RE, Hornbrok MC, Nichols GA. Replicating the chronic disease score (CDS) from automated pharmacy data. Journal of Clinical Epidemiology 1994; 47(10):1191-1199.

Kaelin JJ. Risk Adjustment for State Medicaid Programs: Lessons Learned and Prospects for the Future. Center for Health Program Development and Management April 2002; wwwacg.jhsph.edu/library/conference_2002/d2.plenary%203.%20kaelin.ppt

The Kaiser Family Foundation. Prescription Drug Trends. May 2003; www.kff.org/rxdrugs/3057-03-index.cfm

The Kaiser Family Foundation. Prescription Drugs: Facts at a Glance. 2003; www.kaisernetwork.org/static/spotlight_rxdrugs_facts.cfm.

The Kaiser Family Foundation. Medicaid and the Prescription Drug Benefit: Cost Containment Strategies and State Experiences. September 2002; www.kff.org/medicaid/4063.index.cfm

Kleinke JD. The Price of Progress: Prescription Drugs in the Health Care Market. Health Affairs 2001; 20(5):43-60.

Kronick R, Gilmer T, Dreyfus T, Lee L. Improving health-based payment for medicaid beneficiaries: CDPS. Health Care Financ Rev 2000; 21(3):29-64.

Lamers LM. Pharmacy cost groups: a risk-adjuster for capitation payments based on the use of prescribed drugs. Medical Care 1999; 37(8):824-830.

Litton LM, Sisk FA, Akins ME. Managing drug costs: the perception of managed care pharmacy directors. Am J Manag Care 2000; 6(7):805-814.

Mantel N. Why Stepdown procedures in variable Selection. Technometrics 1970; 12:621-625.

The MEDSTAT Group Incorporated. Ann Arbor, Michigan. 2001; www.medstat.com.

Motheral BR, Henderson R. The effects of a closed formulary on prescription drug use and costs. Inquiry 1999-00; 36(4):481-491.

Mullahy J. Much Ado About Two: Reconsidering Retransformation and the Two-Part Model in Health Econometrics. Journal of Health Economics 1998; 17:247-281.

The Multum Lexicon. Multum Information Systems Inc., editor. 2001; www.multum.com.

National Committee for Quality Assurance. Integrated Healthcare Association Pay for Performance Program: 2004 Clinical Measure Specifications and Audit Review. December 2003; www.iha.org

Ozminkowski RJ, Wang S, Marder WD, Azzolini J, Schutt D. Cost implications for the use of inhaled anti-inflammatory medications in the treatment of asthma. Pharmacoeconomics 2000; 18(3):253-264.

Romano PS, Roos LL, Luft HS, et al. A comparison of administrative versus clinical data: coronary artery bypass surgery as an example. Journal of Clinical Epidemiology 1994; 47:249-260.

SAS Institute Inc. 2004; www.sas.com

StataCorp. Stata Statistical Software: Release 6.0. 1999 College Station Texas

United physicians incentive program reduces pharmacy costs. PR Newswire. April 17, 2003.

van Houwelingen JC, le Cessie S. Predictive value of statistical models. Statistics in Medicine 1990; 8:1303-1325.
Von Korff M, Wagner EH, Saunders K. A chronic disease score from automated pharmacy data. J Clin Epidemiol 1992; 45(2):197-203.

Zhao SZ, Arguelles LM, Dedhiya SD, Morgan DG. Healthcare utilization associated with dyspepsia in patients with arthritis. Am J Manag Care 1999; 5(10):1285-1295.

Table 4.1

Data Counts Resulting from MarketScan Eligibility Requirements	
Total unduplicated count	9,043,605
Count after continuous eligibility requirements	3,016,354
Count after deleting people without drug coverage reported	2,320,553
Count after deleting members not 18 to 62 years of age	1,644,988
Count after deleting members admitted to institutions or with inpatient stays > 30 days	1,634,427

Table 4.2

New CCS-Based Classification
1. Infectious and parasitic diseases
1.1 Bacterial infection
1.1.1 Tuberculosis
1.1.2 Septicemia (except in labor)
1.1.3 Sexually transmitted infections (not HIV or hepatitis)
1.1.4 Other bacterial infections
1.2 Mycoses other than candidiasis
1.3 Hepatitis
2. HIV Infection
3. Neoplasms
4. Diabetes Mellitus
4.1 Diabetes mellitus without complication
4.2 Diabetes mellitus with other complications
5. Other endocrine, nutritional, and metabolic diseases and immunity disorders
5.1 Thyroid disorders
5.2 Disorders of lipid metabolism
5.3 Gout and other crystal arthropathies
5.4 Cystic fibrosis
5.5 Immunity disorders
5.6 Other nutritional, endocrine, and metabolic disorders
6. Diseases of the blood and blood-forming organs
6.1 Anemia
6.2 Coagulation and hemorrhagic disorders
6.3 Diseases of white blood cells AND other hematologic conditions
7. Mental retardation
8. Senility and organic mental disorders
9. Other mental disorders
9.1 Affective disorders
9.2 Schizophrenia and related disorders and other psychoses
9.3 Anxiety, somatoform, dissociative, and personality disorders
9.4 Other mental conditions
10. Alcohol and substance-related mental disorders
11. Eye Disorders
11.1 Glaucoma
11.2 Other eye disorders
12. Ear conditions
12.1 Otitis media and related conditions
12.2 Conditions associated with dizziness or vertigo AND other sense organ disorders

13. Diseases of the Central nervous system and other sense organs
13.1 Central nervous system infection
13.2 Hereditary and degenerative nervous system conditions
13.2.1 Parkinson's disease
13.2.2 Multiple sclerosis
13.3 Paralysis
13.4 Epilepsy, convulsions
13.5 Migraine
14. Hypertension
15. Heart valve disorders
16. Acute myocardial infarction
17. Coronary atherosclerosis and other heart disease
18. Congestive heart failure (nonhypertensive) and pulmonary heart disease
19. Cardiac dysrhythmias
20. Cerebrovascular disease
21. Diseases of arteries, arterioles, capillaries, veins and lymphatics
21.1 Peripheral and visceral atherosclerosis
21.2 Hypotention and Other unspecified circulatory disease
21.3 Phlebitis, thrombophlebitis and thromboembolism AND Varicose veins of lower extremity
22. Other diseases of the circulatory system
22.1 Peri-, endo-, and myocarditis, cardiomyopathy (except that caused by TB or STD)
22.2 Conduction disorders
23. Diseases of the respiratory system
23.1 Respiratory infections
23.1.1 Pneumonia (except that caused by TB or STD)
23.1.2 Influenza, Acute and chronic tonsillitis, Acute bronchitis AND other URIs
23.2 Chronic obstructive pulmonary disease and bronchiectasis
23.3 Asthma
23.4 Lung disease due to external agents
23.5 Other lower respiratory disease
23.6 Other upper respiratory disease
24. Diseases of the digestive system
24.1 Intestinal infection
24.2 Upper gastrointestinal disorders
24.2.1 Esophageal disorders
24.2.2 Gastroduodenal ulcer (except hemorrhage) AND Gastritis and duodenitis
24.2.3 Other disorders of stomach and duodenum
24.3 Lower gastrointestinal disorders AND Biliary tract disease
24.4 Pancreatic disorders (not diabetes)
24.5 Constipation, Dysphagia, and Other unspecified GI disorders
25. Liver disease

26. Diseases of the urinary system
26.1 Nephritis, nephrosis, renal sclerosis
26.2 Chronic renal failure
26.3 Urinary tract infections, Calculus of the urinary tract AND Other diseases of the kidney and ureters
27. Diseases of the genitourinary system
27.1 Diseases of male genital organs
27.2 Diseases of female genital organs
27.2.1 Nonmalignant breast conditions, Inflammatory diseases of female pelvic organs, Endometriosis, AND Prolapse of female genital organs
27.2.2 Menstrual disorders AND Menopausal disorders
27.2.3 Female infertility
28. Complications of pregnancy, childbirth, and the puerperium
28.1 Contraceptive and procreative management
28.2 Complications mainly related to pregnancy
28.2.1 Hypertension complicating pregnancy, childbirth and the puerperium
28.2.2 Diabetes or abn. glucose tolerance complicating pregn., childbirth, or the puerperium
29. Diseases of the skin and subcutaneous tissue
29.1 Skin and subcutaneous tissue infections AND Other inflammatory conditions of skin
29.2 Chronic ulcer of skin
30. Diseases of the musculoskeletal system and connective tissue
30.1 Infective arthritis and osteomyelitis (except that caused by TB or STD)
30.2 Non-traumatic joint disorders
30.3 Spondylosis, intervertebral disc disorders, other back problems
30.4 Osteoporosis AND Pathological fracture
30.5 Systemic lupus erythematosus and connective tissue disorders, Other connective tissure disease AND Other bone disease and musculoskeletal deformities
31. Congenital anomalies
31.1 Cardiac and circulatory congenital anomalies
31.2 Nervous system congenital anomalies
32. Certain conditions originating in the perinatal period
33. Injury and poisoning
33.1 Spinal cord injury
33.2 Intracranial injury
33.3 Crushing injury or internal injury
33.4 Open wounds
33.5 Burns
33.6 Complications of surgical procedures or medical care
34. Symptoms, signs, and ill-defined conditions and factors influencing health status

Table 4.3

Commercial RxCost Model: All Diagnostic Variables Submitted to OLS Estimation

#	Variable
1	Tuberculosis (CCS 1.1.1)
2	Septicemia (except in labor) (CCS 1.1.2)
3	Sexually transmitted infections (not HIV or hepatitis) (CCS 1.1.3)
4	Other bacterial infections (CCS 1.1.4)
5	Mycoses other than candidiasis (CCS 1.2)
6	Hepatitis (CCS 1.3)
7	HIV Infection (CCS 2)
8	Neoplasms (CCS 3)
9	Diabetes mellitus without complication (CCS 4.1)
10	Diabetes mellitus with other complications (CCS 4.2)
11	Thyroid disorders (CCS 5.1)
12	Disorders of lipid metabolism (CCS 5.2)
13	Gout and other crystal arthropathies (CCS 5.3)
14	Cystic fibrosis (CCS 5.4)
15	Immunity disorders (CCS 5.5)
16	Other nutritional, endocrine, and metabolic disorders (CCS 5.6)
17	Anemia (CCS 6.1)
18	Coagulation and hemorrhagic disorders (CCS 6.2)
19	Diseases of white blood cells AND other hematologic conditions (CCS 6.3)
20	Mental retardation (CCS 7)
21	Senility and organic mental disorders (CCS 8)
22	Affective disorders (CCS 9.1)
23	Schizophrenia and related disorders and other psychoses (CCS 9.2)
24	Anxiety, somatoform, dissociative, and personality disorders (CCS 9.3)
25	Other mental conditions (CCS 9.4)
26	Alcohol and substance-related mental disorders (CCS 10)
27	Glaucoma (CCS 11.1)]
28	Other eye disorders (CCS 11.2)
29	Otitis media and related conditions (CCS 12.1)
30	Conditions associated with dizziness or vertigo AND other sense organ disorders (CCS 12.2)
31	Central nervous system infection (CCS 13.1)
32	Parkinson's disease (CCS 13.2.1)
33	Multiple sclerosis (CCS 13.2.2)
34	Paralysis (CCS 13.3)
35	Epilepsy, convulsions (CCS 13.4)
36	Migraine (CCS 13.5)

37 Hypertension (CCS 14)
38 Heart valve disorders (CCS 15)
39 Acute myocardial infarction (CCS 16)
40 Coronary atherosclerosis and other heart disease (CCS 17)
41 Congestive heart failure (nonhypertensive) and pulmonary heart disease (CCS 18)
42 Cardiac dysrhythmias (CCS 19)
43 Cerebrovascular disease (CCS 20)
44 Peripheral and visceral atherosclerosis (CCS 21.1)
45 Hypotention and Other unspecified circulatory disease (CCS 21.2)
46 Phlebitis, thrombophlebitis and thromboembolism AND
Varicose veins of lower extremity (CCS 21.3)
47 Peri-, endo-, and myocarditis, cardiomyopathy (except that caused by
TB or STD) (CCS 22.1)
48 Conduction disorders (CCS 22.2)
49 Pneumonia (except that caused by TB or STD) (CCS 23.1.1)
50 Influenza, Acute and chronic tonsillitis, Acute bronchitis AND other URIs (CCS 23.1.2)
51 Chronic obstructive pulmonary disease and bronchiectasis (CCS 23.2)
52 Asthma (CCS 23.3)
53 Lung disease due to external agents (CCS 23.4)
54 Other lower respiratory disease (CCS 23.5)
55 Other upper respiratory disease (CCS23.6)
56 Intestinal infection (CCS 24.1)
57 Esophageal disorders (CCS 24.2.1)
58 Gastroduodenal ulcer (except hemorrhage) AND Gastritis and duodenitis (CCS 24.2.2)
59 Other disorders of stomach and duodenum (CCS 24.2.3)
60 Lower gastrointestinal disorders AND Biliary tract disease (CCS 24.3)
61 Pancreatic disorders (not diabetes) (CCS 24.4)
62 Constipation, Dysphagia, and Other unspecified GI disorders (CCS 24.5)
63 Liver disease (CCS 25)
64 Nephritis, nephrosis, renal sclerosis (CCS 26.1)
65 Chronic renal failure (CCS 26.2)
66 Urinary tract infections, Calculus of the urinary tract AND Other disesases of
the kidney and ureters (CCS 26.3)
67 Diseases of male genital organs (CCS 27.1)
68 Nonmalignant breast conditions, Inflammatory diseases of female pelvic organs, (CCS 27.2.1)
69 Menstrual disorders AND Menopausal disorders (CCS 27.2.2)
70 Female infertility (CCS 27.2.3)
71 Contraceptive and procreative management (CCS 28.1)
72 Hypertension complicating pregnancy, childbirth and the puerperium (CCS 28.2.1)
73 Diabetes or abn. glucose tolerance complicating pregn., childbirth,
or the puerperium (CCS 28.2.2)

74 Skin and subcutaneous tissue infections AND Other inflammatory conditions of skin (CCS 29.1)
75 Chronic ulcer of skin (CCS 29.2)
76 Infective arthritis and osteomyelitis (except that caused by TB or STD) (CCS 30.1)
77 Non-traumatic joint disorders (CCS 30.2)
78 Spondylosis, intervertebral disc disorders, other back problems (CCS 30.3)
79 Osteoporosis AND Pathological fracture (CCS 30.4)
80 Systemic lupus erythematosus and connective tissue disorders, Other connective tissure disease (CCS 30.5)
81 Cardiac and circulatory congenital anomalies (CCS 31.1)
82 Nervous system congenital anomalies (CCS 31.2)
83 Certain conditions originating in the perinatal period (CCS 32)
84 Spinal cord injury (CCS 33.1)
85 Intracranial injury (CCS 33.2)
86 Crushing injury or internal injury (CCS 33.3)
87 Open wounds (CCS 33.4)
88 Burns (CCS 33.5)
89 Complications of surgical procedures or medical care (CCS 33.6)
90 Symptoms, signs, and ill-defined conditions and factors influencing health status (CCS 34)

Table 4.4

New Drug Classification		
Rx #	**Rx Category**	**Drug Description and (Multum Classes)**
1	Alzheimers	Tacrine, Donepezil, Rivastigmine, Galantamine (in 80)
2	Anticoagulants	Heparins (261), Warfarin (262)
3	Asthma	Antiasthmatic combinations (131), Adrenergic bronchodilators (180), Bronchodialators combinations (181), Methylxantihines (126), Leukotriene modifier (243), Respiratory Inhalants (130)
4	Autoimmune	Azathioprine (in 104 and 192)
5	Burns	Silver Sulfadiazine (in140)
6	Cardiovascular	ACE inhibitors (42), Beta blockers (274 & 275), Nitrates (45), Vasodilators (52 & 53), Calcium channel blockers (48), Digoxin (in 50)
7	Arrhythmias	Antiarrhythmic agents (46)
8	Cystic fibrosis	Pancrelipase (in 91)
9	Depression/anxiety	Antidepressants (76, 208 & 209), Antianxiety (in 69 - alprazolam, chlordiazepoxide, clonazepam, clorazepate, diazepam, halazepam, lorazepam & oxazepam) and (in 70 - buspirone, doxepin, ethchlorvynol, & meprobamate)
10	Diabetes	Insulin (215), Sulfonylureas (213), Alpha-glucosidase inhibitors(216), Thiazolidinediones (271), Metformin (214)
11	ESRD/renal	Epoietin Alfa (in 36), Calcitriol (in 119)
12	Gastric acid disorders	H2 Blockers (272), PPIs (94)
13	Gout	Colchicine (in 194), Allopurinol (in 194)
14	Hepatitis	Interferon beta (in 256), Peginterferon (in 256 & 177), Ribavirin (in 229)
15	Herpes	Acyclovir (in 229)
16	HIV/AIDS	Antiretrovirals (175, 176, & 227)
17	Hyperlipidemia	Antihyperlipidemics (173 & 174)

18	Infections	Quinolones (14), Cephalosporins (159 - 162), Penicillins (223, 224, & 226), Macrolides (11), Sulfonamides (15), Tetracyclines (16), Penicillinase resistant Penicillins (222), Beta-lactamase inhibitors (225), Urinary anti-infectives (17)
19	Insomnia	Sedatives, Hypnotics (in 69 estazolam, flurazepam, midazolam, quazepam, temazepam & triazolam) (in 70 acetylcarbromal, chloral hydrate, chlormezanone, dexmedetomidine, doxylamine, hydroxyzine, paraldehyde, propiomazine, pyrilamine, zaleplon, & zolpidem)
20	Iron deficiency	Iron (in 116) (B8"Carbonyl Iron", "Iron Dextran", "Iron Polysaccharide", & "Multivitamin with Iron", "Iron Sucrose")
21	Liver disease	Lactulose (in 95)
22	Malignancies	Antineoplastics (22 - 25)
23	Multiple sclerosis/paralysis	Baclofen (178)
24	Nausea	Antiemetics (195 - 198 minus diphenhydramine)
25	Neurogenic bladder	Oxybutynin (in 264)
26	Osteoporosis/pagets	Etidronate/calcium regulators (217)
27	Pain	Narcotics (60 & 191)
28	Parkinsons/tremor	Dopaminergic antiparkinson agents (276), Benztropine (in 205), Trihexyphenidyl (in 205),
29	Psychotic illness/bipolar	Antipsychotics, Lithium (77 & 79)
30	Seizure disorders	Anticonvulsants (199 - 204)
31	Thyroid disorder	Thyroid hormones (103)
32	Transplant	Immunosuppressive agents (104)
33	Tuberculosis AND PCP pneumonia	Rifampin (in 232), Isoniazid (in231)

Table 4.5

New Drug Classification - Combined with CCS Categories		
Rx Category	**Drug Description and (Multum Classes)**	**Combined with CCS Category***
Asthma (Rx 3)	Antiasthmatic combinations (131), Adrenergic bronchodilators (180), Bronchodialators combinations (181), Methylxantihines (126), Leukotriene modifier (243), Respiratory Inhalants (130)	CCS 23.3 - Asthma
Cardiovascular (Rx6)	ACE inhibitors (42), Beta blockers (274 & 275), Nitrates (45), Vasodilators (52 & 53), Calcium channel blockers (48), Digoxin (in 50)	CCS 14 - Hypertension CCS 17 - Coronary atherosclerosis and other heart disease **Logic: 1.** Multum class 42 (ACE inhibitors), 48 (Ca channel blockers), 52 & 53 (Vasodilators), 274 & 275 (Beta blockers) combined with CCS 14. **2.** Multum class 45 (Nitrates) and Digoxin combined with CCS 17
Cystic fibrosis (Rx 8)	Pancrelipase (in 91)	CCS 5.4 Cystic fibrosis
Depression/anxiety (Rx 9)	Antidepressants (76, 208 & 209), Antianxiety (in 69 - alprazolam, chlordiazepoxide, clonazepam, clorazepate, diazepam, halazepam, lorazepam & oxazepam) and (in 70 - buspirone, doxepin, ethchlorvynol, & meprobamate)	CCS 9.1 - Affective disorders CCS 9.3 - Anxiety, somatoform, dissociative, and personality disorders CCS 9.4 - Other mental conditions **Logic:** if no diagnosis is present within CCS 9.1, 9.3, or 9.4 then combined with CCS 9.1

Diabetes (Rx10)	Insulin (215), Sulfonylureas (213), Alpha-glucosidase inhibitors(216), Thiazolidinediones (271), Metformin (214)	CCS 4.1 - Diabetes w/o complications CCS 4.2 - Diabetes w/ complications **Logic:** if no diagnosis is present within CCS 4.1 or 4.2 then combined with CCS 4.1
Gastric acid disorders (Rx12)	H2 Blockers (272), PPIs (94)	CCS 24.2.1 - Esophageal disorder CCS 24.2.2 - Gastroduodenal ulcer and gastritis and duodenitis **Logic:** if no diagnosis is present within CCS 24.2.1 or 24.2.2 then combined with CCS 24.2.2
Gout (Rx13)	Colchicine (in 194), Allopurinol (in 194)	CCS 5.3 - Gout and other crystal arthropathies
Hepatitis (Rx14)	Interferon beta (in 256), Peginterferon (in 256 & 177), Ribavirin (in229)	CCS 1.3 - Hepatitis
HIV/AIDS (Rx16)	Antiretrovirals (175, 176, & 227)	CCS 2 - HIV Infection
Hyperlipidemia (Rx 17)	Antihyperlipidemics (173 & 174)	CCS 5.2 - Disorders of lipid metabolism
Infections (Rx18)	Quinolones (14), Cephalosporins (159 - 162), Penicillins (223, 224, & 226), Macrolides (11), Sulfonamides (15), Tetracyclines (16), Penicillinase resistant Penicillins (222), Beta-lactamase inhibitors (225), Urinary anti-infectives (17)	CCS 1.1.1, 1.1.2, 1.1.3, 1.1.4, CCS 1.2, 1.3 CCS 2, CCS 12.1, 12.2 CCS 13.1, CCS 23.1.1, 23.1.2, CCS 24.1 CCS 26.3, CCS 29.1 - Infections **Logic:** If no diagnosis within CCS 1.1.1, 1.1.2, 1.1.3, 1.2, 1.3, 2, 12.1, 12.2, 13.1, 23.1.1, 23.1.2, 24.1, 26.3, or 29.1 then combined with CCS 1.1.4 - other bacterial infections

Malignancies (Rx 22)	Antineoplastics (22 - 25)	CCS 3 Neoplasms
Multiple sclerosis/paralysis (Rx23)	Baclofen (178)	CCS 13.2.2 - Multiple sclerosis
Osteoporosis/pagets (Rx 26)	Etidronate/calcium regulators (217)	CCS 30.4 - Osteoporosis and pathological fracture
Psychotic illness/bipolar (Rx 29)	Antipsychotics, Lithium (77 & 79)	CCS 9.2 - Schizophrenia and related disorders and other psychoses
Seizure disorders (Rx 30)	Anticonvulsants (199 - 204)	CCS 13.4 - Epilepsy, convulsions
Thyroid disorder (Rx31)	Thyroid hormones (103)	CCS 5.1 - Thyroid disorders

* Model variables are binomial 1/0 (1 if present). If diagnosis is present within combined CCS groups (variable = 1) then the presence of a Rx group will not affect the model.

Table 4.6

Commercial MRxCost Model: All Diagnostic Variables Submitted to OLS Estimation

#	Variable
1	Tuberculosis (CCS 1.1.1)
2	Septicemia (except in labor) (CCS 1.1.2)
3	Sexually transmitted infections (not HIV or hepatitis) (CCS 1.1.3)
4	Other bacterial infections (CCS 1.1.4)
5	Mycoses other than candidiasis (CCS 1.2)
6	Hepatitis (CCS 1.3)
7	HIV Infection (CCS 2)
8	Neoplasms (CCS 3)
9	Diabetes mellitus without complication (CCS 4.1)
10	Diabetes mellitus with other complications (CCS 4.2)
11	Thyroid disorders (CCS 5.1)
12	Disorders of lipid metabolism (CCS 5.2)
13	Gout and other crystal arthropathies (CCS 5.3)
14	Cystic fibrosis (CCS 5.4)
15	Immunity disorders (CCS 5.5)
16	Other nutritional, endocrine, and metabolic disorders (CCS 5.6)
17	Anemia (CCS 6.1)
18	Coagulation and hemorrhagic disorders (CCS 6.2)
19	Diseases of white blood cells AND other hematologic conditions (CCS 6.3)
20	Mental retardation (CCS 7)
21	Senility and organic mental disorders (CCS 8)
22	Affective disorders (CCS 9.1)
23	Schizophrenia and related disorders and other psychoses (CCS 9.2)
24	Anxiety, somatoform, dissociative, and personality disorders (CCS 9.3)
25	Other mental conditions (CCS 9.4)
26	Alcohol and substance-related mental disorders (CCS 10)
27	Glaucoma (CCS 11.1)]
28	Other eye disorders (CCS 11.2)
29	Otitis media and related conditions (CCS 12.1)
30	Conditions associated with dizziness or vertigo AND other sense organ disorders (CCS 12.2)
31	Central nervous system infection (CCS 13.1)
32	Parkinson's disease (CCS 13.2.1)
33	Multiple sclerosis (CCS 13.2.2)
34	Paralysis (CCS 13.3)
35	Epilepsy, convulsions (CCS 13.4)
36	Migraine (CCS 13.5)
37	Hypertension (CCS 14)
38	Heart valve disorders (CCS 15)

39 Acute myocardial infarction (CCS 16)
40 Coronary atherosclerosis and other heart disease (CCS 17)
41 Congestive heart failure (nonhypertensive) and pulmonary heart disease (CCS 18)
42 Cardiac dysrhythmias (CCS 19)
43 Cerebrovascular disease (CCS 20)
44 Peripheral and visceral atherosclerosis (CCS 21.1)
45 Hypotention and Other unspecified circulatory disease (CCS 21.2)
46 Phlebitis, thrombophlebitis and thromboembolism AND
Varicose veins of lower extremity (CCS 21.3)
47 Peri-, endo-, and myocarditis, cardiomyopathy (except that caused by
TB or STD) (CCS 22.1)
48 Conduction disorders (CCS 22.2)
49 Pneumonia (except that caused by TB or STD) (CCS 23.1.1)
50 Influenza, Acute and chronic tonsillitis, Acute bronchitis AND other URIs (CCS 23.1.2)
51 Chronic obstructive pulmonary disease and bronchiectasis (CCS 23.2)
52 Asthma (CCS 23.3)
53 Lung disease due to external agents (CCS 23.4)
54 Other lower respiratory disease (CCS 23.5)
55 Other upper respiratory disease (CCS23.6)
56 Intestinal infection (CCS 24.1)
57 Esophageal disorders (CCS 24.2.1)
58 Gastroduodenal ulcer (except hemorrhage) AND Gastritis and duodenitis (CCS 24.2.2)
59 Other disorders of stomach and duodenum (CCS 24.2.3)
60 Lower gastrointestinal disorders AND Biliary tract disease (CCS 24.3)
61 Pancreatic disorders (not diabetes) (CCS 24.4)
62 Constipation, Dysphagia, and Other unspecified GI disorders (CCS 24.5)
63 Liver disease (CCS 25)
64 Nephritis, nephrosis, renal sclerosis (CCS 26.1)
65 Chronic renal failure (CCS 26.2)
66 Urinary tract infections, Calculus of the urinary tract AND Other disesases of
the kidney and ureters (CCS 26.3)
67 Diseases of male genital organs (CCS 27.1)
68 Nonmalignant breast conditions, Inflammatory diseases of female pelvic organs, (CCS 27.2.1)
69 Menstrual disorders AND Menopausal disorders (CCS 27.2.2)
70 Female infertility (CCS 27.2.3)
71 Contraceptive and procreative management (CCS 28.1)
72 Hypertension complicating pregnancy, childbirth and the puerperium (CCS 28.2.1)
73 Diabetes or abn. glucose tolerance complicating pregn., childbirth,
or the puerperium (CCS 28.2.2)
74 Skin and subcutaneous tissue infections AND Other inflammatory
conditions of skin (CCS 29.1)
75 Chronic ulcer of skin (CCS 29.2)

76 Infective arthritis and osteomyelitis (except that caused by TB or STD) (CCS 30.1)
77 Non-traumatic joint disorders (CCS 30.2)
78 Spondylosis, intervertebral disc disorders, other back problems (CCS 30.3)
79 Osteoporosis AND Pathological fracture (CCS 30.4)
80 Systemic lupus erythematosus and connective tissue disorders, Other connective tissure disease (CCS 30.5)
81 Cardiac and circulatory congenital anomalies (CCS 31.1)
82 Nervous system congenital anomalies (CCS 31.2)
83 Certain conditions originating in the perinatal period (CCS 32)
84 Spinal cord injury (CCS 33.1)
85 Intracranial injury (CCS 33.2)
86 Crushing injury or internal injury (CCS 33.3)
87 Open wounds (CCS 33.4)
88 Burns (CCS 33.5)
89 Complications of surgical procedures or medical care (CCS 33.6)
90 Symptoms, signs, and ill-defined conditions and factors influencing health status (CCS 34)
91 Alzheimers (RX 1)
92 Anticoagulants (RX 2)
93 Autoimmune (RX 4)
94 Burns (RX 5)
95 Arrhythmias (RX 8)
96 ESRD/renal disorders (RX 11)
97 Herpes (RX 15)
98 Insomnia (RX 19)
99 Iron deficiency (RX 20)
100 Liver disease (RX 21)
101 Nausea (RX 24)
102 Neurogenic bladder (RX 25)
103 Pain (RX 27)
104 Parkinson's/tremor (RX 28)
105 Transplant (RX 32)
106 Tuberculosis AND PCP pneumonia (RX 33)
107 Number of Asthma drug classes
108 Number of Cardiovascular drug classes
109 Number of Seizure disorder drug classes

Table 4.7

Descriptive Statistics for MarketScan Training and Validation Data

	Training		Validation	
n	1,308,705		218,383	
Male	609,282	46.6%	101,882	46.7%
Female	699,423	53.4%	116,501	53.4%
Age (mean)	42.0		42.0	
FFS	858,166	65.6%	143,103	65.5%
Encounter	450,539	34.4%	75,280	34.5%
# of Months Eligible (mean)	23.1		23.1	
Demographic Model Variables				
male age 18-20	49,347	3.8%	8,311	3.8%
male age 23-30	70,413	5.4%	11,662	5.3%
male age 31-40	137,162	10.5%	22,799	10.4%
male age 41-50	165,401	12.6%	27,692	12.7%
male age 51-60	161,628	12.4%	27,120	12.4%
male age 61-64	25,331	1.9%	4,298	2.0%
female age 18-20	48,324	3.7%	8,062	3.7%
female age 23-30	80,874	6.2%	13,460	6.2%
female age 31-40	160,513	12.3%	26,474	12.1%
female age 41-50	202,136	15.4%	33,936	15.5%
female age 51-60	180,623	13.8%	30,120	13.8%
female age 61-64	26,953	2.1%	4,449	2.0%
Year 1 Rx Cost	$400.90		$397.84	
Year 2 Rx Cost	$468.67		$465.63	

Table 4.8

Results: RxCost Model Estimation

Number of variables = 87
Number of obs = 1,308,705
R-square = 0.17
*Adjusted R-square = 0.17
Root MSE = 1312.3

Variable	Coefficient	p-value	Lower 95% CI	Upper 95% CI
Male				
18 to 22 Years	Reference	-	-	-
23 to 30 Years	0.25	-	-	-
31 to 40 Years	24.04	-	-	-
41 to 50 Years	62.59	-	-	-
51 to 60 Years	153.25	-	-	-
61 to 64 Years	229.02	-	-	-
Female				
18 to 22 Years	38.50	-	-	-
23 to 30 Years	55.81	-	-	-
31 to 40 Years	65.27	-	-	-
41 to 50 Years	105.34	-	-	-
51 to 60 Years	213.32	-	-	-
61 to 64 Years	282.23	-	-	-
Dxn Categories				
Tuberculosis (CCS 1.1.1)	148.05	0.01	43.31	253.79
Septicemia (except in labor) (CCS 1.1.2)	1,022.94	0.00	943.87	1,102.01
Sexually transmitted infections (not HIV or hepatitis) (CCS 1.1.3)	168.54	0.00	124.84	212.25
Mycoses other than candidiasis (CCS 1.2)	57.87	0.00	43.39	72.34

Variable	Coefficient	p-value	Lower 95% CI	Upper 95% CI
Hepatitis (CCS 1.3)	601.50	0.00	566.63	636.37
HIV Infection (CCS 2)	7,025.12	0.00	6,946.74	7,103.50
Neoplasms (CCS 3)	137.80	0.00	130.93	144.67
[Hierarchy 1: estimated highest to least cost]				
[1] Diabetes mellitus with other complications (CCS 4.2)	956.16	0.00	936.43	975.88
[2] Diabetes mellitus without complication (CCS 4.1)	575.95	0.00	562.00	589.91
Thyroid disorders (CCS 5.1)	92.61	0.00	81.28	103.94
Disorders of lipid metabolism (CCS 5.2)	236.35	0.00	228.52	244.18
Cystic fibrosis (CCS 5.4)	3,415.35	0.00	3264.53	3566.18
Immunity disorders (CCS 5.5)	1,470.86	0.00	1387.52	1554.25
Other nutritional, endocrine, and metabolic disorders (CCS 5.6)	93.70	0.00	79.00	108.41
Anemia (CCS 6.1)	69.82	0.00	54.51	85.13
Coagulation and hemorrhagic disorders (CCS 6.2)	553.05	0.00	513.70	592.40
Diseases of white blood cells AND other hematologic conditions (CCS 6.3)	364.11	0.00	330.00	398.21
Senility and organic mental disorders (CCS 8)	476.29	0.00	415.42	537.16
Schizophrenia and related disorders and other psychoses (CCS 9.2)	816.91	0.00	754.50	879.31
[Hierarchy 2: estimated highest to least cost]				
[1] Affective disorders (CCS 9.1)	734.67	0.00	721.09	748.25
[2] Anxiety, somatoform, dissociative, and personality disorders (CCS 9.3)	255.70	0.00	239.53	271.87
[3] Other mental conditions (CCS 9.4)	227.76	0.00	213.03	242.50
Glaucoma (CCS 11.1)]	158.68	0.00	138.75	178.61
Other eye disorders (CCS 11.2)	94.42	0.00	85.85	102.99
Otitis media and related conditions (CCS 12.1)	44.25	0.00	31.07	57.44
Central nervous system infection (CCS 13.1)	547.88	0.00	470.77	624.99
Parkinson's disease (CCS 13.2.1)	2,323.29	0.00	2,209.92	2,436.66
Multiple sclerosis (CCS 13.2.2)	3,873.69	0.00	3,821.96	3,925.43
Paralysis (CCS 13.3)	404.17	0.00	336.95	471.40
Epilepsy, convulsions (CCS 13.4)	634.39	0.00	602.57	666.20
Migraine (CCS 13.5)	603.57	0.00	585.11	622.03
Hypertension (CCS 14)	299.35	0.00	291.98	306.72
Heart valve disorders (CCS 15)	52.84	0.00	33.16	72.52

Variable	Coefficient	p-value	Lower 95% CI	Upper 95% CI
Coronary atherosclerosis and other heart disease (CCS 17)	577.10	0.00	561.51	592.68
Acute myocardial infarction (CCS 16)	475.88	0.00	427.98	523.78
Congestive heart failure (nonhypertensive) and pulmonary heart disease (CCS 18)	672.47	0.00	637.78	707.15
Cardiac dysrhythmias (CCS 19)	108.86	0.00	94.37	123.36
Cerebrovascular disease (CCS 20)	311.86	0.00	283.83	339.90
Peripheral and visceral atherosclerosis (CCS 21.1)	251.98	0.00	213.31	290.65
Hypotention and Other unspecified circulatory disease (CCS 21.2)	110.12	0.00	93.64	126.60
Phlebitis, thrombophlebitis and thromboembolism AND Varicose veins of lower extremity (CCS 21.3)	200.79	0.00	173.84	227.74
Peri-, endo-, and myocarditis, cardiomyopathy (except that caused by TB or STD) (CCS 22.1)	461.17	0.00	418.37	503.98
Conduction disorders (CCS 22.2)	103.64	0.00	46.53	160.76
Pneumonia (except that caused by TB or STD) (CCS 23.1.1)	290.91	0.00	268.92	312.91
[Hierarchy 3: estimated highest to least cost]				
[1] Other upper respiratory disease (CCS23.6)	192.33	0.00	183.76	200.91
[2] Influenza, Acute and chronic tonsillitis, Acute bronchitis AND other URIs (CCS 23.1.2)	39.20	0.00	33.39	45.00
Chronic obstructive pulmonary disease and bronchiectasis (CCS 23.2)	169.27	0.00	155.90	182.63
Asthma (CCS 23.3)	422.20	0.00	407.64	436.75
Lung disease due to external agents (CCS 23.4)	76.87	0.08	-7.68	161.43
Other lower respiratory disease (CCS 23.5)	130.17	0.00	120.15	140.20
Intestinal infection (CCS 24.1)	108.33	0.00	75.64	141.03
[Hierarchy 4: estimated highest to least cost]				
[1] Esophageal disorders (CCS 24.2.1)	411.79	0.00	398.03	425.56
[2] Gastroduodenal ulcer (except hemorrhage) AND Gastritis and duodenitis (CCS 24.2.2)	174.14	0.00	152.35	195.94
Other disorders of stomach and duodenum (CCS 24.2.3)	95.35	0.00	68.22	122.47
Lower gastrointestinal disorders AND Biliary tract disease (CCS 24.3)	197.08	0.00	183.09	211.06
Pancreatic disorders (not diabetes) (CCS 24.4)	540.91	0.00	476.98	604.85

Variable	Coefficient	p-value	Lower 95% CI	Upper 95% CI
Constipation, Dysphagia, and Other unspecified GI disorders (CCS 24.5)	96.96	0.00	84.53	109.38
Liver disease (CCS 25)	198.77	0.00	175.86	21.68
Nephritis, nephrosis, renal sclerosis (CCS 26.1)	222.18	0.00	152.13	292.23
Chronic renal failure (CCS 26.2)	2,607.01	0.00	2,548.38	2,665.64
Urinary tract infections, Calculus of the urinary tract AND Other disesases of the kidney and ureters (CCS 26.3)	52.68	0.00	43.00	62.35
Diseases of male genital organs (CCS 27.1)	67.14	0.00	54.95	79.32
Female infertility (CCS 27.2.3)	548.54	0.00	513.26	583.82
Skin and subcutaneous tissue infections AND Other inflammatory conditions of skin (CCS 29.1)	141.93	0.00	130.75	153.11
Chronic ulcer of skin (CCS 29.2)	326.73	0.00	277.60	375.87
[Hierarchy 5: estimated highest to least cost]				
[1] Infective arthritis and osteomyelitis (except that caused by TB or STD) (CCS 30.1)	510.67	0.00	447.82	573.53
[2] Osteoporosis AND Pathological fracture (CCS 30.4)	427.04	0.00	406.19	447.90
[3] Non-traumatic joint disorders (CCS 30.2)	286.61	0.00	278.63	294.59
[4] Systemic lupus erythematosus and connective tissue disorders, Other connective tissure disease (CCS 30.5)	133.62	0.00	125.53	141.70
[5] Spondylosis, intervertebral disc disorders, other back problems (CCS 30.3)	112.49	0.00	102.27	122.70
Cardiac and circulatory congenital anomalies (CCS 31.1)	46.43	0.07	-3.88	96.73
Certain conditions originating in the perinatal period (CCS 32)	85.26	0.00	53.56	116.96
Spinal cord injury (CCS 33.1)	242.85	0.00	144.01	341.69
Open wounds (CCS 33.4)	23.15	0.00	7.23	39.07
Complications of surgical procedures or medical care (CCS 33.6)	438.26	0.00	406.48	470.04
Symptoms, signs, and ill-defined conditions and factors influencing health status (CCS 34)	54.88	0.00	48.40	61.36

* R-square adjusted for shrinkage using the van Howelingen method:
R^2 adj = 1 - (1-R^2) x (n-1) / (n-p-1)

Table 4.9

Results: MRxCost Model Estimation

Number of variables = 83
Number of obs = 1,308,705
R-square = 0.26
Adjusted R-square = 0.26
Root MSE = 1227

Variable	Coefficient	p-value	Lower 95% CI	Upper 95% CI
Male				
18 to 22 Years	Reference	-	-	-
23 to 30 Years	3.79	-	-	-
31 to 40 Years	15.49	-	-	-
41 to 50 Years	26.23	-	-	-
51 to 60 Years	48.42	-	-	-
61 to 64 Years	63.28	-	-	-
Female				
18 to 22 Years	45.50	-	-	-
23 to 30 Years	44.77	-	-	-
31 to 40 Years	42.65	-	-	-
41 to 50 Years	45.53	-	-	-
51 to 60 Years	68.60	-	-	-
61 to 64 Years	80.52	-	-	-
Dxn Categories				
Tuberculosis (CCS 1.1.1)	104.63	0.039	5.40	203.86
Septicemia (except in labor) (CCS 1.1.2)	521.98	0.000	448.04	595.91
Other bacterial infections (CCS 1.1.4)	33.11	0.000	27.56	38.67
Mycoses other than candidiasis (CCS 1.2)	57.14	0.000	43.62	70.67

Variable	Coefficient	p-value	Lower 95% CI	Upper 95% CI
Hepatitis (CCS 1.3)	485.21	0.000	452.59	517.82
HIV Infection (CCS 2)	6,574.96	0.000	6,508.03	6,641.90
Neoplasms (CCS 3)	457.62	0.000	451.19	464.06
[Hierarchy 1: estimated highest to least cost]				
[1] Diabetes mellitus with other complications (CCS 4.2)	731.75	0.000	713.22	750.27
[2] Diabetes mellitus without complication (CCS 4.1)	484.84	0.000	473.02	496.67
Thyroid disorders (CCS 5.1)	130.45	0.000	121.50	139.41
Disorders of lipid metabolism (CCS 5.2)	339.27	0.000	332.42	346.12
Gout and other crystal arthropathies (CCS 5.3)	156.60	0.000	135.61	177.59
Cystic fibrosis (CCS 5.4)	3,332.83	0.000	3,191.00	3,474.66
Immunity disorders (CCS 5.5)	1,211.87	0.000	1,133.92	1,289.82
Coagulation and hemorrhagic disorders (CCS 6.2)	361.83	0.000	324.68	398.98
Diseases of white blood cells AND other hematologic conditions (CCS 6.3)	220.15	0.000	188.30	252.01
Senility and organic mental disorders (CCS 8)	224.88	0.000	167.43	282.33
Schizophrenia and related disorders and other psychoses (CCS 9.2)	773.58	0.000	746.08	801.09
[Hierarchy 2: estimated highest to least cost]				
[1] Affective disorders (CCS 9.1)	467.85	0.000	461.11	474.59
[2] Anxiety, somatoform, dissociative, and personality disorders (CCS 9.3)	240.62	0.000	225.48	255.75
[3] Other mental conditions (CCS 9.4)	237.23	0.000	223.44	251.02
Glaucoma (CCS 11.1)]	133.81	0.000	115.17	152.44
Other eye disorders (CCS 11.2)	60.76	0.000	52.77	68.75
Central nervous system infection (CCS 13.1)	287.67	0.000	215.54	359.79
Parkinson's disease (CCS 13.2.1)	1,845.71	0.000	1,737.58	1,953.83
Multiple sclerosis (CCS 13.2.2)	2,578.22	0.000	2,539.13	2,617.31
Epilepsy, convulsions (CCS 13.4)	76.26	0.000	41.22	111.30
Migraine (CCS 13.5)	311.92	0.000	294.54	329.31
Hypertension (CCS 14)	89.86	0.000	81.47	98.25
Coronary atherosclerosis and other heart disease (CCS 17)	108.01	0.000	93.65	122.38
Congestive heart failure (nonhypertensive) and pulmonary heart disease (CCS 18)	28.70	0.027	3.29	54.12

Variable	Coefficient	p-value	Lower 95% CI	Upper 95% CI
Cerebrovascular disease (CCS 20)	78.20	0.000	52.03	104.37
Peripheral and visceral atherosclerosis (CCS 21.1)	89.18	0.000	53.01	125.34
Hypotention and Other unspecified circulatory disease (CCS 21.2)	37.71	0.000	22.34	53.09
Phlebitis, thrombophlebitis and thromboembolism AND Varicose veins of lower extremity (CCS 21.3)	86.38	0.000	60.84	111.92
Peri-, endo-, and myocarditis, cardiomyopathy (except that caused by TB or STD) (CCS 22.1)	160.51	0.000	120.45	200.57
Conduction disorders (CCS 22.2)	107.12	0.000	86.66	127.58
Chronic obstructive pulmonary disease and bronchiectasis (CCS 23.2)	20.34	0.001	7.84	32.85
[Hierarchy 3: estimated highest to least cost]				
[1] Esophageal disorders (CCS 24.2.1)	447.05	0.000	434.41	459.71
[2] Gastroduodenal ulcer (except hemorrhage) AND Gastritis and duodenitis (CCS 24.2.2)	597.27	0.000	588.55	606.00
Lower gastrointestinal disorders AND Biliary tract disease (CCS 24.3)	60.40	0.000	47.47	73.32
Liver disease (CCS 25)	93.55	0.000	72.18	114.91
Chronic renal failure (CCS 26.2)	510.14	0.000	452.02	568.26
Diseases of male genital organs (CCS 27.1)	72.29	0.000	60.90	83.69
Female infertility (CCS 27.2.3)	465.93	0.000	432.88	498.99
Skin and subcutaneous tissue infections AND Other inflammatory conditions of skin (CCS 29.1)	99.88	0.000	89.42	110.35
Chronic ulcer of skin (CCS 29.2)	221.68	0.000	175.71	267.64
[Hierarchy 4: estimated highest to least cost]				
[1] Infective arthritis and osteomyelitis (except that caused by TB or STD) (CCS 30.1)	301.72	0.000	242.92	360.52
[2] Osteoporosis AND Pathological fracture (CCS 30.4)	404.76	0.000	387.47	422.06
[3] Non-traumatic joint disorders (CCS 30.2)	143.34	0.000	135.80	150.87
[4] Systemic lupus erythematosus and connective tissue disorders, Other connective tissure disease (CCS 30.5)	27.55	0.000	19.98	35.13
[5] Spondylosis, intervertebral disc disorders, other back problems (CCS 30.3)	5.89	0.229	-3.71	15.49
Complications of surgical procedures or medical care (CCS 33.6)	114.33	0.000	84.56	144.09

Variable	Coefficient	p-value	Lower 95% CI	Upper 95% CI
Alzheimers (RX 1)	1,235.11	0.000	1,077.16	1,393.06
Anticoagulants (RX 2)	382.42	0.000	354.60	410.24
Autoimmune (RX 4)	36.25	0.000	1,141.04	1,283.15
Burns (RX 5)	186.05	0.000	142.73	229.37
Arrhythmias (RX 8)	1,808.99	0.000	1,722.01	1,895.98
ESRD/renal disorders (RX 11)	3,055.79	0.000	2,972.55	31.39.03
Herpes (RX 15)	233.83	0.000	213.30	254.36
Insomnia (RX 19)	357.30	0.000	345.18	369.43
Iron deficiency (RX 20)	105.66	0.000	69.12	142.20
Liver disease (RX 21)	675.05	0.000	613.68	736.42
Nausea (RX 24)	185.34	0.000	174.91	195.78
Neurogenic bladder (RX 25)	608.79	0.000	567.90	649.67
Pain (RX 27)	84.12	0.000	78.28	89.95
Parkinson's/tremor (RX 28)	381.96	0.000	353.90	410.02
Transplant (RX 32)	610.85	0.000	518.91	702.78
Tuberculosis AND PCP pneumonia (RX 33)	4,874.83	0.000	4,793.32	4,956.34
No. classes within Asthma	331.54	0.000	326.09	337.00
No. classes within Cardiovascular	277.25	0.000	270.27	284.22
No. classes within Seizure disorders	675.70	0.000	648.87	702.53

* R-square adjusted for shrinkage using the van Howelingen method:

R^2 adj = 1 - (1-R^2) x (n-1) / (n-p-1)

Table 4.10

Model R-Square Values - MarketScan with Untrimmed Data (Coefficients Estimated using Untrimmed Training Sample)			
Model	**Training Sample Untrimmed**	**Validation Sample Untrimmed**	**Validation Sample Trimmed**
Demographics Model	0.03	0.03	0.05
RxCost Model	0.17	0.22	0.26
Concurrent RxCost Model	-	0.34	0.53
MRxCost Model	0.26	0.34	0.40
DxCG Model	-	0.16	-

Table 4.11

Pooled R-Square and Predicted Ratio Values for Groups of Members

	RxCost Model		MRxCost Model	
Group Size	**R-Square Value**	**Predicted Ratio (95% CI)**	**R-Square Value**	**Predicted Ratio (95% CI)**
10 members	0.23	1.04 (1.03, 1.05)	0.29	1.04 (1.03, 1.05)
20 members	0.23	1.04 (1.03, 1.05)	0.29	1.04 (1.03, 1.05)
50 members	0.23	1.04 (1.03, 1.05)	0.29	1.04 (1.03, 1.05)
100 members	0.23	1.03 (1.02, 1.04)	0.29	1.04 (1.03, 1.05)
200 members	0.24	1.03 (1.02, 1.04)	0.30	1.04 (1.03, 1.05)
300 members	0.26	1.03 (1.02, 1.04)	0.34	1.04 (1.03, 1.05)
400 members	0.30	1.03 (1.02, 1.04)	0.38	1.04 (1.03, 1.05)
500 members	0.30	1.03 (1.02, 1.04)	0.38	1.04 (1.03, 1.05)

CHAPTER 5

DEVELOPMENT AND VALIDATION OF A CLAIMS-BASED RISK ASSESSMENT MODEL TO PREDICT PHARMACY EXPENDITURES IN A MEDICAID POPULATION[1]

[1]CR Cantrell, BC Martin. To be submitted to *Medical Care.*

ABSTRACT

Objectives: Empirically develop and validate the Medicaid RxCost Model, a risk assessment model that uses claims-based, diagnostic information to predict future pharmacy expenditures for a Medicaid population. Additionally, another model, the Medicaid MRxCost Model, was empirically developed and validated as well to explore the gain in predictive power associated with adding drug information to the Medicaid RxCost Model. The performance of these models was compared to each other as well as a demographic-only model, a demographic and Medicaid eligibility model and the CDPS model.

Study Design: Retrospective longitudinal cohort study

Data Sources: Three years, 1998 through 2000, of a 20% sample of California Medicaid claims data.

Subjects: All persons enrolled in the dataset who had continuous enrollment for at least 13 months, who were 18 to 64 years of age and did not meet the exclusion criteria

Methods: A training sample consisting of 138,454 lives was utilized to develop the models. Initially, both OLS and a two-part model were evaluated but OLS was chosen to estimate the model coefficients. A random holdout sample of 92,621 lives was utilized to validate the models and to compare the performance of each model.

Results: The R-square value for the Medicaid RxCost Model, the Medicaid MRxCost Model and the CDPS using the validation sample was 0.24, 0.30 and 0.04 respectively.

Conclusions: The Medicaid RxCost Model was successfully developed and it substantially outperformed the CDPS model in terms of R-square even after re-calibrating the CDPS model. The Medicaid MRxCost Model also proved that

supplementing drug information can improve discriminatory power although discretion should be utilized.

Key Words: Risk assessment, risk adjustment, prescription cost, risk models, drug cost

INTRODUCTION

Spending on prescription drugs in the U.S. rose at double-digit rates throughout the past decade. (Employee Benefit Research Institute 1999) Although outpatient prescription drug spending represented only a small portion (11% or $140.6 billion) of personal health spending in 2001, it was one of the fastest growing components (The Kaiser Family Foundation 2003). From 1997 to 2001, three primary factors are attributed to the growth in prescription drug expenditure; increased utilization (47%), use of newer, more expensive medications (27%) and prices increases (26%). (The Kaiser Family Foundation 2003).

The Medicaid program, which plays a fundamental role in providing the low-income population with prescription drug benefits, has faced challenges with increasing prescription expenditures. Spending for prescription drugs by the Medicaid program has grown at an annual average of 14.8% from 1990 to 1998. (The Kaiser Family Foundation new – Prescription drug trends: chartbook 2001) Medicaid spent an estimated $21 billion on for outpatient prescription drugs in 2000 representing roughly 10 percent of the total Medicaid budget. (The Kaiser Family Foundation 2002) The average drug expenditure per Medicaid enrollee was $358 in 1998.

As a result of the rising cost of drugs and the increased utilization, especially of newer, more expensive drugs, Medicaid programs have explored various strategies for managing pharmacy benefits including drug utilization review, federally allowed

exclusions and generic substitutions, prior authorizations and preferred drug formularies. (The Kaiser Family Foundation 2002)

Although the techniques described above remain an important component in managing prescription expenditures, another more innovative way to control drug costs and utilization has been to shift the risk or otherwise incentivize providers. There is a growing trend toward paying more for higher quality. Two such movements of late are the Pay for Performance (P4P) initiative in California and the Bridges to Excellence (BTE) program for diabetes care (National Committee for Quality Assurance 2003, Bridges to Excellence 2004). Both programs provide physicians with financial incentives (bonuses) to provide high quality care. This approach may well be used in the future to promote judicious prescribing behaviors since it is the physician that has the best knowledge of the patient and acts as the patient's agent to select prescription drugs. Creating structures that financially affect the patient through copays, formularies, and other patient incentives only indirectly influences physician prescribing decisions since the physicians themselves are not directly rewarded for more judicious prescribing. Physician profiling and physician capitation have both been used in the past as a means of shifting risk to providers. (Burton 2001; Carroll 2000; United 2003) Recently, physician profiling was used to introduce financial incentives for physicians to prescribe effectively and efficiently. (AFSCME 2003; United 2003; Graden 1998) Bonus structures reward physicians who prescribe judiciously and hold accountable those that do not. In addition, capitation for prescription drug costs passes financial incentives to control drug spending on to physicians, who are paid a set allowance for each patient that must cover the patient's yearly prescription drug costs, with any excess coming out

of the physician's own pocket. (Burton 2001) These new techniques have not been used extensively but may become more commonplace in the future as health plans increasingly focus on prescription expenditures. The advantage of using these methods to shift costs to providers is that it gives physicians an incentive to resist excessive consumer demand for expensive or overprescribed drugs. (Burton 2001) Disadvantages can include an increased emphasis on cost rather than medical necessity in drug choice and the creation of adverse selection incentives where more healthy patients are chosen over less healthy patients by physicians. To combat adverse selection, these profiling techniques should incorporate risk assessment in order to adjust for differences in patient case-mix. Risk assessment techniques attempt to control for clinical differences in patient risk and to isolate quality differences. Risk adjustment techniques can then be utilized to allow physicians who treat a more severe patient-mix higher prescription drug expenditures. When shifting risk to providers, the goal of risk assessment/adjustment is to develop a system that doesn't penalize physicians who treat high-risk patients. Without risk assessment/adjustment physicians will argue that their patients are sicker than plan averages.

Risk assessment/adjustment models are not new to state Medicaid programs. As of 1992, ten states (CO, DE, MD, MI, MN, NJ, OR, TN, UT and WA) utilized risk adjustment models mostly for payment purposes. (Kaelin 2002) Numerous other states have evaluated risk adjustment models in an effort to promote more efficient payment. Four states, CO, MD, MI and WA, utilized risk adjustment models for purposes other than payment such as profiling and disease management. With increased attention focused on prescription drug expenditures, Medicaid programs may begin offering

financial incentives for judicious prescribing behaviors. Their knowledge of risk assessment/adjustment would allow them to utilize these methods to build in incentives based on a physician's patient-mix.

To date there are no published literature on risk assessment models specifically designed to predict prescription expenditures. The objective of this research is to empirically develop a diagnostic-based risk assessment model to predict prescription expenditures for a Medicaid population. Additionally, a model using both diagnostic and drug information will be developed to explore the advantages of supplementing a diagnostic model with drug information. The models could potentially be used to set prescribing goals for physicians or physician groups based on their patient population. These goals could then be used to build in financial incentives or bonuses for those physicians that achieve their goals.

METHODS

A retrospective longitudinal review was employed to empirically develop risk assessment models to predict pharmacy expenditures for a Medicaid population. Four new models were developed separately utilizing claims data from the California Medicaid administrative claims database; a demographic-only model, a demographic and Medicaid eligibility model, a diagnostic-based model (Medicaid RxCost Model) and a model incorporating both diagnostic and drug information (Medicaid MRxCost Model).

The risk assessment models were developed in a prospective fashion and utilized information from one plan year to predict pharmacy expenditures for the following year. The demographic model and the demographic and Medicaid eligibility model were used to establish a baseline to observe the predictive power of a model that

can be derived with minimal effort. The Medicaid RxCost Model was the primary model for this research. The Medicaid MRxCost Model was used to explore how much predictive power would be obtained by adding drug information. A 33% random sample was utilized to validate both the diagnostic and mixed models (two-thirds of a 50% random sample; one-third was set aside as a spare dataset).

Data

Three years, 1998 through 2000, of California Medicaid (MediCal) administrative data was utilized to develop and validate the risk assessment indices. The data consists of a 20% sample of California Medicaid claims and is available through the California Department of Human services. (CDHS 2001) The data consists of enrollment information, prescription claims, inpatient and outpatient hospital claims, and ambulatory claims in a recipient level linkable format. This database provides the level of detail on costs and utilization needed to develop and validate risk assessment models. The data have been found to be valid in previous epidemiological and health policy studies and have been used in developing other risk assessment indices. (Adams 2001; Perkins 2001; Croghan 1999) The total unduplicated count for the MediCal dataset was 1,689,019 lives. All methods employed to mine the data and generate the model samples were performed using SAS Version 8.2 (SAS Inc. 2004).

Subjects

Persons meeting the following criteria were included in the MediCal sample to develop and validate the Medicaid risk assessment indices:

- Not dually eligible for Medicare
- Continuously eligible for a minimum of thirteen months from January 1998 through December 2000
- Age 18 to 62 years at the beginning of the observational period
- Eligible for Aid to Families with Dependent Children (AFDC), Temporary Assistance for Needy Children (TANF) or Supplemental Security Income (SSI)
- Not admitted to institutions or nursing home facilities and who do not have periods of inpatient care in excess of 30 consecutive days.

The member count for the MediCal sample dropped from approximately 1.68 million to 276 thousand after excluding members who did not meet the above criteria Table 5.1. Most members were lost due to the age requirement and the continuous eligibility requirement. From the 276 thousand members who met the criteria for inclusion, a random 50% sample was selected, referred to as the training sample, to develop the risk assessment models. Of the remaining 50%, two-thirds of the sample was randomly chosen to validate the models. The remaining one-third was set aside as spare data in case problems were encountered. The percentages for each sample were chosen to maximize the number of members available for development of the models while maintaining an adequate number in which to validate the models. This study was reviewed and approved by the University of Georgia Institutional Review Board; IRB # H2002-10473-2.

Risk Assessment

The demographic-only model was empirically developed using the following mix of age-sex dummy variables where 1 indicates presence and 0 indicates absence:

» (Male 18-22 years), (Male 23-30 years), (Male 31-40 years)
(Male 41-50 years), (Male 51-60 years)

» (Female 18-22 years), (Female 23-30 years), (Female 31-40 years),
(Female 41-50 years), (Female 51-60 years), (Female 61-64 years)

Medicaid eligibility status is a contributing factor in prescription drug utilization. (Baugh 1999) Members were included in the Medicaid sample if they were eligible for TANF, AFDC or SSI. The Medicaid Demographic and Eligibility Model adds Medicaid eligibility information to the Demographic Model defined above. Due to TANF replacing AFDC, the aid categories were very similar or even identical in the MediCal system. The Medicaid Demographic and Eligibility Model treats TANF and AFDC as one category and SSI as a separate category. The model includes a dummy (0 / 1) variable to indicate Medicaid eligibility status for SSI. When an individual is eligible for both TANF/AFDC and SSI they are treated as being eligible for SSI.

The Medicaid RxCost Model was the primary model for this research and was developed utilizing diagnostic information in addition to demographic information and Medicaid Eligibility. With more than 12,000 diagnostic codes making up the ICD-9-CM classification system, the first challenge is to aggregate these diagnostic codes into clinically meaningful categories that reflect similar prescription expenditures. Some of the techniques used by Ash and Kronick in developing the Diagnostic Cost Groups-Hierarchical Coexisting Conditions (DCG-HCC) (Ash 2000) and Chronic Illness and

Disability Payment System (CDPS) (Kronick 2000) were used to guide the development of this project. The initial classification system to organize diagnostic codes was based on AHRQ's Clinical Classification Software (CCS). (Clinical Classification Software 2001) The CCS is a multi-level classification scheme that aggregates individual ICD-9-CM codes into clinically meaningful categories that group similar conditions. (Agency for Health Care Policy and Research) The CCS Software is available for download at www.ahrq.org. The multi-level scheme used by CCS aggregates ICD-9-CM codes into 17 broad categories (e.g., Infectious Diseases, Neoplasms, and Mental Disorders) excluding the residual E codes. Within each of the 17 broad categories of the CCS, a multilevel categorization scheme divides each of the broad categories into more specific refined categories. The determining factor in creating these categories was the extent to which conditions could be grouped into relatively homogeneous clusters of interest to public policy researchers. These CCS categories were not created specifically for prescription expenditures and, therefore, do not necessarily reflect conditions with similar prescription expenditures.

Based on the prior use of diagnostic models to predict total health care costs, (Ash 2000, Kronick 2000) and a clinical panel review (consisting of two practicing clinical pharmacists) 34 major categories with 91 subcategories were proposed for aggregating diagnostic codes. The low-cost and ill-defined conditions were excluded and numerous CCS subcategories were collapsed into broader categories. For example, all subcategories under Neoplasms were collapsed into 1 major category “Neoplasms”. The final CCS-based diagnostic classification is presented in Table 5.2.

All categories that are missing in the final models were either discarded or collapsed into the broader categories.

The new diagnostic classification system was applied to the MediCal data. The prevalence of the diagnostic categories were checked to ensure that an adequate number of cases exist in each. When the prevalence was low in any category, it was either dropped or merged with another category. Based on these frequencies, female infertility was dropped from the model presumably because MediCal does not cover this condition resulting in 89 subcategories (diagnostic variables) that were used to estimate the Medicaid RxCost Model. (Table 5.3)

The Medicaid RxCost Model was empirically derived using dummy variables (1 / 0) to indicate presence or absence of a diagnostic category. Both inpatient and outpatient claims (ICD-9CM codes) were utilized for diagnostic information. Two separate methods were initially utilized to capture diagnostic information within the Medicaid RxCost models. One method, the full method, identifies and enters into the model all diagnostic categories for which there was a corresponding ICD-9-CM code present regardless if there were other similar diagnostic categories coded for each individual. The other method utilizes a hierarchical approach for certain variables where some categories are grouped together in clusters and only the most costly category within each cluster is entered into the model. In order to use the hierarchical method, relevant categories must first be grouped into clusters. A separate data set, MEDSTAT MarketScan data 1998 through 2000, was utilized to identify variables that would be appropriate to cluster in a hierarchical fashion. (MEDSTAT Inc. 2001) This data set contains commercial claims data of over 1.3 million lives. The MarketScan sample was

utilized to identify variables with a Pearson Correlation ≥ 0.2. After these variables were identified a 2x2 contingency table for each of the correlated variable pairs was output and the phi coefficient was estimated. Based on the contingency results, 16 variables were considered appropriate for a total 6 hierarchical clusters. This method was utilized because it identifies variables that are often coded together for the same person. These 6 clusters each contain variables that are often coded for the same condition. When a recipient has more than one variable coded within a cluster only the most costly variable should be captured by the model not multiple variables for the same condition to minimize code proliferation and gaming. For example, if a patient is coded for diabetes mellitus without complications and diabetes mellitus with other complications only the most costly of these variables, diabetes with complications, should be captured by the model. To determine the most costly variable for each cluster a random 20% sample of the Medical training data was utilized to calculate the average annual prescription cost associated with each of the 16 variables. Based on these results the variables within each cluster were assigned a hierarchical classification. Utilizing the hierarchical approach, only the single-highest category within each hierarchical cluster will be captured by the model.

This hierarchical approach was initially adopted to prevent code proliferation and "gaming" of the risk assessment system where clinicians may have incentive to code more conditions which may be justified but do not substantially add to the prescription costs of those persons. This hierarchical method of counting has been shown to simplify the model, strengthen its resistance to additional coding, and produce only small decreases in the accuracy. (Kronick 2000)

Because only 16 variables were affected and only 6 clusters were formed using the hierarchical counting scheme, these 6 hierarchical clusters were incorporated into the full method. The 6 clusters chosen are potential areas for gaming an incentive-based system. Additionally, the minor changes should result in very little loss of predictive power. However, to ensure this was the case, the non-hierarchical full method was later run in addition to the method incorporating the hierarchical clusters for comparison. In both populations the R-square values remained the same.

Diagnosis-based models have some shortcomings that are evident regardless of counting method. These models rely on administrative claims-based data that are not always complete. Diagnosis codes often suffer from left censoring where an individual may be treated for a chronic condition but only diagnosed once for the condition or diagnosed sporadically over long periods of time. Also, diagnostic coding is often an uncertain practice where many ill-defined conditions may be coded in numerous ways. Adding drug information to a diagnosis-based model can attempt to alleviate some of the problems with these diagnosis-based models. One advantage to utilizing drug information is that the data is much more complete than diagnostic information. Secondly, drug information can help identify individuals who are being treated for a chronic condition but who are not diagnosed at each physician-patient encounter. Another advantage is that drug information can potentially help explain severity. The number of classes of medication prescribed to treat a condition can sometimes be used to indicate severity. The disadvantages to utilizing drug information to predict drug expenditures is that a certain degree of endogeneity exists because prescription fills lead directly to greater prescription cost. However, endogeneity is less of a concern

using models prospectively because drug costs in one period are only indirectly related to drug costs in a previous period. (Clark 1995) Another disadvantage of using drug information occurs when the risk assessment models are used for shifting some of the risk to providers (i.e. financial incentives). Here, there will be some incentive for providers to prescribe medications more liberally, particularly for medications that can potentially be prescribed for conditions with varying levels of severity or even prevention.

An alternative risk assessment model, the Medicaid MRxCost Model, based on both diagnostic information and drug information was developed. This “Mixed Model” was utilized to ascertain how much predictive ability can be gained by adding drug information to the Medicaid RxCost Model explained above. The Medicaid RxCost Model was supplemented with drug information based on Gilmer's updated version of Clark's revised Chronic Disease Score to develop the Medicaid MRxCost Model. (Gilmer 2001; Clark 1995)

Virtually all of the published literature utilizing prescription drugs to predict costs has been based on the Von Korff's Chronic Disease Score. (Fishman 2003; Gilmer 2001; Fishman 1999; Lamers 1999; Clark 1995; Johnson 1994; Von Korff 1992) Gilmer's Medicaid Rx model provides one of the latest approaches to updating the CDS and utilizing prescription drugs to model overall healthcare expenditures. (Gilmer 2001) The Medicaid Rx categories were utilized as a starting point; however, these categories were not created specifically for prescription expenditures and, therefore, do not necessarily reflect conditions with similar prescription expenditures.

The Multum Classification system (Multum 2001), publicly available at www.multum.com, was utilized both to refine Gilmer's Medicaid Rx categories and to classify drug claims. Multum updates their database monthly and as of February 2002 there were 284 therapeutic classes. However, the Multum class numbers are not entirely sequential due to former classes being dropped or split out into multiple classes.

The Clinical panel reviewed Gilmer's Medicaid Rx model categories and revised the medication classes to reflect a proxy of severity and comorbidity that would consistently reflect an individual's expected outpatient prescription expenditure. Next, the prevalence of the drug categories in each population were checked to ensure that an adequate number of cases exist in each. When the prevalence was low in any category, it was either dropped or merged with another category. The new drug classification system deleted 6 original categories, combined Tuberculosis with PCP pneumonia and added numerous drugs to the drug list of some categories (Table 5.4). Two additional categories, Alzheimer's disease and Iron deficiency were dropped from the Medicaid Medicaid MRxCost Model due to MediCal formulary restrictions.

The Medicaid MRxCost Model was empirically derived using dummy variables (1 / 0) to indicate presence or absence of a diagnostic category or drug category. Here, the presence of a drug in each category will be counted rather than number of drugs in each category. However, three drug categories were identified by the clinical panel as categories that could potentially help explain severity based on the number of drug classes prescribed. These categories are asthma, cardiovascular, and seizure disorders. Here, the dummy variables for each class were not included in the model but were summed in a single additive variable. Both asthma and seizure disorders

categories have 6 classes of medication that could potentially be prescribed. A variable was created for each and was coded as 0 through 6 depending upon how many drug classes are prescribed. The Cardiovascular category has 8 classes that could potentially be prescribed. Here a variable was created in which a value of 0 to 8 was assigned. For example, if a patient filled medication that is classified into Multum categories 131, 180 and 243 then he/she would be assigned a 3 for their asthma additive variable.

The new drug classification system, that was used to complement the Medicaid RxCost Model, contains some categories that are very similar or identical to the Medicaid RxCost Model. Although similar (and collinear), these variables will be helpful to identify patients with conditions who did not receive a diagnosis within the timeframe the model is run but filled a prescription for the condition. These variables were combined so that either a diagnosis or a drug would indicate the presence of a condition. Some of these variables, such as asthma, are straightforward. Here both the diagnostic portion and the drug portion of the mixed model contain a category for asthma. However, other categories are more complicated. The diagnostic classification system uses two categories for diabetes while the drug classification system uses one. Other categories are concise in one classification and broad in the other. Drug categories that were empirically correlated were inspected and combined in a manner previously. Fifteen drug categories were chosen and combined with their diagnostic counterparts. These categories and the logic used are presented in Table 5.5. The final diagnostic and Rx drug classification system used 106 variables to estimate the Medicaid MRxCost Model. (Table 5.6) Because dummy variables are used in both

the diagnostic portion and the drug portion of the Medicaid MRxCost Model, the presence of a drug in a drug category that was combined with a diagnostic category will only add to the model when there is no presence of a diagnosis. For example, if a patient is diagnosed with asthma the presence of a drug in the asthma category (patient filled an asthma drug) will not affect the model because these categories were combined and the diagnosis is already present.

Analysis

A minimum of 13 months of data and up to 24 months was analyzed for each enrollee where the first twelve months served as the index year and was used to collect ICD-9CM and drug class information. Information gathered from the first 12 months was used to predict pharmacy expenditures for months 13 through 24. Where possible, the most recent twenty-four months of continuously eligible data was used for each enrollee. When twenty-four months of continuously eligible data was not available, the longest span of continuously eligible data was utilized. Here, the first 12 months of data was used as the index year and the remaining months were used for the cost year. Prescription costs for individuals with partial second-year data was annualized using the method described in Ash et al. (Ash 1989)

$$\text{Total Rx Expenditure} = \text{Rx Cost} \times (12 \div \text{Months Eligible})$$

Both inpatient and outpatient claims (ICD-9CM codes) were utilized for diagnostic information and pharmacy claims (NDC codes) were utilized for drug information.

The models were initially estimated using Ordinary Least Squares regression (OLS) and a Two-Part Model (2PM). OLS is the most common method for analyzing health utilization data and have been shown to work just as well as other more

complicated models for predicting future costs (Diehr 1999). However, health care utilization data typically consists of a high proportion of persons with zero expenditure and a small number of individuals with very high expenditures. Because of this, a 2PM based on Duan et al. (Duan 1983) was used as an alternative to OLS. A two-part regression model was chosen because this type of model attempts to correct the problems associated with non-spenders by first using a logit equation for the dichotomous event of having zero or positive pharmacy expenses. The next part of the model is a regression equation that is conditional on having positive pharmacy expenses and is used to model the level of positive expenses. A log-linear regression model was utilized for the second part of the model. A "smearing" estimator was used to retransform the predicted values back to the mean of the original distribution. (Duan 1983)

Both models utilized a stepwise selection process where a variable was allowed to enter at a significance level of 0.20 and remain in the model at 0.10 was utilized. (Mantel 1970) The age-sex variables were considered essential variables and were included in all models regardless of their significance to minimize model mispecification errors and to ensure every member will have a predicted cost for the following year. The hierarchical variables in the Medicaid RxCost Model and Medicaid MRxCost Model were grouped into clusters for the stepwise procedure so that they either all remained in the model or all fell out of the model.

Variables included in the models are as follows:

The *Demographic Model* was estimated with:

» Intercept

» dummy variables for age-sex classification (1 = year; 0 = absence)

The *Demographic and Medicaid Eligibility Model* was estimated with:

» same as demographic model

» dummy variable for Medicaid SSI eligibility (1 = yes; 0 = no)

The *Medicaid RxCost Model* was estimated with:

» same as demographic model Medicaid eligibility model

» dummy variables for the presence or absence of a given condition in the revised diagnostic classification system
(1 = diagnosis present in a clinical classification; 0 = absence)

The *Medicaid MRxCost Model* (both diagnostic and drug info) was estimated with:

» same as Medicaid RxCost Model

» dummy variable for the presence of absence of a drug from a given therapeutic class in the revised drug classification system
(1 = drug present in therapeutic class; 0 = absence)

» additive variable for the 3 conditions identified as indicating severity based on the number of medication classes prescribed

Model Derivation

All model development was carried out using STATA Intercooled Version 6.0. (STATACORP 1999) The models estimated using OLS were compared to the models estimated using the two-part modeling technique. Based on this comparison the OLS model was chosen to estimate all models.

Even though the two-part model seems to fit the distribution better, OLS performed slightly better when predicting future prescription costs. The Root Mean Squared Error (RMSE) of each technique was compared to assess model performance (RMSE Medicaid RxCost Model: OLS = 2,003 2PM = 2,260). This is consistent with other work using models to predict cost. (Diehr 1999) One problem encountered with OLS was that it estimated a negative intercept for the Demographic and Medicaid Eligibility model, the Medicaid RxCost Model and Medicaid MRxCost Model. With a negative intercept, the reference case (male 61 to 64) with no diagnostic information present would have a negative cost prediction for the following year which would be unacceptable. To circumvent this problem the intercept was constrained to zero and the age-sex variables were constrained to their mean prediction value when diagnostic information was absent. More specifically, members with no diagnostic information classified (all diagnostic variables were coded 0) were used to calculate the mean prediction value of each age-sex variable on following year's prescription cost. These values were then used to constrain the age-sex variables for everyone. This allowed OLS to predict only positive costs and resulted in no loss in predictive power in terms of R-square. OLS was chosen to estimate all models.

Two main tactics were used to guard against overfitting the models. First, the stepwise variable selection (discussed above) was employed to either keep or drop variables from the model based on their significance to the model. Secondly, the number of candidate variables was limited to a ratio of at least 10 cases to each predictor variable. (Harrell 1996) The use of large database kept data reduction techniques from being employed except in a three cases. Female infertility was

dropped from the Medicaid RxCost Model due to low prevalence presumably because MediCal did not cover female infertility and Alzheimer's and iron deficiency categories were dropped from the Medicaid MRxCost Model because of formulary restrictions within MediCal.

Model Validation

All models were frozen and tested on the validation sample. Model discrimination, the amount of variance explained by each model, was evaluated by calculating the model's R-square value.

All models have been trained using all costs with no removal or trimming techniques utilized. The frozen model estimates were also applied to the validation sample after Year 2 prescription costs were trimmed at $20K. Here, any individual's prescription drug cost of over $20K was set equal to $20K. The value of $20K was chosen because it trims only the most sever outliers that could have a substantial effect on the models. The trimming procedure only affects 0.11% (100 observations) of the MediCal validation sample.

Additionally, the Medicaid RxCost Model was also run in a concurrent fashion to see how well the frozen coefficients predict year 1 prescription expenditures. The Medicaid MRxCost Model was not run using the concurrent approach because endogeneity was deemed to be too great. Here, the model would draw upon prescription drug information to predict prescription drug expenditures.

Model Performance

The one universally reported, single-number summary performance measure for risk-assessment payment models is the R-square value; the proportion of variance in

costs that the model explains. (Ash 2000) Model R-square values are reported for each of the models. Additionally, R-square values were also adjusted for shrinkage, the flattening of the plot of predicted versus observed away from the line, caused by overfitting, using the heuristic shrinkage estimator of van Howelingen (van Houwelingen 1990)

$$R^2_{adj} = 1 - (1- R^2) \times (n-1) / (n-p-1)$$ where n = the number of subjects and p = the number of candidate variables

Both R^2 and R^2_{adj} values are reported for each of the training sample models and each of the validation sample models.

The R-square value described above assesses the amount of variance explained by the model for each individual member applied to the model. These models are designed to be utilized in a physician or physician group setting where a group of patients (e.g. 10, 50 or 100 patients) could be modeled to predict a single prescribing expenditure for the entire groups of patients. When predicting a prescribing expenditure for a physician or physician group based on their patients the amount of variance explained for individual patients is less important than the amount of variance explained for the entire group. Here a pooled R-square value will help assess the amount of variance the model explains for a group (or pool) of patients. Using these models in a physician or physician group setting would require the model to perform well for varying numbers of patients. To assess how well the model performs for varying numbers of patients, pooled R-square values were calculated based on hypothetical physician groups of 10, 20, 50, 100, 200, 300, 400, and 500 patients utilizing the validation sample. To calculate the pooled R-square value for groups of 10 patients, the groups were randomly selected without replacement and each of the patients actual year 2

prescription costs are summed into one group value. Additionally, each of their predicted cost was summed into one value. Next, each group was then counted as 1 observation and an R-square value was calculated for the entire sample based on the groups of 10 patients. This procedure was replicated for groups of 10, 20, 50, 100, 200, 300, 400 and 500 patients. The ultimate goal here was to assess how many patients a physician or physician group needed to have before the model could be used to predict prescription costs effectively.

Another measure of model performance is a predictive ratio. (Ash 2000) Here, the model is applied to a subgroup of people and the predictive ratio is calculated by dividing the model-predicted costs for the group by their actual costs. Each model applied to the validation sample was used to predict costs of specific subgroups including, asthma, depression, diabetes, HIV infection and hypertension. These subgroups were identified by ICD-9-CM codes (and NDC codes in the Mixed Model) during year 1. The predicted costs were then utilized to calculate a predictive ratio. The predictive ratio was used to evaluate how well the models perform for the chosen specific subgroups. An ideal predictive ratio of 1.0 would indicate that the predicted costs and actual costs were exactly the same. As with R-square values, pooled predictive ratios were also calculated for the hypothetical physician groups of patients. Here, the random groups identified above in the validation sample were used to calculate a predictive ratio for each group and an average predictive ratio was calculated for groups of 10, 20, 50, 100, 200, 300, 400 and 500 patients. Once again, this was utilized to assess how well the models predict a single prescription expenditure for varying groups of patients.

Although there are no other risk assessment indices publicly reported for predicting pharmacy expenditures there are publicly available approaches for predicting overall health care costs. One such index is Kronick's Chronic Illness and Disability Payment System (CDPS) developed for Medicaid populations is one of the more sophisticated models available. This model was also applied prospectively to the validation sample to predict pharmacy expenditures. The R-square value for this model was then compared to the Medicaid RxCost Model and the Medicaid MRxCost Model for each population. Comparing the models allowed further evaluation of model performance. The CDPS model and the software to implement the model are free of charge and publicly available online at www.medicine.ucsd.edu/fpm/cdps. Because this model was not developed specifically for predicting prescription expenditures, it was recalibrated to estimate prescription costs. Without recalibration the CDPS model would substantially over predict costs. The benchmark weights were used for each model; however the models were recalibrated using a proportional calibration method where the mean of the individual predicted expense is set equal to the mean observed year 1 expense. This method is explained in DxCGs Analytic Manual for their own hierarchical model (DxCG Inc 2001). Employing this method, the model generates costs that are in line with prescription costs instead of overall health costs.

RESULTS

Population Sample

The population description for both the MediCal training and validation samples are presented in Table 5.7. The samples were randomly chosen and both are very similar in terms of age, gender, eligibility and prescription cost. Almost three quarters of

the population for both samples are made up of females and individuals that are eligible for TANF/AFDC.

Models

The results of the Medicaid RxCost Model including the coefficients, their p-values and the 95% confidence intervals are presented in Table 5.8. The R-square value was equal to 0.27 in the training sample and 56 of the possible 101 variables remained in the model after the stepwise selection procedure. The reference group was male 61 to 64 with TANF/AFDC eligibility. Variable costs ranged from $4.48 for Male 18 to 22 up to $9,175.23 for CCS 5.4 (Cystic fibrosis). With the exception of males 61 to 64, costs of the age-sex variables increased with age. Females age 61 to 64 had the highest cost estimate of the age-sex variables. Eight diagnostic variables had cost estimates over one thousand dollars; HIV infection, cystic fibrosis, immunity disorders, coagulation and hemorrhagic disorders, Schizophrenia and related disorders, Parkinson's disease, multiple sclerosis and chronic renal failure. The average predicted cost for all members was 550.42 (95% CI: 542.91, 557.94). The average actual cost was 497.48 (95% CI: 482.09, 512.87). The predictive ratios, the ratio of predicted costs versus actual costs, of the Medicaid RxCost Model for each of the subgroup populations were as follows: hypertension = 1.00 (95% CI: 0.96, 1.04), depression = 0.99 (95% CI: 0.92, 1.06), HIV = 0.99 (95% CI: 0.91, 1.06), diabetes = 0.89 (95% CI: 0.85, 0.92) and asthma = 1.47 (95% CI: 1.29, 1.66). The overall predicted ratio for all members = 1.11 (95% CI: 1.10, 1.12).

The Medicaid MRxCost Model results from the training sample are presented in Table 5.9. The R-square value = 0.33 and 58 of the possible 118 variables remained in

the model after the stepwise selection procedure. Only 1 of the 17 prescription-based variables, Rx 25 (pain), was dropped from due to the selection procedure. The reference group was male 61 to 64 with TANF/AFDC eligibility. Variable costs ranged from $3.70 for Male 18 to 22 up to $9,523.26 for CCS 5.4 (Cystic fibrosis). Once again, with the exception of males 61 to 64, costs of the age-sex variables increased with age and females age 61 to 64 had the highest cost estimate. Six diagnostic variables and six prescription variables had cost estimates over one thousand dollars; HIV infection, cystic fibrosis, immunity disorders, coagulation and hemorrhagic disorders, Schizophrenia and related disorders, Parkinson's disease, multiple sclerosis and chronic renal failure, anticoagulants, ESRD/renal disorders, herpes, liver disease, tuberculosis and PCP pneumonia. The average predicted cost for all members was 524.01 (95% CI: 516.07, 531.94). The average actual cost was 497.48 (95% CI: 482.09, 512.87). The predictive ratios of the Medicaid MRxCost Model for each of the subgroup populations were as follows: depression = 0.98 (95% CI: 0.95, 1.00), diabetes = 0.98 (95% CI: 0.94, 1.02), HIV = 0.90 (95% CI: 0.79, 1.00), hypertension = 1.28 (95% CI: 1.21, 1.36) and asthma = 2.01 (95% CI: 1.33, 2.69). The overall predicted ratio for all members = 1.05 (95% CI: 1.03, 1.07).

The R-square values adjusted for shrinkage did not differ from the original R-square values in any of the models. All of the model R-square values are presented in Table 5.10. The Demographics Model with Medicaid Eligibility provided an R-square equal to 0.10 in the training sample and 0.10 in the validation sample, over thee times that of the Demographics Model. This serves as a baseline for what can be achieved with minimal effort.

The R-square value for Medicaid RxCost Model decreased from 0.27, as estimated using the training sample, to 0.24 when applied to the validation sample. The R-square value was six times higher for the Medicaid RxCost Model as compared to the CDPS. When run concurrently on the validation sample the R-square value of the Medicaid RxCost Model doubled as compared to the prospective model. The Medicaid MRxCost Model yielded a higher R-square value than the Medicaid RxCost Model in both the training and validation samples.

The pooled R-square values and average predicted ratios for each model applied to various groups of patients are presented in Table 5.11. The R-square values are fairly high even with groups of 10 members (0.24) but increase as the group size increases.

Trimming the outliers in the validation sample proved to be especially useful. Here, only 100 (0.11%) observations were affected by the trimming technique; however the R-square value more than doubled for both prospective models. The highest R-square value for any model was achieved by the Medicaid MRxCost Model when applied to the trimmed validation sample with an R-square value equal to 0.63.

DISCUSSION

The objective of this study was to empirically develop a claims-based risk assessment model to predict prescription expenditures for a Medicaid population. Coupling the recent trend towards using incentives to drive physician behavior with the ever-growing attention to prescription drug expenditures, this model could potentially be used to set prescribing goals for physicians based on their patient-mix. Also, due to the states familiarity with risk assessment this type of model may be particularly attractive.

Currently, there are no other published models developed specifically for predicting prescription expenditures. The primary model of interest is the Medicaid RxCost Model which uses diagnostic claims-based information to predict prescription expenditures. Additionally, the Medicaid MRxCost Model was developed that supplements the Medicaid RxCost Model with prescription drug information.

The Medicaid RxCost Model seemed to perform well when estimated on the training sample (in terms of R-square) as compared to other risk assessment models used to predict overall health costs. When applied to the validation sample the R-square value decreased only slightly. A decrease in discrimination is typical but only dropping slightly shows that the model suffers little from overfitting and that it is potentially generalizable to other Medicaid populations. The Medicaid RxCost Model outperformed the CDPS model by a magnitude of six in terms of R-square value. Even after re-calibrating the model to predict prescription expenditures, the CDPS model performed quite poorly; about what you would expect from a demographics model. However, the CDPS model was not developed specifically to predict prescription expenditures and even after re-calibration the Medicaid RxCost Model is expected to have higher discriminatory power. While no model is available to directly compare in terms of prescription cost, this comparison was striking and provides a benchmark of how much predictive power can be gained over one of the more utilized models to date in regards to predicting prescription expenditures.

The predictive power of the Medicaid RxCost Model was also tested for subgroups of the population with select disease states including asthma, depression, diabetes, HIV, and hypertension. Here, the predictive ratio served as a marker for

predictive power. The Medicaid RxCost Model performed quite well for depression, diabetes, HIV and hypertension, only deviating at most .11 from 1.0. The model did not fair quite as well with the asthma population; it predicted approximately 47% as much cost as was actually incurred.

To explore the potential use of the Medicaid RxCost Model as a method for profiling physicians it was also run concurrently on the validation sample. Here year 1 diagnostic information was utilized to predict year 1 costs. The model performed extremely well in comparison to the prospective model as the R-square value doubled. With an R-square value of 0.48, the model could potentially be very useful in profiling physicians to examine current prescribing behaviors.

In addition to a diagnostic model, an alternative model was developed, the Medicaid MRxCost Model, to explore how much predictive power could be gained by supplementing the Medicaid RxCost Model with drug information. Drug information proved to be quite powerful as the R-square values increased substantially for both the training and validation samples as compared to the Medicaid RxCost Model. The predictive ratios for this model were very similar for depression, diabetes and HIV as compared to the Medicaid RxCost Model. The predictive ratios for asthma and hypertension were worse for the Medicaid MRxCost Model with asthma predicting twice as much expense as was incurred.

Viewing the pooled R-square value and predicted ratio for varying group sizes allows one to evaluate the performance of the models when utilized with different group sizes. Here, the models performed quite well in terms of both R-square and predicted ratio for a group size as small as 10 members. The R-square values for both models

steadily improved as the group size increased. The Medicaid MRxCost Model slightly outperformed the Medicaid RxCost Model for groups of 10 and 20 members however the Medicaid RxCost Model performed equally as well with a group size of 50 members and outperformed the Medicaid MRxCost model for groups of 100 members or more. The Medicaid MRxCost model did outperform the Medicaid RxCost Model in terms of predicted ratio across all group sizes, however neither model showed a benefit beyond a group size of 50 members. These pooled values indicate that the models could potentially be utilized to a predict prescription expenditure for a physician or physician group with a patient base as small as 10 patients. In fact, the predicted ratio did not increase beyond 50 members. However, the pooled R-square values do indicate gains in prediction power as the group sizes increase.

Despite gains in R-square values, the advantages of supplementing the diagnostic model with drug information may be tempered with some potential disadvantages. One advantage of drug information is that it can be more timely, reliable and complete than diagnostic information. This model identifies numerous individuals with conditions that were not diagnosed within the timeframe of the study period. Additionally, drug information helped explain severity in three conditions, asthma, cardiovascular, and seizure disorders. When supplemented with drug information, the diagnostic model performed much better in terms of discrimination. However, this model is associated with a certain amount of endogeneity as it uses drug information in a prior period which may be influenced by relative over or under prescribing to predict subsequent drug costs. To minimize the problem of endogeneity, the model only used binary variable to indicate whether a drug class was prescribed and did not count the

number of prescriptions prescribed within a predefined class. Also, if this model was used to influence prescribing behavior, risk assessment models supplemented with drug information could potentially reward physicians for prescribing medication that would lead to the appearance of a more expensive patient-mix. This type of "gaming" happens with diagnostic model as well but there is no real utilization being consumed as there is with prescribing a medication. We did, however, attempt to attend to this problem by utilizing hierarchies and combining drug categories with diagnostic categories when plausible. Regardless of the situation, the advantages and disadvantages of this type of mixed model should be carefully considered if this model were to be used to influence physician compensation.

The Heuristic shrinkage estimator was utilized to adjust the R-square value of each model. This estimator tries to account for overfitting of the model by taking into consideration the number of candidate variables utilized as compared to the number of observations. Because the ratio of candidate variables to observations was extremely small the estimator did not change the R-square value of any model estimated.

The effect of outliers on the models was also of interest. We wanted to see the effect of trimming (capping) year 2 prescriptions costs at $20K. Here, the extreme values were not dropped from the model completely; only capped at a high level. This technique only affected 100 (0.11%) of the observations but had a substantial effect on R-square values as they more than doubled for both prospective models. This may be particularly useful if some states or plans employ re-insurance or stop-loss protections. These plans would be able to utilize this approach to even further increase the predictive power.

Overall, the R-square values obtained with the Medicaid RxCost Model and the Medicaid MRxCost Model were relatively high. Most general health risk assessment models produce R-square values at or below 0.18. However, prescription drug costs are, on average, less costly and have less variance than do general health care costs. For these reasons, the obtainable R-square values will be larger especially for models specifically derived to predict prescription drug costs. Despite this, the Medicaid RxCost model substantially outperformed the CDPS model even after recalibrating to predict prescription drug costs.

Limitations

No risk assessment model will ever fully adjust for all differences in patient populations, nor will any model perfectly predict costs. The problem with trying to predict future costs is that random variation is so prevalent however, the prescription cost risk assessment models should performed much better than the traditional age-and-sex method. The use of ICD-9-CM diagnosis and NDC drug codes presents another limitation. The use of this information has been found to be reliable but does not compare to the accuracy of clinical patient-specific data. Overall, there has been a lot of concern with accuracy, unreliability and clinical specificity of diagnostic information. (Iezzoni 1997; Romano 1994; Hannan 1992) Diagnosis-based indices rely solely on administrative claims-based coding of physician-patient encounters. Although ICD-9-CM diagnostic coding is not complete, the use of claims-based diagnostic coding in the past by both government and commercial programs for billing has resulted in substantial gains in completeness of data. A disadvantage, however, is that the completeness of data depends on the manner in which physicians are paid. If

physicians are paid based on a discounted fee-for-service system, the data will be more complete than if physicians are paid on a sub-capitated system or fixed annual salary. However, the negative effect of reimbursement based on diagnosis is that physician now have incentives to code more prolifically and to code ill-defined conditions in a manner to increase reimbursement. The hierarchical method utilized attempted to account for some of the potential up-coding. Finally, another disadvantage to using diagnosis-based models is that there can be a substantial lag time between when a patient receives services and when the diagnostic information is ready for use. Often a 6-month lag time is necessary to have reasonably complete diagnostic information available from health care encounters.

Another limitation is that the risk assessment weights proposed in this research may not be appropriate for each health plan. Due to varying formularies and benefit design structures the weights may need to be re-estimated or at least re-calibrated to perform well in a plan setting. Because risk assessment is based on averages and future cost uncertainty will always be high. A goal must be set to allow for some deviation from the predicted value. This is particularly important for any risk adjustment model used to adjust or otherwise profile small groups of patients or physician groups with relatively small practices.

CONCLUSION

In summary, the Medicaid RxCost Model seems to do a nice job of predicting prescription expenditures both prospectively and concurrently. This model may serve well both setting prescribing goals for the coming year as well as profiling physician habits the previous year. The Medicaid MRxCost Model also performed well when

utilized prospectively, however this model should be weighed against the greater potential for gaming and rewarding practices for previously high prescription use.

REFERENCES

American Federation of State, County and M Employees. Prescription Drugs and State Budgets. Health Focus 2003; www.afscme.org/publications/healht_focus/focus302.htm.

Agency for Health Care Policy and Research. Department of Health and Human Services. www.ahrq.gov

Ash A, Porell F, Gruenberg L, Sawitz E, Beiser A. Adjusting Medicare capitation payments using prior hospitalization data. Health Care Financ Rev 1989; 10(4):17-29.

Ash AS, Ellis RP, Pope GC, Ayanian JZ, Bates DW, Burstin H et al. Using diagnoses to describe populations and predict costs. Health Care Financ Rev 2000; 21(3):7-28.

Bridges to Excellence. Diabetes Care Link for Physicians and Providers. 2004; www.bridgestoexcellence.com/bte/diabetescarelink/gty_physicians.htm.

Burton SL, Randel L, Titlow K, Emanuel EJ. The ethics of Pharmaceutical benefit management. Health Aff (Millwood) 2001; 20(5):150-163.

The California Department of Health Services. 2001; www.dhs.ca.gov.

Carroll J. Physicians Reconsider Taking on Pharmacy Risk. Managed Care July 2000.

Clark DO, Von Korff M, Saunders K, Baluch WM, Simon GE. A chronic disease score with empirically derived weights. Med Care 1995; 33(8):783-795.

Clinical Classifications Software (ICD-9-CM) Summary and Download. Summary and Downloading Information. 2001 Agency for Health Care Policy and Research, Rockville, MD. http://www.ahrq.gov/data/hcup/ccs.htm

Diehr P, Yanez D, Ash A, Hornbrook M, Lin DY. Methods for Analyzing Health Care Utilization and Costs. Annual Review of Public Health 1999; 20:25-144.

Duan N, Manning WG, Morris CN, Newhouse JP. A Comparison of Alternative Models for the Demand of Medical Care. Journal of Business and Economic Statistics 1983; 1(2):115-126.

DxCG Inc. DxCG Risk Adjustment Software: Analytic Guide Release 6.0. 2001; Boston MA. support@dxcg.com

Employee Benefit Research Institute. Prescription Drug Costs Up Sharply - but Still Small Overall. Press Release 470 Washington EBRI . 1999.

Ettner SL, Frank RG, McGuire TG, Newhouse JP, Notman EH. Risk Adjustment of Mental Health and Substance Abuse Payment. Inquiry 1998; 35: 223-239.

Fishman PA, Shay DK. Development and estimation of a pediatric chronic disease score using automated pharmacy data. Medical Care 1999; 37:874-883.

Fishman PA, Goodman MJ, Hornbrook MC, Meenan RT, Bachman DJ, O'Keefe Rosetti MC. Risk Adjustment Usng Automated Ambulatory Pharmacy Data: The Risk Model. Medical Care 2003; 41(1): 84-99.

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

Graden SE, Schafermeyer KW. Performance Reporting for Managed Care Prescription Programs. Journal of Managed Care Pharmacy 1998; 4(2): 160-166.

Hannan EL, Kilburn H, Lindsey M, et al. Clinical versus administrative databases for CABG surgery: does it matter? Medical Care 1992; 30:892-907.

Iezzoni L, Ash AS, Daley J, Hughes JS, Shwartz M. Risk adjustment for measuring healthcare outcomes. 2nd ed. Chicago: Health Administration Press, 1997.

IMS Health I. Integrated Promotional Services (Office Promotion Reports, Hospital Promotion Reports, National Journal Audit) and Integrated Share of Voice Report (with data from Competitive Media Reporting). 1999.

Johnson RE, Hornbrok MC, Nichols GA. Replicating the chronic disease score (CDS) from automated pharmacy data. Journal of Clinical Epidemiology 1994; 47(10):1191-1199.

Kaelin JJ. Risk Adjustment for State Medicaid Programs: Lessons Learned and Prospects for the Future. Center for Health Program Development and Management April 2002; wwwacg.jhsph.edu/library/conference_2002/d2.plenary%203.%20kaelin.ppt

The Kaiser Family Foundation. Prescription Drug Trends. May 2003; www.kff.org/rxdrugs/3057-03-index.cfm

The Kaiser Family Foundation. Prescription Drug Trends. September 2000; www.kff.org

The Kaiser Family Foundation. Prescription Drugs: Facts at a Glance. 2003; www.kaisernetwork.org/static/spotlight_rxdrugs_facts.cfm.

The Kaiser Family Foundation. Medicaid and the Prescription Drug Benefit: Cost Containment Strategies and State Experiences. September 2002; www.kff.org/medicaid/4063.index.cfm

Kronick R, Gilmer T, Dreyfus T, Lee L. Improving health-based payment for medicaid beneficiaries: CDPS. Health Care Financ Rev 2000; 21(3):29-64.

Lamers LM. Pharmacy cost groups: a risk-adjuster for capitation payments based on the use of prescribed drugs. Medical Care 1999; 37(8):824-830.

Mantel N. Why Stepdown procedures in variable Selection. Technometrics 1970; 12:621-625.

Mullahy J. Much Ado About Two: Reconsidering Retransformation and the Two-Part Model in Health Econometrics. Journal of Health Economics 1998; 17:247-281.

The Multum Lexicon. Multum Information Systems Inc., editor. 2001; www.multum.com.

National Committee for Quality Assurance. Integrated Healthcare Association Pay for Performance Program: 2004 Clinical Measure Specifications and Audit Review. December 2003; www.iha.org

Romano PS, Roos LL, Luft HS, et al. A comparison of administrative versus clinical data: coronary artery bypass surgery as an example. Journal of Clinical Epidemiology 1994; 47:249-260.

SAS Institute Inc. 2004; www.sas.com
StataCorp. Stata Statistical Software: Release 6.0. 1999 College Station Texas

United physicians incentive program reduces pharmacy costs. PR Newswire. April 17, 2003.

van Houwelingen JC, le Cessie S. Predictive value of statistical models. Statistics in Medicine 1990; 8:1303-1325.

Von Korff M, Wagner EH, Saunders K. A chronic disease score from automated pharmacy data. J Clin Epidemiol 1992; 45(2):197-203.

Wray NP, Hollingsworth JC, Peterson NJ, Asthon CM. Case-mix adjustment using administrative databases: a paradigm to guide future research. Medical Care Res Rev1997; 54(3): 326-256.

Table 5.1

Data Counts Resulting from MediCal Eligibility Requirements	
Total unduplicated count	1,689,019
Count after deleting members with Medicare eligibility	1,474,660
Count after continuous eligibility requirements	874,102
Count after deleting members not 18 to 62 years of age	305,877
Count after deleting members not eligible for AFDC/TANF or SSI/Disability	280,474
Count after deleting members admitted to institutions or with inpatient stays > 30 days	276,518

Table 5.2

New CCS-Based Classification
1. Infectious and parasitic diseases
1.1 Bacterial infection
1.1.1 Tuberculosis
1.1.2 Septicemia (except in labor)
1.1.3 Sexually transmitted infections (not HIV or hepatitis)
1.1.4 Other bacterial infections
1.2 Mycoses other than candidiasis
1.3 Hepatitis
2. HIV Infection
3. Neoplasms
4. Diabetes Mellitus
4.1 Diabetes mellitus without complication
4.2 Diabetes mellitus with other complications
5. Other endocrine, nutritional, and metabolic diseases and immunity disorders
5.1 Thyroid disorders
5.2 Disorders of lipid metabolism
5.3 Gout and other crystal arthropathies
5.4 Cystic fibrosis
5.5 Immunity disorders
5.6 Other nutritional, endocrine, and metabolic disorders
6. Diseases of the blood and blood-forming organs
6.1 Anemia
6.2 Coagulation and hemorrhagic disorders
6.3 Diseases of white blood cells AND other hematologic conditions
7. Mental retardation
8. Senility and organic mental disorders
9. Other mental disorders
9.1 Affective disorders
9.2 Schizophrenia and related disorders and other psychoses
9.3 Anxiety, somatoform, dissociative, and personality disorders
9.4 Other mental conditions
10. Alcohol and substance-related mental disorders
11. Eye Disorders
11.1 Glaucoma
11.2 Other eye disorders
12. Ear conditions
12.1 Otitis media and related conditions
12.2 Conditions associated with dizziness or vertigo AND other sense organ disorders

13. Diseases of the Central nervous system and other sense organs
13.1 Central nervous system infection
13.2 Hereditary and degenerative nervous system conditions
13.2.1 Parkinson's disease
13.2.2 Multiple sclerosis
13.3 Paralysis
13.4 Epilepsy, convulsions
13.5 Migraine
14. Hypertension
15. Heart valve disorders
16. Acute myocardial infarction
17. Coronary atherosclerosis and other heart disease
18. Congestive heart failure (nonhypertensive) and pulmonary heart disease
19. Cardiac dysrhythmias
20. Cerebrovascular disease
21. Diseases of arteries, arterioles, capillaries, veins and lymphatics
21.1 Peripheral and visceral atherosclerosis
21.2 Hypotention and Other unspecified circulatory disease
21.3 Phlebitis, thrombophlebitis and thromboembolism AND Varicose veins of lower extremity
22. Other diseases of the circulatory system
22.1 Peri-, endo-, and myocarditis, cardiomyopathy (except that caused by TB or STD)
22.2 Conduction disorders
23. Diseases of the respiratory system
23.1 Respiratory infections
23.1.1 Pneumonia (except that caused by TB or STD)
23.1.2 Influenza, Acute and chronic tonsillitis, Acute bronchitis AND other URIs
23.2 Chronic obstructive pulmonary disease and bronchiectasis
23.3 Asthma
23.4 Lung disease due to external agents
23.5 Other lower respiratory disease
23.6 Other upper respiratory disease
24. Diseases of the digestive system
24.1 Intestinal infection
24.2 Upper gastrointestinal disorders
24.2.1 Esophageal disorders
24.2.2 Gastroduodenal ulcer (except hemorrhage) AND Gastritis and duodenitis
24.2.3 Other disorders of stomach and duodenum
24.3 Lower gastrointestinal disorders AND Biliary tract disease
24.4 Pancreatic disorders (not diabetes)
24.5 Constipation, Dysphagia, and Other unspecified GI disorders
25. Liver disease

26. Diseases of the urinary system

26.1 Nephritis, nephrosis, renal sclerosis

26.2 Chronic renal failure

26.3 Urinary tract infections, Calculus of the urinary tract AND Other diseases of the kidney and ureters

27. Diseases of the genitourinary system

27.1 Diseases of male genital organs

27.2 Diseases of female genital organs

27.2.1 Nonmalignant breast conditions, Inflammatory diseases of female pelvic organs, Endometriosis, AND Prolapse of female genital organs

27.2.2 Menstrual disorders AND Menopausal disorders

27.2.3 Female infertility

28. Complications of pregnancy, childbirth, and the puerperium

28.1 Contraceptive and procreative management

28.2 Complications mainly related to pregnancy

28.2.1 Hypertension complicating pregnancy, childbirth and the puerperium

28.2.2 Diabetes or abn. glucose tolerance complicating pregn., childbirth, or the puerperium

29. Diseases of the skin and subcutaneous tissue

29.1 Skin and subcutaneous tissue infections AND Other inflammatory conditions of skin

29.2 Chronic ulcer of skin

30. Diseases of the musculoskeletal system and connective tissue

30.1 Infective arthritis and osteomyelitis (except that caused by TB or STD)

30.2 Non-traumatic joint disorders

30.3 Spondylosis, intervertebral disc disorders, other back problems

30.4 Osteoporosis AND Pathological fracture

30.5 Systemic lupus erythematosus and connective tissue disorders, Other connective tissure disease AND Other bone disease and musculoskeletal deformities

31. Congenital anomalies

31.1 Cardiac and circulatory congenital anomalies

31.2 Nervous system congenital anomalies

32. Certain conditions originating in the perinatal period

33. Injury and poisoning

33.1 Spinal cord injury

33.2 Intracranial injury

33.3 Crushing injury or internal injury

33.4 Open wounds

33.5 Burns

33.6 Complications of surgical procedures or medical care

34. Symptoms, signs, and ill-defined conditions and factors influencing health status

Table 5.3

Medicaid RxCost Model: All Diagnostic Variables Submitted to OLS Estimation

#	Variable
1	Tuberculosis (CCS 1.1.1)
2	Septicemia (except in labor) (CCS 1.1.2)
3	Sexually transmitted infections (not HIV or hepatitis) (CCS 1.1.3)
4	Other bacterial infections (CCS 1.1.4)
5	Mycoses other than candidiasis (CCS 1.2)
6	Hepatitis (CCS 1.3)
7	HIV Infection (CCS 2)
8	Neoplasms (CCS 3)
9	Diabetes mellitus without complication (CCS 4.1)
10	Diabetes mellitus with other complications (CCS 4.2)
11	Thyroid disorders (CCS 5.1)
12	Disorders of lipid metabolism (CCS 5.2)
13	Gout and other crystal arthropathies (CCS 5.3)
14	Cystic fibrosis (CCS 5.4)
15	Immunity disorders (CCS 5.5)
16	Other nutritional, endocrine, and metabolic disorders (CCS 5.6)
17	Anemia (CCS 6.1)
18	Coagulation and hemorrhagic disorders (CCS 6.2)
19	Diseases of white blood cells AND other hematologic conditions (CCS 6.3)
20	Mental retardation (CCS 7)
21	Senility and organic mental disorders (CCS 8)
22	Affective disorders (CCS 9.1)
23	Schizophrenia and related disorders and other psychoses (CCS 9.2)
24	Anxiety, somatoform, dissociative, and personality disorders (CCS 9.3)
25	Other mental conditions (CCS 9.4)
26	Alcohol and substance-related mental disorders (CCS 10)
27	Glaucoma (CCS 11.1)]
28	Other eye disorders (CCS 11.2)
29	Otitis media and related conditions (CCS 12.1)
30	Conditions associated with dizziness or vertigo AND other sense organ disorders (CCS 12.2)
31	Central nervous system infection (CCS 13.1)
32	Parkinson's disease (CCS 13.2.1)
33	Multiple sclerosis (CCS 13.2.2)
34	Paralysis (CCS 13.3)
35	Epilepsy, convulsions (CCS 13.4)
36	Migraine (CCS 13.5)
37	Hypertension (CCS 14)
38	Heart valve disorders (CCS 15)

39 Acute myocardial infarction (CCS 16)
40 Coronary atherosclerosis and other heart disease (CCS 17)
41 Congestive heart failure (nonhypertensive) and pulmonary heart disease (CCS 18)
42 Cardiac dysrhythmias (CCS 19)
43 Cerebrovascular disease (CCS 20)
44 Peripheral and visceral atherosclerosis (CCS 21.1)
45 Hypotention and Other unspecified circulatory disease (CCS 21.2)
46 Phlebitis, thrombophlebitis and thromboembolism AND
Varicose veins of lower extremity (CCS 21.3)
47 Peri-, endo-, and myocarditis, cardiomyopathy (except that caused by
TB or STD) (CCS 22.1)
48 Conduction disorders (CCS 22.2)
49 Pneumonia (except that caused by TB or STD) (CCS 23.1.1)
50 Influenza, Acute and chronic tonsillitis, Acute bronchitis AND other URIs (CCS 23.1.2)
51 Chronic obstructive pulmonary disease and bronchiectasis (CCS 23.2)
52 Asthma (CCS 23.3)
53 Lung disease due to external agents (CCS 23.4)
54 Other lower respiratory disease (CCS 23.5)
55 Other upper respiratory disease (CCS23.6)
56 Intestinal infection (CCS 24.1)
57 Esophageal disorders (CCS 24.2.1)
58 Gastroduodenal ulcer (except hemorrhage) AND Gastritis and duodenitis (CCS 24.2.2)
59 Other disorders of stomach and duodenum (CCS 24.2.3)
60 Lower gastrointestinal disorders AND Biliary tract disease (CCS 24.3)
61 Pancreatic disorders (not diabetes) (CCS 24.4)
62 Constipation, Dysphagia, and Other unspecified GI disorders (CCS 24.5)
63 Liver disease (CCS 25)
64 Nephritis, nephrosis, renal sclerosis (CCS 26.1)
65 Chronic renal failure (CCS 26.2)
66 Urinary tract infections, Calculus of the urinary tract AND Other disesases of
the kidney and ureters (CCS 26.3)
67 Diseases of male genital organs (CCS 27.1)
68 Nonmalignant breast conditions, Inflammatory diseases of female pelvic organs, (CCS 27.2.1)
69 Menstrual disorders AND Menopausal disorders (CCS 27.2.2)
70 Contraceptive and procreative management (CCS 28.1)
71 Hypertension complicating pregnancy, childbirth and the puerperium (CCS 28.2.1)
72 Diabetes or abn. glucose tolerance complicating pregn., childbirth,
or the puerperium (CCS 28.2.2)
73 Skin and subcutaneous tissue infections AND Other inflammatory
conditions of skin (CCS 29.1)
74 Chronic ulcer of skin (CCS 29.2)
75 Infective arthritis and osteomyelitis (except that caused by TB or STD) (CCS 30.1)

76 Non-traumatic joint disorders (CCS 30.2)
77 Spondylosis, intervertebral disc disorders, other back problems (CCS 30.3)
78 Osteoporosis AND Pathological fracture (CCS 30.4)
79 Systemic lupus erythematosus and connective tissue disorders, Other connective tissure disease (CCS 30.5)
80 Cardiac and circulatory congenital anomalies (CCS 31.1)
81 Nervous system congenital anomalies (CCS 31.2)
82 Certain conditions originating in the perinatal period (CCS 32)
83 Spinal cord injury (CCS 33.1)
84 Intracranial injury (CCS 33.2)
85 Crushing injury or internal injury (CCS 33.3)
86 Open wounds (CCS 33.4)
87 Burns (CCS 33.5)
88 Complications of surgical procedures or medical care (CCS 33.6)
89 Symptoms, signs, and ill-defined conditions and factors influencing health status (CCS 34)

Table 5.4

New Drug Classification - MediCal (All variables utilized*)		
Rx #	**Category**	**Drug Description and (Multum Classes)**
1	Anticoagulants	Heparins (261), Warfarin (262)
2	Asthma	Antiasthmatic combinations (131), Adrenergic bronchodilators (180), Bronchodialators combinations (181), Methylxantihines (126), Leukotriene modifier (243), Respiratory Inhalants (130)
3	Autoimmune	Azathioprine (in 104 and 192)
4	Burns	Silver Sulfadiazine (in140)
5	Cardiovascular	ACE inhibitors (42), Beta blockers (274 & 275), Nitrates (45), Vasodilators (52 & 53), Calcium channel blockers (48), Digoxin (in 50)
6	Arrhythmias	Antiarrhythmic agents (46)
7	Cystic fibrosis	Pancrelipase (in 91)
8	Depression/anxiety	Antidepressants (76, 208 & 209), Antianxiety (in 69 - alprazolam, chlordiazepoxide, clonazepam, clorazepate, diazepam, halazepam, lorazepam & oxazepam) and (in 70 - buspirone, doxepin, ethchlorvynol, & meprobamate)
9	Diabetes	Insulin (215), Sulfonylureas (213), Alpha-glucosidase inhibitors (216), Thiazolidinediones (271), Metformin (214)
10	ESRD/renal	Epoietin Alfa (in 36), Calcitriol (in 119)
11	Gastric acid disorders	H2 Blockers (272), PPIs (94)
12	Gout	Colchicine (in 194), Allopurinol (in 194)
13	Hepatitis	Interferon beta (in 256), Peginterferon (in 256 & 177), Ribavirin (in 229)
14	Herpes	Acyclovir (in 229)
15	HIV/AIDS	Antiretrovirals (175, 176, & 227)
16	Hyperlipidemia	Antihyperlipidemics (173 & 174)

17	Infections	Quinolones (14), Cephalosporins (159 - 162), Penicillins (223, 224, & 226), Macrolides (11), Sulfonamides (15), Penicillinase resistant Penicillins (222), Tetracyclines (16), Beta-lactamase inhibitors (225), Urinary anti-infectives (17)
18	Insomnia	Sedatives, Hypnotics (in 69 estazolam, flurazepam, midazolam, quazepam, temazepam & triazolam) (in 70 acetylcarbromal, chloral hydrate, chlormezanone, dexmedetomidine, doxylamine, hydroxyzine, paraldehyde, propiomazine, pyrilamine, zaleplon, & zolpidem)
19	Liver disease	Lactulose (in 95)
20	Malignancies	Antineoplastics (22 - 25)
21	Multiple sclerosis/paralysis	Baclofen (178)
22	Nausea	Antiemetics (195 - 198 minus diphenhydramine)
23	Neurogenic bladder	Oxybutynin (in 264)
24	Osteoporosis/paget s	Etidronate/calcium regulators (217)
25	Pain	Narcotics (60 & 191)
26	Parkinsons/tremor	Dopaminergic antiparkinson agents (276), Benztropine (in 205), Trihexyphenidyl (in 205),
27	Psychotic illness/bipolar	Antipsychotics, Lithium (77 & 79)
28	Seizure disorders	Anticonvulsants (199 - 204)
29	Thyroid disorder	Thyroid hormones (103)
30	Transplant	Immunosuppressive agents (104)
31	Tuberculosis AND PCP pneumonia	Rifampin (in 232), Isoniazid (in231)

*Some variables are added to the Diagnostic Model and some variables are combined with Diagnostic Model categories

† The categories of Alzheimer's disease and iron deficiency were dropped due to formulary restrictions in the MediCal system

Table 5.5

New Drug Classification - Combined with CCS Categories		
Rx Category	**Drug Description and (Multum Classes)**	**Combined with CCS Category***
Asthma (Rx 2)	Antiasthmatic combinations (131), Adrenergic bronchodilators (180), Bronchodialators combinations (181), Methylxantihines (126), Leukotriene modifier (243), Respiratory Inhalants (130)	CCS 23.3 - Asthma
Cardiovascular (Rx5)	ACE inhibitors (42), Beta blockers (274 & 275), Nitrates (45), Vasodilators (52 & 53), Calcium channel blockers (48), Digoxin (in 50)	CCS 14 - Hypertension CCS 17 - Coronary atherosclerosis and other heart disease **Logic: 1.** Multum class 42 (ACE inhibitors), 48 (Ca channel blockers), 52 & 53 (Vasodilators), 274 & 275 (Beta blockers) combined with CCS 14. **2.** Multum class 45 (Nitrates) and Digoxin combined with CCS 17
Cystic fibrosis (Rx 7)	Pancrelipase (in 91)	CCS 5.4 Cystic fibrosis
Depression/anxiety (Rx 8)	Antidepressants (76, 208 & 209), Antianxiety (in 69 - alprazolam, chlordiazepoxide, clonazepam, clorazepate, diazepam, halazepam, lorazepam & oxazepam) and (in 70 - buspirone, doxepin, ethchlorvynol, & meprobamate)	CCS 9.1 - Affective disorders CCS 9.3 - Anxiety, somatoform, dissociative, and personality disorders CCS 9.4 - Other mental conditions **Logic:** if no diagnosis is present within CCS 9.1, 9.3, or 9.4 then combined with CCS 9.1

Diabetes (Rx9)	Insulin (215), Sulfonylureas (213), Alpha-glucosidase inhibitors(216), Thiazolidinediones (271), Metformin (214)	CCS 4.1 - Diabetes w/o complications CCS 4.2 - Diabetes w/ complications **Logic:** if no diagnosis is present within CCS 4.1 or 4.2 then combined with CCS 4.1
Gastric acid disorders (Rx11)	H2 Blockers (272), PPIs (94)	CCS 24.2.1 - Esophageal disorder CCS 24.2.2 - Gastroduodenal ulcer and gastritis and duodenitis **Logic:** if no diagnosis is present within CCS 24.2.1 or 24.2.2 then combined with CCS 24.2.2
Gout (Rx12)	Colchicine (in 194), Allopurinol (in 194)	CCS 5.3 - Gout and other crystal arthropathies
Hepatitis (Rx13)	Interferon beta (in 256), Peginterferon (in 256 & 177), Ribavirin (in229)	CCS 1.3 - Hepatitis
HIV/AIDS (Rx15)	Antiretrovirals (175, 176, & 227)	CCS 2 - HIV Infection
Hyperlipidemia (Rx 16)	Antihyperlipidemics (173 & 174)	CCS 5.2 - Disorders of lipid metabolism
Infections (Rx17)	Quinolones (14), Cephalosporins (159 - 162), Penicillins (223, 224, & 226), Macrolides (11), Sulfonamides (15), Tetracyclines (16), Penicillinase resistant Penicillins (222), Beta-lactamase inhibitors (225), Urinary anti-infectives (17)	CCS 1.1.1, 1.1.2, 1.1.3, 1.1.4, CCS 1.2, 1.3 CCS 2, CCS 12.1, 12.2 CCS 13.1, CCS 23.1.1, 23.1.2, CCS 24.1 CCS 26.3, CCS 29.1 - Infections **Logic:** If no diagnosis within CCS 1.1.1, 1.1.2, 1.1.3, 1.2, 1.3, 2, 12.1, 12.2, 13.1, 23.1.1, 23.1.2, 24.1, 26.3, or 29.1 then combined with CCS 1.1.4 - other bacterial infections
Malignancies (Rx 20)	Antineoplastics (22 - 25)	CCS 3 Neoplasms

Multiple sclerosis/paralysis (Rx21)	Baclofen (178)	CCS 13.2.2 - Multiple sclerosis
Osteoporosis/pagets (Rx 24)	Etidronate/calcium regulators (217)	CCS 30.4 - Osteoporosis and pathological fracture
Psychotic illness/bipolar (Rx 27)	Antipsychotics, Lithium (77 & 79)	CCS 9.2 - Schizophrenia and related disorders and other psychoses
Seizure disorders (Rx 28)	Anticonvulsants (199 - 204)	CCS 13.4 - Epilepsy, convulsions
Thyroid disorder (Rx29)	Thyroid hormones (103)	CCS 5.1 - Thyroid disorders

* Model variables are binomial 1/0 (1 if present). If diagnosis is present within combined CCS groups (variable = 1) then the presence of a Rx group will not affect the model.

Table 5.6

Medicaid MRxCost Model: All Diagnostic Variables Submitted to OLS Estimation

#	Variable
1	Tuberculosis (CCS 1.1.1)
2	Septicemia (except in labor) (CCS 1.1.2)
3	Sexually transmitted infections (not HIV or hepatitis) (CCS 1.1.3)
4	Other bacterial infections (CCS 1.1.4)
5	Mycoses other than candidiasis (CCS 1.2)
6	Hepatitis (CCS 1.3)
7	HIV Infection (CCS 2)
8	Neoplasms (CCS 3)
9	Diabetes mellitus without complication (CCS 4.1)
10	Diabetes mellitus with other complications (CCS 4.2)
11	Thyroid disorders (CCS 5.1)
12	Disorders of lipid metabolism (CCS 5.2)
13	Gout and other crystal arthropathies (CCS 5.3)
14	Cystic fibrosis (CCS 5.4)
15	Immunity disorders (CCS 5.5)
16	Other nutritional, endocrine, and metabolic disorders (CCS 5.6)
17	Anemia (CCS 6.1)
18	Coagulation and hemorrhagic disorders (CCS 6.2)
19	Diseases of white blood cells AND other hematologic conditions (CCS 6.3)
20	Mental retardation (CCS 7)
21	Senility and organic mental disorders (CCS 8)
22	Affective disorders (CCS 9.1)
23	Schizophrenia and related disorders and other psychoses (CCS 9.2)
24	Anxiety, somatoform, dissociative, and personality disorders (CCS 9.3)
25	Other mental conditions (CCS 9.4)
26	Alcohol and substance-related mental disorders (CCS 10)
27	Glaucoma (CCS 11.1)]
28	Other eye disorders (CCS 11.2)
29	Otitis media and related conditions (CCS 12.1)
30	Conditions associated with dizziness or vertigo AND other sense organ disorders (CCS 12.2)
31	Central nervous system infection (CCS 13.1)
32	Parkinson's disease (CCS 13.2.1)
33	Multiple sclerosis (CCS 13.2.2)
34	Paralysis (CCS 13.3)
35	Epilepsy, convulsions (CCS 13.4)
36	Migraine (CCS 13.5)
37	Hypertension (CCS 14)
38	Heart valve disorders (CCS 15)

39 Acute myocardial infarction (CCS 16)
40 Coronary atherosclerosis and other heart disease (CCS 17)
41 Congestive heart failure (nonhypertensive) and pulmonary heart disease (CCS 18)
42 Cardiac dysrhythmias (CCS 19)
43 Cerebrovascular disease (CCS 20)
44 Peripheral and visceral atherosclerosis (CCS 21.1)
45 Hypotention and Other unspecified circulatory disease (CCS 21.2)
46 Phlebitis, thrombophlebitis and thromboembolism AND
Varicose veins of lower extremity (CCS 21.3)
47 Peri-, endo-, and myocarditis, cardiomyopathy (except that caused by
TB or STD) (CCS 22.1)
48 Conduction disorders (CCS 22.2)
49 Pneumonia (except that caused by TB or STD) (CCS 23.1.1)
50 Influenza, Acute and chronic tonsillitis, Acute bronchitis AND other URIs (CCS 23.1.2)
51 Chronic obstructive pulmonary disease and bronchiectasis (CCS 23.2)
52 Asthma (CCS 23.3)
53 Lung disease due to external agents (CCS 23.4)
54 Other lower respiratory disease (CCS 23.5)
55 Other upper respiratory disease (CCS23.6)
56 Intestinal infection (CCS 24.1)
57 Esophageal disorders (CCS 24.2.1)
58 Gastroduodenal ulcer (except hemorrhage) AND Gastritis and duodenitis (CCS 24.2.2)
59 Other disorders of stomach and duodenum (CCS 24.2.3)
60 Lower gastrointestinal disorders AND Biliary tract disease (CCS 24.3)
61 Pancreatic disorders (not diabetes) (CCS 24.4)
62 Constipation, Dysphagia, and Other unspecified GI disorders (CCS 24.5)
63 Liver disease (CCS 25)
64 Nephritis, nephrosis, renal sclerosis (CCS 26.1)
65 Chronic renal failure (CCS 26.2)
66 Urinary tract infections, Calculus of the urinary tract AND Other disesases of
the kidney and ureters (CCS 26.3)
67 Diseases of male genital organs (CCS 27.1)
68 Nonmalignant breast conditions, Inflammatory diseases of female pelvic organs, (CCS 27.2.1)
69 Menstrual disorders AND Menopausal disorders (CCS 27.2.2)
70 Contraceptive and procreative management (CCS 28.1)
71 Hypertension complicating pregnancy, childbirth and the puerperium (CCS 28.2.1)
72 Diabetes or abn. glucose tolerance complicating pregn., childbirth,
or the puerperium (CCS 28.2.2)
73 Skin and subcutaneous tissue infections AND Other inflammatory
conditions of skin (CCS 29.1)
74 Chronic ulcer of skin (CCS 29.2)
75 Infective arthritis and osteomyelitis (except that caused by TB or STD) (CCS 30.1)

76 Non-traumatic joint disorders (CCS 30.2)
77 Spondylosis, intervertebral disc disorders, other back problems (CCS 30.3)
78 Osteoporosis AND Pathological fracture (CCS 30.4)
79 Systemic lupus erythematosus and connective tissue disorders, Other connective tissure disease (CCS 30.5)
80 Cardiac and circulatory congenital anomalies (CCS 31.1)
81 Nervous system congenital anomalies (CCS 31.2)
82 Certain conditions originating in the perinatal period (CCS 32)
83 Spinal cord injury (CCS 33.1)
84 Intracranial injury (CCS 33.2)
85 Crushing injury or internal injury (CCS 33.3)
86 Open wounds (CCS 33.4)
87 Burns (CCS 33.5)
88 Complications of surgical procedures or medical care (CCS 33.6)
89 Symptoms, signs, and ill-defined conditions and factors influencing health status (CCS 34)
90 Anticoagulants (RX 1)
91 Autoimmune (RX 3)
92 Burns (RX 4)
93 Arrhythmias (RX 6)
94 ESRD/renal disorders (RX 10)
95 Herpes (RX 14)
96 Insomnia (RX 18)
97 Liver disease (RX 19)
98 Nausea (RX 22)
99 Neurogenic bladder (RX 23)
100 Pain (RX 25)
101 Parkinson's/tremor (RX 26)
102 Transplant (RX 30)
103 Tuberculosis AND PCP pneumonia (RX 31)
104 Number of Asthma drug classes
105 Number of Cardiovascular drug classes
106 Number of Seizure disorder drug classes

Table 5.7

Descriptive Statistics for MediCal Training and Validation Data				
	Training		**Validation**	
n	138,454		92,621	
Male	38,970	28.2%	26,093	28.2%
Female	99,484	71.8%	66,528	71.8%
Age (mean)	35.2		35.2	
TANF/AFDC	102,077	73.7%	68,455	73.9%
SSI, Blind or Disabled	36,377	26.3%	24,166	26.1%
# of Months Eligible (mean)	22.0		22.0	
Demographic Model Variables				
male age 18-20	5,691	4.1%	3,887	4.2%
male age 23-30	6,144	4.4%	4,087	4.4%
male age 31-40	10,530	7.6%	6,978	7.5%
male age 41-50	9,481	6.8%	6,370	6.9%
male age 51-60	6,088	4.4%	4,086	4.4%
male age 61-64	1,036	0.7%	685	0.7%
female age 18-20	14,729	10.6%	9,915	10.7%
female age 23-30	25,414	18.4%	17,133	18.5%
female age 31-40	30,280	21.9%	20,022	21.6%
female age 41-50	18,566	13.4%	12,635	13.6%
female age 51-60	8,963	6.5%	5,794	6.3%
female age 61-64	1,532	1.1%	1,029	1.1%
Year 1 Rx Cost	$400.23		$394.67	
Year 2 Rx Cost	$468.38		$459.65	

Table 5.8

Results: Medicaid RxCost Model Estimation

Number of variables = 56
Number of obs = 138,454
R-square = 0.27
Adjusted R-square = 0.27
Root MSE = 2002.9

Variable	Coefficient	p-value	Lower 95% CI	Upper 95% CI
Male				
18 to 22 Years	4.48	-	-	-
23 to 30 Years	11.64	-	-	-
31 to 40 Years	13.34	-	-	-
41 to 50 Years	21.66	-	-	-
51 to 60 Years	30.61	-	-	-
61 to 64 Years	Reference	-	-	-
Female				
18 to 22 Years	11.91	-	-	-
23 to 30 Years	16.02	-	-	-
31 to 40 Years	20.02	-	-	-
41 to 50 Years	23.10	-	-	-
51 to 60 Years	30.57	-	-	-
61 to 64 Years	32.03	-	-	-
Medicaid Eligibility				
Disability or SSI	691.48	0.000	666.65	716.32
Dxn Categories				
Other bacterial infections (CCS 1.1.4)	419.51	0.000	258.67	580.35
Mycoses other than candidiasis (CCS 1.2)	132.02	0.000	68.69	195.34

Variable	Coefficient	p-value	Lower 95% CI	Upper 95% CI
Hepatitis (CCS 1.3)	167.99	0.000	85.23	250.74
HIV Infection (CCS 2)	8,936.68	0.000	8,790.05	9,083.31
Neoplasms (CCS 3)	274.05	0.000	216.84	331.27
[Hierarchy 1: estimated highest to least cost]				
[1] Diabetes mellitus with other complications (CCS 4.2)	723.97	0.000	639.88	808.06
[2] Diabetes mellitus without complication (CCS 4.1)	493.37	0.000	424.43	562.30
Disorders of lipid metabolism (CCS 5.2)	385.86	0.000	328.02	443.69
Cystic fibrosis (CCS 5.4)	9,175.23	0.000	8,440.92	9,909.54
Immunity disorders (CCS 5.5)	4,362.92	0.000	4,029.86	4,695.97
Other nutritional, endocrine, and metabolic disorders (CCS 5.6)	330.02	0.000	261.40	398.65
Coagulation and hemorrhagic disorders (CCS 6.2)	1,107.94	0.000	908.18	1,307.70
Diseases of white blood cells AND other hematologic conditions (CCS 6.3)	826.95	0.000	648.16	1,005.74
Mental retardation (CCS 7)	306.40	0.000	179.55	433.25
Schizophrenia and related disorders and other psychoses (CCS 9.2)	2,098.71	0.000	2,009.19	2,188.23
[Hierarchy 2: estimated highest to least cost]				
[1] Other mental conditions (CCS 9.4)	604.21	0.000	521.96	686.46
[2] Affective disorders (CCS 9.1)	454.08	0.000	370.75	537.40
[3] Anxiety, somatoform, dissociative, and personality disorders (CCS 9.3)	90.10	0.007	24.86	155.34
Glaucoma (CCS 11.1)]	312.52	0.000	184.64	440.41
Other eye disorders (CCS 11.2)	210.56	0.000	173.96	247.16
Central nervous system infection (CCS 13.1)	621.60	0.000	356.55	886.66
Parkinson's disease (CCS 13.2.1)	1,747.34	0.000	1,286.60	2,208.07
Multiple sclerosis (CCS 13.2.2)	1,473.50	0.000	1,172.34	1,774.66
Epilepsy, convulsions (CCS 13.4)	565.66	0.000	485.38	645.95
Migraine (CCS 13.5)	197.20	0.000	110.13	284.26
Hypertension (CCS 14)	255.09	0.000	208.36	301.82
Coronary atherosclerosis and other heart disease (CCS 17)	284.82	0.000	208.46	361.18
Pneumonia (except that caused by TB or STD) (CCS 23.1.1)	508.93	0.000	415.64	602.22
Chronic obstructive pulmonary disease and bronchiectasis (CCS 23.2)	251.42	0.000	198.92	303.93
Other lower respiratory disease (CCS 23.5)	136.95	0.000	89.98	183.91

Variable	Coefficient	p-value	Lower 95% CI	Upper 95% CI
[Hierarchy 3: estimated highest to least cost]				
[1] Esophageal disorders (CCS 24.2.1)	661.13	0.000	580.76	741.50
[2] Gastroduodenal ulcer (except hemorrhage) AND Gastritis and duodenitis (CCS 24.2.2)	53.65	0.120	-14.05	121.36
Other disorders of stomach and duodenum (CCS 24.2.3)	127.24	0.014	25.52	228.96
Pancreatic disorders (not diabetes) (CCS 24.4)	473.48	0.000	284.16	662.80
Constipation, Dysphagia, and Other unspecified GI disorders (CCS 24.5)	387.10	0.000	325.77	448.44
Liver disease (CCS 25)	242.99	0.000	150.46	335.52
Chronic renal failure (CCS 26.2)	2,366.89	0.000	2,175.29	2,558.49
Skin and subcutaneous tissue infections AND Other inflammatory Conditions of skin (CCS 29.1)	176.56	0.000	115.55	237.57
Chronic ulcer of skin (CCS 29.2)	540.49	0.000	360.42	720.55
[Hierarchy 4: estimated highest to least cost]				
[1] Osteoporosis AND Pathological fracture (CCS 30.4)	589.12	0.000	452.30	725.95
[2] Systemic lupus erythematosus and connective tissue disorders, Other connective tissure disease (CCS 30.5)	230.09	0.000	175.07	285.11
[3] Non-traumatic joint disorders (CCS 30.2)	214.96	0.000	171.05	258.88
[4] Infective arthritis and osteomyelitis (except that caused by TB or STD) (CCS 30.1)	54.49	0.684	-207.69	316.68
[5] Spondylosis, intervertebral disc disorders, other back problems (CCS 30.3)	10.67	0.713	-46.21	67.55

* R-square adjusted for shrinkage using the van Howelingen method:
R^2 adj = 1 - (1-R^2) x (n-1) / (n-p-1)

Table 5.9

Results: Medicaid MRxCost Model Estimation

Number of variables = 58
Number of obs = 138,454
R-square = 0.33
Adjusted R-square = 0.33
Root MSE = 1911.2

Variable	Coefficient	p-value	Lower 95% CI	Upper 95% CI
Male				
18 to 22 Years	3.7	-	-	-
23 to 30 Years	7.84	-	-	-
31 to 40 Years	10.44	-	-	-
41 to 50 Years	12.73	-	-	-
51 to 60 Years	24.29	-	-	-
61 to 64 Years	Reference	-	-	-
Female				
18 to 22 Years	11.3	-	-	-
23 to 30 Years	12.62	-	-	-
31 to 40 Years	12.85	-	-	-
41 to 50 Years	15.45	-	-	-
51 to 60 Years	23.89	-	-	-
61 to 64 Years	30.42	-	-	-
Medicaid Eligibility				
Disability or SSI	275.84	0.000	251.26	300.43
Dxn Categories				
HIV Infection (CCS 2)	8,333.10	0.000	8,200.85	8,465.34
Neoplasms (CCS 3)	782.87	0.000	728.07	837.67

Variable	Coefficient	p-value	Lower 95% CI	Upper 95% CI
Hierarchy 1: estimated highest to least cost]				
[1] Diabetes mellitus with other complications (CCS 4.2)	605.70	0.000	525.17	686.22
[2] Diabetes mellitus without complication (CCS 4.1)	349.36	0.000	287.79	410.93
Disorders of lipid metabolism (CCS 5.2)	422.86	0.000	372.79	472.94
Cystic fibrosis (CCS 5.4)	9,523.26	0.000	8,821.17	10,225.35
Immunity disorders (CCS 5.5)	3,536.32	0.000	3,217.40	3,855.23
Other nutritional, endocrine, and metabolic disorders (CCS 5.6)	243.31	0.000	177.96	308.67
Coagulation and hemorrhagic disorders (CCS 6.2)	592.82	0.000	398.40	787.25
Diseases of white blood cells AND other hematologic conditions (CCS 6.3)	481.95	0.000	311.15	652.75
Mental retardation (CCS 7)	220.04	0.000	99.79	340.28
Schizophrenia and related disorders and other psychoses (CCS 9.2)	1,478.65	0.000	1,422.44	1,534.86
Hierarchy 2: estimated highest to least cost]				
[1] Other mental conditions (CCS 9.4)	608.46	0.000	529.13	687.79
[2] Affective disorders (CCS 9.1)	662.16	0.000	619.44	704.88
[3] Anxiety, somatoform, dissociative, and personality disorders (CCS 9.3)	130.60	0.000	67.40	193.80
Glaucoma (CCS 11.1)]	219.15	0.000	97.09	341.22
Other eye disorders (CCS 11.2)	53.55	0.000	18.88	88.22
Central nervous system infection (CCS 13.1)	511.04	0.000	258.24	763.85
Parkinson's disease (CCS 13.2.1)	1,393.80	0.000	951.99	1,835.62
Multiple sclerosis (CCS 13.2.2)	1,429.34	0.000	1,255.54	1,603.15
Epilepsy, convulsions (CCS 13.4)	295.18	0.000	192.14	398.22
Pneumonia (except that caused by TB or STD) (CCS 23.1.1)	290.92	0.001	202.55	379.28
Hierarchy 3: estimated highest to least cost]				
[1] Esophageal disorders (CCS 24.2.1)	581.47	0.000	504.21	658.73
[2] Gastroduodenal ulcer (except hemorrhage) AND Gastritis and duodenitis (CCS 24.2.2)	223.53	0.000	177.32	269.74
Constipation, Dysphagia, and Other unspecified GI disorders (CCS 24.5)	214.08	0.000	155.62	272.53
Liver disease (CCS 25)	77.07	0.014	-9.28	163.43
Chronic renal failure (CCS 26.2)	656.49	0.000	458.17	854.81
Skin and subcutaneous tissue infections AND Other inflammatory conditions of skin (CCS 29.1)	94.00	0.000	36.13	151.87

Variable	Coefficient	p-value	Lower 95% CI	Upper 95% CI
Chronic ulcer of skin (CCS 29.2)	584.32	0.000	412.23	756.41
Complications of surgical procedures or medical care (CCS 33.6)	672.44	0.000	545.91	798.96
Anticoagulants (RX 1)	1,420.10	0.000	1,274.56	1,565.63
Autoimmune (RX 3)	470.96	0.000	78.98	862.93
Burns (RX 4)	210.19	0.001	14.38	406.01
Arrhythmias (RX 6)	702.63	0.000	405.71	999.56
ESRD/renal disorders (RX 10)	3,504.71	0.000	3,220.78	3,788.65
Herpes (RX 14)	1,006.84	0.000	885.12	1,128.55
Insomnia (RX 18)	292.06	0.000	239.65	344.48
Liver disease (RX 19)	1,634.84	0.000	1,219.77	2,049.92
Nausea (RX 22)	320.69	0.000	270.93	370.45
Neurogenic bladder (RX 23)	351.44	0.000	162.37	540.51
Parkinson's/tremor (RX 26)	639.59	0.000	566.10	713.08
Transplant (RX 30)	457.80	0.000	204.95	710.62
Tuberculosis AND PCP pneumonia (RX 31)	5,987.62	0.000	5,622.26	6,352.98
No. classes within Asthma	270.99	0.000	244.97	297.01
No. classes within Cardiovascular	236.29	0.000	209.78	262.80
No. classes within Seizure disorders	509.76	0.000	436.17	583.35

* R-square adjusted for shrinkage using the van Howelingen method:

R^2 adj = 1 - (1-R^2) x (n-1) / (n-p-1)

Table 5.10

Model R-Square Values - MediCal with Untrimmed Data (Coefficients Estimated using Untrimmed Training Sample)

Model	Training Sample Untrimmed	Validation Sample Untrimmed	Validation Sample Trimmed
Demographics Model	0.03	0.03	0.06
Demographics and Medicaid Eligibility Model	0.10	0.10	0.17
Medicaid RxCost Model	0.27	0.24	0.51
Concurrent MedicaidRxCost Model	-	0.48	0.58
Medicaid MRxCost Model	0.34	0.30	0.63
CDPS Model	-	0.04	-

Table 5.11

Pooled R-Square and Predicted Ratio Values for Groups of Members

	RxCost Model		MRxCost Model	
Group Size	**R-Square Value**	**Predicted Ratio (95% CI)**	**R-Square Value**	**Predicted Ratio (95% CI)**
10 members	0.24	0.86 (0.84, 0.88)	0.26	0.90 (0.88, 0.92)
20 members	0.24	0.88 (0.86, 0.90)	0.26	0.92 (0.90, 0.94)
50 members	0.26	0.90 (0.88, 0.92)	0.26	0.94 (0.92, 0.96)
100 members	0.29	0.90 (0.88, 0.93)	0.27	0.95 (0.92, 0.97)
200 members	0.33	0.90 (0.88, 0.93)	0.28	0.95 (0.92, 0.97)
300 members	0.38	0.90 (0.88, 0.93)	0.30	0.95 (0.92, 0.97)
400 members	0.41	0.90 (0.88, 0.93)	0.30	0.95 (0.92, 0.97)
500 members	0.52	0.90 (0.88, 0.93)	0.37	0.95 (0.92, 0.98)

CHAPTER 6

CONCLUSION

The RxCost Model was the main focus of this research. This model was developed to predict prescription expenditures for both a commercial and a Medicaid population. Although no other risk assessment prescription models are publicly reported in the literature, these models appear to perform well as compared to other models that predict overall health costs. Additionally, these models performed well when directly compared to DCG-HCC on the validation sample of the commercial population and the CDPS on the validation sample of the Medicaid population even after both models were re-calibrated to predict prescription costs. The RxCost Model could potentially be used to set prescription goals for physicians and physician groups based on their patient-mix. These goals could then be tied to financial incentives where physicians would receive compensation for meeting their goals.

This model also proved to perform well when run in a concurrent fashion. Run concurrently, the model can potentially be used to profile physician prescribing behaviors. Here, prescribing behaviors can be used to educate physicians or address prescribing concerns.

Another model, the MRxCost Model, was also developed to predict prescription expenditures for both populations. This exploratory model showed that substantial

predictive power can be gained from adding drug information to the RxCost Model. However, the advantages and disadvantages of this model should be carefully weighed before using this type of mixed model to predict pharmacy expenditures. These models are the first of their kind that specifically predicts prescription expenditures. Further research in this area needs to be done to examine just how much predictive power can be achieved and how much utility these models actually have.

APPENDIX A

ORIGINAL CCS CLASSIFICATION

Original CCS Classification

1 Infectious and parasitic diseases
- 1.1 Bacterial infection
  - 1.1.1 Tuberculosis
  - 1.1.2 Septicemia (except in labor)
    - 1.1.2.1 Streptococcal septicemia
    - 1.1.2.2 Staphylococcal septicemia
    - 1.1.2.3 E. Coli septicemia
    - 1.1.2.4 Other gram negative septicemia
    - 1.1.2.5 Other specified septicemia
    - 1.1.2.6 Unspecified septicemia
  - 1.1.3 Sexually transmitted infections (not HIV or hepatitis)
  - 1.1.4 Other bacterial infections
- 1.2 Mycoses
  - 1.2.1 Candidiasis of the mouth (thrush)
  - 1.2.2 Other mycoses
- 1.3 Viral infection
  - 1.3.1 HIV infection
  - 1.3.2 Hepatitis
  - 1.3.3 Other viral infections
    - 1.3.3.1 Herpes zoster infection
    - 1.3.3.2 Herpes simplex infection
    - 1.3.3.3 Other and unspecified viral infection
- 1.4 Other infections, including parasitic
- 1.5 Immunizations and screening for infectious disease

2 Neoplasms
- 2.1 Colorectal cancer
  - 2.1.1 Cancer of colon
  - 2.1.2 Cancer of rectum and anus
- 2.2 Other gastrointestinal cancer
  - 2.2.1 Cancer of esophagus
  - 2.2.2 Cancer of stomach
  - 2.2.3 Cancer of liver and intrahepatic bile duct
  - 2.2.4 Cancer of pancreas
  - 2.2.5 Cancer of other GI organs, peritoneum
- 2.3 Cancer of bronchus, lung
- 2.4 Cancer of skin
  - 2.4.1 Melanomas of skin
  - 2.4.2 Other non-epithelial cancer of skin

- 2.5 Cancer of breast
- 2.6 Cancer of uterus and cervix
  - 2.6.1 Cancer of uterus
  - 2.6.2 Cancer of cervix
- 2.7 Cancer of ovary and other female genital organs
  - 2.7.1 Cancer of ovary
  - 2.7.2 Cancer of other female genital organs
- 2.8 Cancer of male genital organs
  - 2.8.1 Cancer of prostate
  - 2.8.2 Cancer of testis
  - 2.8.3 Cancer of other male genital organs
- 2.9 Cancer of urinary organs
  - 2.9.1 Cancer of bladder
  - 2.9.2 Cancer of kidney and renal pelvis
  - 2.9.3 Cancer of other urinary organs
- 2.10 Cancer of lymphatic and hematopoietic tissue
  - 2.10.1 Hodgkin's disease
  - 2.10.2 Non-Hodgkin's lymphoma
  - 2.10.3 Leukemias
  - 2.10.4 Multiple myeloma
- 2.11 Cancer, other primary
  - 2.11.1 Cancer of head and neck
  - 2.11.2 Cancer, other respiratory and intrathoracic
  - 2.11.3 Cancer of bone and connective tissue
  - 2.11.4 Cancer of brain and nervous system
  - 2.11.5 Cancer of thyroid
  - 2.11.6 Cancer, other and unspecified primary
- 2.12 Secondary malignancies
  - 2.12.1 Secondary malignancy of lymph nodes
  - 2.12.2 Secondary malignancy of lung
  - 2.12.3 Secondary malignancy of liver
  - 2.12.4 Secondary malignancy of brain/spine
  - 2.12.5 Secondary malignancy of bone
  - 2.12.6 Other secondary malignancy
- 2.13 Malignant neoplasm without specification of site
- 2.14 Neoplasms of unspecified nature or uncertain behavior
- 2.15 Maintenance chemotherapy, radiotherapy
  - 2.15.1 Radiotherapy
  - 2.15.2 Chemotherapy
- 2.16 Benign neoplasms
  - 2.16.1 Benign neoplasm of uterus
  - 2.16.2 Other and unspecified benign neoplasm

- 2.16.2.1 Benign neoplasm of ovary
- 2.16.2.2 Benign neoplasm of colon
- 2.16.2.3 Benign neoplasm of the thyroid
- 2.16.2.4 Benign neoplasm of cerebral meninges
- 2.16.2.5 Other and unspecified benign neoplasms

3 Endocrine, nutritional, and metabolic diseases and immunity disorders

- 3.1 Thyroid disorders
  - 3.1.1 Thyrotoxicosis with or without goiter
  - 3.1.2 Other thyroid disorders
- 3.2 Diabetes mellitus without complication
- 3.3 Diabetes mellitus with complications
  - 3.3.1 Diabetes with ketoacidosis or uncontrolled diabetes
  - 3.3.2 Diabetes with renal manifestations
  - 3.3.3 Diabetes with ophthalmic manifestations
  - 3.3.4 Diabetes with neurological manifestations
  - 3.3.5 Diabetes with circulatory manifestations
  - 3.3.6 Diabetes with unspecified complications
  - 3.3.7 Diabetes with other manifestations
- 3.4 Other endocrine disorders
- 3.5 Nutritional deficiencies
  - 3.5.1 Unspecified protein-calorie malnutrition
  - 3.5.2 Other malnutrition
- 3.6 Disorders of lipid metabolism
- 3.7 Gout and other crystal arthropathies
- 3.8 Fluid and electrolyte disorders
  - 3.8.1 Hyposmolality
  - 3.8.2 Hypovolemia
  - 3.8.3 Hyperpotassemia
  - 3.8.4 Hypopotassemia
  - 3.8.5 Other fluid and electrolyte disorders
- 3.9 Cystic fibrosis
- 3.10 Immunity disorders
- 3.11 Other nutritional, endocrine, and metabolic disorders]
  - 3.11.1 Disorders of mineral metabolism
  - 3.11.2 Obesity
  - 3.11.3 Other and unspecified metabolic, nutritional, and endocrine disorders

4 Diseases of the blood and blood-forming organs

- 4.1 Anemia
  - 4.1.1 Acute posthemorrhagic anemia
  - 4.1.2 Sickle cell anemia
  - 4.1.3 Deficiency and other anemia

- 4.1.3.1 Iron deficiency anemia
- 4.1.3.2 Other deficiency anemia
- 4.1.3.3 Aplastic anemia
- 4.1.3.4 Chronic blood loss anemia
- 4.1.3.5 Acquired hemolytic anemia
- 4.1.3.6 Other specified anemia
- 4.1.3.7 Anemia, unspecified

4.2 Coagulation and hemorrhagic disorders
- 4.2.1 Coagulation defects
- 4.2.2 Thrombocytopenia
- 4.2.3 Other coagulation and hemorrhagic disorders

4.3 Diseases of white blood cells

4.4 Other hematologic conditions

5 Mental disorders

5.1 Mental retardation

5.2 Alcohol and substance-related mental disorders
- 5.2.1 Alcohol-related mental disorders
  - 5.2.1.1 Acute alcoholic intoxication
  - 5.2.1.2 Other and unspecified alcohol dependence
  - 5.2.1.3 Nondependent alcohol abuse
  - 5.2.1.4 Other alcohol-related mental disorders
- 5.2.2 Substance-related mental disorders
  - 5.2.2.1 Opioid dependence
  - 5.2.2.2 Cocaine dependence
  - 5.2.2.3 Other, combined, and unspecified drug dependence
  - 5.2.2.4 Cocaine abuse
  - 5.2.2.5 Other, mixed, or unspecified drug abuse
  - 5.2.2.6 Other substance-related mental disorders

5.3 Senility and organic mental disorders
- 5.3.1 Senile dementia, uncomplicated
- 5.3.2 Arteriosclerotic dementia
- 5.3.3 Transient organic psychotic conditions
- 5.3.4 Specific nonpsychotic mental disorders due to organic brain damage
- 5.3.5 Presenile dementia, uncomplicated
- 5.3.6 Senile dementia with delirium
- 5.3.7 Other senility and organic mental disorders

5.4 Affective disorders Affective disorders
- 5.4.1 Major depressive disorder, single episode
- 5.4.2 Major depressive disorder, recurrent episode
- 5.4.3 Neurotic depression
- 5.4.4 Bipolar affective disorder
- 5.4.5 Manic-depressive psychosis
- 5.4.6 Other affective disorders

- 5.5 Schizophrenia and related disorders
  - 5.5.1 Paranoid schizophrenia
  - 5.5.2 Schizo-affective type
  - 5.5.3 Other schizophrenia
- 5.6 Other psychoses
- 5.7 Anxiety, somatoform, dissociative, and personality disorders
  - 5.7.1 Anxiety states
  - 5.7.2 Personality disorders
  - 5.7.3 Other anxiety, somatoform, dissociative, and personality disorders
- 5.8 Preadult disorders
- 5.9 Other mental conditions
  - 5.9.1 Adjustment reaction
    - 5.9.1.1 Brief depressive reaction
    - 5.9.1.2 Other adjustment reaction
  - 5.9.2 Depressive disorder, not elsewhere classified
  - 5.9.3 Other and unspecified mental conditions
- 5.10 Personal history of mental disorder, screening for mental condition

6 Diseases of the nervous system and sense organs

- 6.1 Central nervous system infection
  - 6.1.1 Meningitis (except that caused by TB or STD)
  - 6.1.2 Encephalitis (except that caused by TB or STD)
  - 6.1.3 Other CNS infection and poliomyelitis
- 6.2 Hereditary and degenerative nervous system conditions
  - 6.2.1 Parkinson's disease
  - 6.2.2 Multiple sclerosis
  - 6.2.3 Other hereditary and degenerative nervous system conditions
    - 6.2.3.1 Disorders of the autonomic nervous system
    - 6.2.3.2 Other and unspecified hereditary and degenerative nervous conditions
- 6.3 Paralysis
  - 6.3.1 Hemiplegia
  - 6.3.2 Other paralysis
- 6.4 Epilepsy, convulsions
  - 6.4.1 Epilepsy
  - 6.4.2 Convulsions
- 6.5 Headache, including migraine
  - 6.5.1 Migraine
  - 6.5.2 Other headache
- 6.6 Coma, stupor, and brain damage
- 6.7 Eye disorders
  - 6.7.1 Cataract
  - 6.7.2 Retinal detachments, defects, vascular occlusion, and retinopathy

- 6.7.2.1 Retinal detachment with defect
- 6.7.2.2 Other retinal detachment or defect
- 6.7.2.3 Other retinal disorders
- 6.7.3 Glaucoma
- 6.7.4 Blindness and vision defects
- 6.7.5 Inflammation, infection of eye (except that caused by TB or STD)
- 6.7.6 Other eye disorders

6.8 Ear conditions

- 6.8.1 Otitis media and related conditions
  - 6.8.1.1 Suppurative and unspecified otitis media
  - 6.8.1.2 Other otitis media and related conditions
- 6.8.2 Conditions associated with dizziness or vertigo
- 6.8.3 Other ear and sense organ disorders

6.9 Other nervous system disorders

- 6.9.1 Disorders of the peripheral nervous system
- 6.9.2 Other central nervous system disorders
- 6.9.3 Other nervous system symptoms and disorders

7 Diseases of the circulatory system

7.1 Hypertension

- 7.1.1 Essential hypertension
- 7.1.2 Hypertension with complications and secondary hypertension
  - 7.1.2.1 Hypertensive heart and/or renal disease
  - 7.1.2.2 Other hypertensive complications

7.2 Diseases of the heart

- 7.2.1 Heart valve disorders
  - 7.2.1.1 Chronic rheumatic disease of the heart valves
  - 7.2.1.2 Nonrheumatic mitral valve disorders
  - 7.2.1.3 Nonrheumatic aortic valve disorders
  - 7.2.1.4 Other heart valve disorders
- 7.2.2 Peri-, endo-, and myocarditis, cardiomyopathy (except that caused caused by TB or STD)
  - 7.2.2.1 Cardiomyopathy
  - 7.2.2.2 Other peri-, endo-, and myocarditis
- 7.2.3 Acute myocardial infarction
- 7.2.4 Coronary atherosclerosis and other heart disease
  - 7.2.4.1 Angina pectoris
  - 7.2.4.2 Unstable angina (intermediate coronary syndrome)
  - 7.2.4.3 Other acute and subacute forms of ischemic heart disease
  - 7.2.4.4 Coronary atherosclerosis
  - 7.2.4.5 Other forms of chronic heart disease
- 7.2.5 Nonspecific chest pain
- 7.2.6 Pulmonary heart disease

- 7.2.7 Other and ill-defined heart disease
- 7.2.8 Conduction disorders
  - 7.2.8.1 Atrioventricular block
  - 7.2.8.2 Bundle branch block
  - 7.2.8.3 Anomalous atrioventricular excitation
  - 7.2.8.4 Other conduction disorders
- 7.2.9 Cardiac dysrhythmias
  - 7.2.9.1 Paroxysmal supraventricular tachycardia
  - 7.2.9.2 Paroxysmal ventricular tachycardia
  - 7.2.9.3 Atrial fibrillation
  - 7.2.9.4 Atrial flutter
  - 7.2.9.5 Premature beats
  - 7.2.9.6 Sinoatrial node dysfunction
  - 7.2.9.7 Other cardiac dysrhythmias
- 7.2.10 Cardiac arrest and ventricular fibrillation
- 7.2.11 Congestive heart failure, nonhypertensive
  - 7.2.11.1 Congestive heart failure
  - 7.2.11.2 Heart failure

7.3 Cerebrovascular disease
- 7.3.1 Acute cerebrovascular disease
  - 7.3.1.1 Intracranial hemorrhage
  - 7.3.1.2 Occlusion of cerebral arteries
  - 7.3.1.3 Acute but ill-defined cerebrovascular accident
- 7.3.2 Occlusion or stenosis of precerebral arteries
- 7.3.3 Other and ill-defined cerebrovascular disease
- 7.3.4 Transient cerebral ischemia
- 7.3.5 Late effects of cerebrovascular disease

7.4 Diseases of arteries, arterioles, and capillaries
- 7.4.1 Peripheral and visceral atherosclerosis
  - 7.4.1.1 Atherosclerosis of arteries of extremities
  - 7.4.1.2 Peripheral vascular disease unspecified
  - 7.4.1.3 Other peripheral and visceral atherosclerosis
- 7.4.2 Aortic, peripheral, and visceral artery aneurysms
  - 7.4.2.1 Abdominal aortic aneurysm, without rupture
  - 7.4.2.2 Other aneurysm
- 7.4.3 Aortic and peripheral arterial embolism or thrombosis
  - 7.4.3.1 Arterial embolism and thrombosis of lower extremity artery
  - 7.4.3.2 Other arterial embolism and thrombosis
- 7.4.4 Other circulatory disease
  - 7.4.4.1 Hypotension
  - 7.4.4.2 Other and unspecified circulatory disease

7.5 Diseases of veins and lymphatics

- 7.5.1 Phlebitis, thrombophlebitis and thromboembolism
  - 7.5.1.1 Phlebitis and thrombophlebitis
  - 7.5.1.2 Other venous embolism and thrombosis
- 7.5.2 Varicose veins of lower extremity
- 7.5.3 Hemorrhoids
- 7.5.4 Other diseases of veins and lymphatics

8 Diseases of the respiratory system

- 8.1 Respiratory infections
  - 8.1.1 Pneumonia (except that caused by TB or STD)
    - 8.1.1.1 Pneumococcal pneumonia
    - 8.1.1.2 Other bacterial pneumonia
    - 8.1.1.3 Pneumonia, organism unspecified
    - 8.1.1.4 Other pneumonia
  - 8.1.2 Influenza
  - 8.1.3 Acute and chronic tonsillitis
  - 8.1.4 Acute bronchitis
  - 8.1.5 Other upper respiratory infections
    - 8.1.5.1 Acute upper respiratory infections of multiple or unspecified sites
    - 8.1.5.2 Chronic sinusitis
    - 8.1.5.3 Croup
    - 8.1.5.4 Other and unspecified upper respiratory infections
- 8.2 Chronic obstructive pulmonary disease and bronchiectasis
  - 8.2.1 Emphysema
  - 8.2.2 Chronic airway obstruction, not otherwise specified
  - 8.2.3 Obstructive chronic bronchitis
  - 8.2.4 Other chronic pulmonary disease
- 8.3 Asthma
  - 8.3.1 Chronic obstructive asthma
    - 8.3.1.1 Chronic obstructive asthma without status asthmaticus or exacerbation
    - 8.3.1.2 Chronic obstructive asthma with status asthmaticus
    - 8.3.1.3 Chronic obstructive asthma with acute exacerbation
  - 8.3.2 Other and unspecified asthma
    - 8.3.2.1 Other asthma without status asthmaticus or exacerbation
    - 8.3.2.2 Other asthma with status asthmaticus
    - 8.3.2.3 Other asthma with acute exacerbation
- 8.4 Aspiration pneumonitis, food/vomitus
- 8.5 Pleurisy, pneumothorax, pulmonary collapse
  - 8.5.1 Pleurisy, pleural effusion
  - 8.5.2 Pulmonary collapse, interstitial and compensatory emphysema
  - 8.5.3 Empyema and pneumothorax

- 8.6 Respiratory failure, insufficiency, arrest (adult)
  - 8.6.1 Respiratory failure
  - 8.6.2 Other respiratory insufficiency
- 8.7 Lung disease due to external agents
- 8.8 Other lower respiratory disease
  - 8.8.1 Postinflammatory pulmonary fibrosis
  - 8.8.2 Painful respiration
  - 8.8.3 Other and unspecified lower respiratory disease
- 8.9 Other upper respiratory disease

9 Diseases of the digestive system

- 9.1 Intestinal infection
- 9.2 Disorders of teeth and jaw
- 9.3 Diseases of mouth, excluding dental
- 9.4 Upper gastrointestinal disorders
  - 9.4.1 Esophageal disorders
    - 9.4.1.1 Esophagitis
    - 9.4.1.2 Other esophageal disorders
  - 9.4.2 Gastroduodenal ulcer (except hemorrhage)
    - 9.4.2.1 Gastric ulcer
    - 9.4.2.2 Duodenal ulcer
    - 9.4.2.3 Peptic ulcer, site unspecified
    - 9.4.2.4 Gastrojejunal ulcer
  - 9.4.3 Gastritis and duodenitis
    - 9.4.3.1 Acute gastritis
    - 9.4.3.2 Other specified gastritis
    - 9.4.3.3 Unspecified gastritis and gastroduodenitis
    - 9.4.3.4 Duodenitis
  - 9.4.4 Other disorders of stomach and duodenum
- 9.5 Abdominal hernia
  - 9.5.1 Inguinal hernia
    - 9.5.1.1 Inguinal hernia with obstruction or gangrene
    - 9.5.1.2 Inguinal hernia without obstruction or gangrene
  - 9.5.2 Diaphragmatic hernia
  - 9.5.3 Other abdominal hernia
    - 9.5.3.1 Femoral hernia with obstruction/gangrene
    - 9.5.3.2 Femoral hernia without obstruction/gangrene
    - 9.5.3.3 Umbilical hernia with obstruction/gangrene
    - 9.5.3.4 Umbilical hernia without obstruction/gangrene
    - 9.5.3.5 Ventral hernia with obstruction/gangrene
    - 9.5.3.6 Ventral hernia without obstruction/gangrene
    - 9.5.3.7 Incisional hernia with obstruction/gangrene
    - 9.5.3.8 Incisional hernia without obstruction/gangrene
    - 9.5.3.9 Other and unspecified hernia

- 9.6 Lower gastrointestinal disorders
  - 9.6.1 Appendicitis and other appendiceal conditions
    - 9.6.1.1 Acute appendicitis with abscess or peritonitis
    - 9.6.1.2 Acute appendicitis without abscess or peritonitis
    - 9.6.1.3 Acute appendicitis, not otherwise specified
    - 9.6.1.4 Other appendiceal conditions
  - 9.6.2 Regional enteritis and ulcerative colitis
  - 9.6.3 Intestinal obstruction without hernia
    - 9.6.3.1 Paralytic ileus
    - 9.6.3.2 Impaction of intestine
    - 9.6.3.3 Peritoneal or intestinal adhesions
    - 9.6.3.4 Other intestinal obstruction
  - 9.6.4 Diverticulosis and diverticulitis
    - 9.6.4.1 Diverticulosis
    - 9.6.4.2 Diverticulitis
  - 9.6.5 Anal and rectal conditions
  - 9.6.6 Peritonitis and intestinal abscess
- 9.7 Biliary tract disease
  - 9.7.1 Cholelithiasis with acute cholecystitis
  - 9.7.2 Cholelithiasis with other cholecystitis
  - 9.7.3 Cholelithiasis without mention of cholecystitis
  - 9.7.4 Calculus of bile duct
  - 9.7.5 Cholecystitis without cholelithiasis
  - 9.7.6 Other biliary tract disease
- 9.8 Liver disease
  - 9.8.1 Liver disease, alcohol-related
  - 9.8.2 Other liver diseases
    - 9.8.2.1 Cirrhosis of liver without mention of alcohol
    - 9.8.2.2 Liver abscess and sequelae of chronic liver disease
    - 9.8.2.3 Ascites
    - 9.8.2.4 Other and unspecified liver disorders
- 9.9 Pancreatic disorders (not diabetes)
  - 9.9.1 Acute pancreatitis
  - 9.9.2 Chronic pancreatitis
  - 9.9.3 Other pancreatic disorders

- 9.10 Gastrointestinal hemorrhage
  - 9.10.1 Hemorrhage from gastrointestinal ulcer
  - 9.10.2 Melena
  - 9.10.3 Gastroesophageal laceration syndrome
  - 9.10.4 Other esophageal bleeding
  - 9.10.5 Hemorrhage of rectum and anus

- 9.10.6 Hematemesis
- 9.10.7 Hemorrhage of gastrointestinal tract
- 9.11 Noninfectious gastroenteritis
- 9.12 Other gastrointestinal disorders
  - 9.12.1 Constipation
  - 9.12.2 Dysphagia
  - 9.12.3 Other and unspecified gastrointestinal disorders

10 Diseases of the genitourinary system

- 10.1 Diseases of the urinary system
  - 10.1.1 Nephritis, nephrosis, renal sclerosis
  - 10.1.2 Acute and unspecified renal failure
    - 10.1.2.1 Acute renal failure
    - 10.1.2.2 Unspecified renal failure
  - 10.1.3 Chronic renal failure
  - 10.1.4 Urinary tract infections
    - 10.1.4.1 Infections of kidney
    - 10.1.4.2 Cystitis and urethritis
    - 10.1.4.3 Urinary tract infection, site not specified
  - 10.1.5 Calculus of urinary tract
    - 10.1.5.1 Calculus of kidney
    - 10.1.5.2 Calculus of ureter
    - 10.1.5.3 Other and unspecified urinary calculus
  - 10.1.6 Other diseases of kidney and ureters
    - 10.1.6.1 Hydronephrosis
    - 10.1.6.2 Other and unspecified diseases of kidney and ureters
  - 10.1.7 Other diseases of bladder and urethra
    - 10.1.7.1 Bladder neck obstruction
    - 10.1.7.2 Other and unspecified diseases of bladder and urethra
  - 10.1.8 Genitourinary symptoms and ill-defined conditions
    - 10.1.8.1 Hematuria
    - 10.1.8.2 Retention of urine
    - 10.1.8.3 Other and unspecified genitourinary symptoms
- 10.2 Diseases of male genital organs
  - 10.2.1 Hyperplasia of prostate
  - 10.2.2 Inflammatory conditions of male genital organs
  - 10.2.3 Other male genital disorders
- 10.3 Diseases of female genital organs
  - 10.3.1 Nonmalignant breast conditions
  - 10.3.2 Inflammatory diseases of female pelvic organs
    - 10.3.2.1 Pelvic peritoneal adhesions
    - 10.3.2.2 Cervicitis and endocervicitis
    - 10.3.2.3 Pelvic inflammatory disease (PID)
    - 10.3.2.4 Other inflammatory diseases of female pelvic organs

- 10.3.3 Endometriosis
- 10.3.4 Prolapse of female genital organs
- 10.3.5 Menstrual disorders
- 10.3.6 Ovarian cyst
- 10.3.7 Menopausal disorders
- 10.3.8 Female infertility
- 10.3.9 Other female genital disorders
  - 10.3.9.1 Female genital pain and other symptoms
  - 10.3.9.2 Other and unspecified female genital disorders

11 Complications of pregnancy, childbirth, and the puerperium

- 11.1 Contraceptive and procreative management
  - 11.1.1 Sterilization
  - 11.1.2 Other contraceptive and procreation management
- 11.2 Abortion-related disorders
  - 11.2.1 Spontaneous abortion
  - 11.2.2 Induced abortion
  - 11.2.3 Postabortion complications
- 11.3 Complications mainly related to pregnancy
  - 11.3.1 Ectopic pregnancy
  - 11.3.2 Hemorrhage during pregnancy, abruptio placenta, placenta previa
    - 11.3.2.1 Placenta previa
    - 11.3.2.2 Abruptio placenta
    - 11.3.2.3 Other hemorrhage during pregnancy, childbirth and the puerperium
  - 11.3.3 Hypertension complicating pregnancy, childbirth and the puerperium
    - 11.3.3.1 Preeclampsia and eclampsia
    - 11.3.3.2 Other hypertension in pregnancy
  - 11.3.4 Early or threatened labor
    - 11.3.4.1 Threatened premature labor
    - 11.3.4.2 Early onset of delivery
    - 11.3.4.3 Other early or threatened labor
  - 11.3.5 Prolonged pregnancy
  - 11.3.6 Diabetes or abn. glucose tolerance complicating pregn., childbirth, or the puerperium
  - 11.3.7 Other complications of pregnancy
    - 11.3.7.1 Infections of genitourinary tract during pregnancy
    - 11.3.7.2 Anemia during pregnancy
    - 11.3.7.3 Mental disorders during pregnancy
    - 11.3.7.4 Missed abortion
    - 11.3.7.5 Hyperemesis gravidarum
    - 11.3.7.6 Infectious and parasitic complications in mother affecting pregnancy
    - 11.3.7.7 Other and unspecified complications of pregnancy

- 11.4 Indications for care in pregnancy, labor, and delivery
  - 11.4.1 Malposition, malpresentation
    - 11.4.1.1 Breech presentation
    - 11.4.1.2 Other malposition, malpresentation
  - 11.4.2 Fetopelvic disproportion, obstruction
    - 11.4.2.1 Fetopelvic disproportion
    - 11.4.2.2 Other disproportion or obstruction
  - 11.4.3 Previous cesarean section
  - 11.4.4 Fetal distress and abnormal forces of labor
    - 11.4.4.1 Fetal distress
    - 11.4.4.2 Uterine inertia
    - 11.4.4.3 Precipitate labor
    - 11.4.4.4 Other abnormal forces of labor
  - 11.4.5 Polyhydramnios and other problems of amniotic cavity
    - 11.4.5.1 Premature rupture of membranes
    - 11.4.5.2 Infection of amniotic cavity
    - 11.4.5.3 Other problems of amniotic cavity
- 11.5 Complications during labor
  - 11.5.1 Umbilical cord complication
    - 11.5.1.1 Cord around neck with compression
    - 11.5.1.2 Other and unspecified cord entanglement with or without compression
    - 11.5.1.3 Other umbilical cord complications
  - 11.5.2 Trauma to perineum and vulva
    - 11.5.2.1 First degree perineal laceration
    - 11.5.2.2 Second degree perineal laceration
    - 11.5.2.3 Third degree perineal laceration
    - 11.5.2.4 Fourth degree perineal laceration
    - 11.5.2.5 Other perineal laceration and trauma
  - 11.5.3 Forceps delivery
- 11.6 Other complications of birth, puerperium affecting management of mother
  - 11.6.1 Postpartum hemorrhage
  - 11.6.2 Complications of the puerperium
  - 11.6.3 Cervical incompetence
  - 11.6.4 Rhesus isoimmunization
  - 11.6.5 Intrauterine death
  - 11.6.6 Failed induction
  - 11.6.7 Other obstetrical trauma
  - 11.6.8 Other and unspecified complications of birth, puerperium affecting management of mother
- 11.7 Normal pregnancy and/or delivery

- 11.7.1 Normal delivery
- 11.7.2 Multiple gestation
- 11.7.3 Outcome of delivery (V codes)

12 Diseases of the skin and subcutaneous tissue

- 12.1 Skin and subcutaneous tissue infections
  - 12.1.1 Cellulitis and abscess
    - 12.1.1.1 Cellulitis and abscess of fingers and toes
    - 12.1.1.2 Cellulitis and abscess of face
    - 12.1.1.3 Cellulitis and abscess of arm
    - 12.1.1.4 Cellulitis and abscess of hand
    - 12.1.1.5 Cellulitis and abscess of leg
    - 12.1.1.6 Cellulitis and abscess of foot
    - 12.1.1.7 Other cellulitis and abscess
  - 12.1.2 Other skin and subcutaneous infections
- 12.2 Other inflammatory condition of skin
- 12.3 Chronic ulcer of skin
  - 12.3.1 Decubitus ulcer
  - 12.3.2 Chronic ulcer of leg or foot
  - 12.3.3 Other chronic skin ulcer
- 12.4 Other skin disorders

13 Diseases of the musculoskeletal system and connective tissue

- 13.1 Infective arthritis and osteomyelitis (except that caused by TB or STD)
- 13.2 Non-traumatic joint disorders
  - 13.2.1 Rheumatoid arthritis and related disease
  - 13.2.2 Osteoarthritis
    - 13.2.2.1 Osteoarthritis, localized
    - 13.2.2.2 Osteoarthritis, generalized and unspecified
  - 13.2.3 Other non-traumatic joint disorders
- 13.3 Spondylosis, intervertebral disc disorders, other back problems
  - 13.3.1 Spondylosis and allied disorders
  - 13.3.2 Intervertebral disc disorders
  - 13.3.3 Other back problems
    - 13.3.3.1 Cervical radiculitis
    - 13.3.3.2 Spinal stenosis, lumbar region
    - 13.3.3.3 Lumbago
    - 13.3.3.4 Sciatica
    - 13.3.3.5 Thoracic or lumbosacral neuritis or radiculitis, unspecified
    - 13.3.3.6 Backache, unspecified
    - 13.3.3.7 Other back pain and disorders
- 13.4 Osteoporosis
- 13.5 Pathological fracture
- 13.6 Acquired deformities

- 13.6.1 Acquired foot deformities
- 13.6.2 Other acquired deformities
- 13.7 Systemic lupus erythematosus and connective tissue disorders
- 13.8 Other connective tissue disease
- 13.9 Other bone disease and musculoskeletal deformities

14 Congenital anomalies

- 14.1 Cardiac and circulatory congenital anomalies
  - 14.1.1 Transposition of great vessels
  - 14.1.2 Tetralogy of Fallot
  - 14.1.3 Ventricular septal defect
  - 14.1.4 Atrial septal defect
  - 14.1.5 Endocardial cushion defects
  - 14.1.6 Pulmonary valve atresia and stenosis
  - 14.1.7 Aortic valve stenosis
  - 14.1.8 Patent ductus arteriosus
  - 14.1.9 Coarctation of aorta
  - 14.1.10 Pulmonary artery anomalies
  - 14.1.11 Cerebrovascular anomalies
  - 14.1.12 Other cardiac and circulatory congenital anomalies
- 14.2 Digestive congenital anomalies
  - 14.2.1 Esophageal atresia/tracheoesophageal fistula
  - 14.2.2 Pyloric stenosis
  - 14.2.3 Rectal and large intestine atresia/stenosis
  - 14.2.4 Hirshsprung's disease
  - 14.2.5 Other digestive congenital anomalies
- 14.3 Genitourinary congenital anomalies
  - 14.3.1 Undescended testicle
  - 14.3.2 Hypospadias and epispadias
  - 14.3.3 Obstructive genitourinary defect
  - 14.3.4 Other genitourinary congenital anomalies
- 14.4 Nervous system congenital anomalies
  - 14.4.1 Spina bifida
  - 14.4.2 Congenital hydrocephalus
  - 14.4.3 Other nervous system congenital anomalies
- 14.5 Other congenital anomalies
  - 14.5.1 Cleft palate without cleft lip
  - 14.5.2 Cleft lip with or without cleft palate
  - 14.5.3 Congenital hip dislocation
  - 14.5.4 All other congenital anomalies

15 Certain conditions originating in the perinatal period

- 15.1 Liveborn
- 15.2 Short gestation, low birth weight, and fetal growth retardation

- 15.3 Intrauterine hypoxia and birth asphyxia
- 15.4 Respiratory distress syndrome
- 15.5 Hemolytic jaundice and perinatal jaundice
- 15.6 Birth trauma
- 15.7 Other perinatal conditions
  - 15.7.1 Respiratory conditions of fetus and newborn, other than respiratory distress
  - 15.7.2 Infections specific to the perinatal period
  - 15.7.3 Endocrine and metabolic disturbances of fetus and newborn
  - 15.7.4 Other and unspecified perinatal conditions

16 Injury and poisoning

- 16.1 Joint disorders and dislocations, trauma-related
- 16.2 Fractures
  - 16.2.1 Fracture of neck of femur (hip)
  - 16.2.2 Skull and face fractures
  - 16.2.3 Fracture of upper limb
    - 16.2.3.1 Fracture of humerus
    - 16.2.3.2 Fracture of radius and ulna
    - 16.2.3.3 Other fracture of upper limb
  - 16.2.4 Fracture of lower limb
    - 16.2.4.1 Fracture of tibia and fibula
    - 16.2.4.2 Fracture of ankle
    - 16.2.4.3 Other fracture of lower limb
  - 16.2.5 Other fractures
    - 16.2.5.1 Fracture of vertebral column without mention of spinal cord injury
    - 16.2.5.2 Fracture of ribs, closed
    - 16.2.5.3 Fracture of pelvis
    - 16.2.5.4 Other and unspecified fracture
- 16.3 Spinal cord injury
- 16.4 Intracranial injury
  - 16.4.1 Concussion
  - 16.4.2 Other intracranial injury
- 16.5 Crushing injury or internal injury
- 16.6 Open wounds
  - 16.6.1 Open wounds of head, neck, and trunk
  - 16.6.2 Open wounds of extremities
- 16.7 Sprains and strains
- 16.8 Superficial injury, contusion
- 16.9 Burns
- 16.10 Complications
  - 16.10.1 Complication of device, implant or graft

- 16.10.1.1 Malfunction of device, implant, and graft
- 16.10.1.2 Infection and inflammation--internal prosthetic device, implant, and graft
- 16.10.1.3 Other complications of internal prosthetic device, implant, and graft
- 16.10.1.4 Complications of transplants and reattached limbs
- 16.10.2 Complications of surgical procedures or medical care
- 16.10.2.1 Cardiac complications
- 16.10.2.2 Respiratory complications
- 16.10.2.3 Gastrointestinal complications
- 16.10.2.4 Urinary complications
- 16.10.2.5 Hemorrhage or hematoma complicating a procedure
- 16.10.2.6 Postoperative infection
- 16.10.2.7 Other complications of surgical and medical procedures

16.1 Poisoning

- 16.11.1 Poisoning by psychotropic agents
- 16.11.2 Poisoning by other medications and drugs
- 16.11.3 Poisoning by nonmedicinal substances

16.1 Other injuries and conditions due to external causes

17 Symptoms, signs, and ill-defined conditions and factors influencing health status

17.1 Symptoms, signs, and ill-defined conditions

- 17.1.1 Syncope
- 17.1.2 Fever of unknown origin
- 17.1.3 Lymphadenitis
- 17.1.4 Gangrene
- 17.1.5 Shock
- 17.1.6 Nausea and vomiting
- 17.1.7 Abdominal pain
- 17.1.8 Malaise and fatigue
- 17.1.9 Allergic reactions

17.2 Factors influencing health care

- 17.2.1 Rehabilitation care, fitting of prostheses, and adjustment of devices
- 17.2.2 Administrative/social admission
- 17.2.3 Medical examination/evaluation
- 17.2.4 Other aftercare
- 17.2.5 Other screening for suspected conditions (not mental disorders or infectious disease)

APPENDIX B

MULTUM NDC CLASSIFICATION (DOWNLOADED FEBRUARY 2002)

Multum Class	Class Description
2	amebicides
3	anthelmintics
8	carbapenems
10	leprostatics
11	macrolides
12	miscellaneous antibiotics
14	quinolones
15	sulfonamides
16	tetracyclines
17	urinary anti-infectives
18	aminoglycosides
21	alkylating agents
22	antibiotics/antineoplastics
23	antimetabolites
24	hormones/antineoplastics
25	miscellaneous antineoplastics
26	mitotic inhibitors
27	radiopharmaceuticals
30	antitoxins and antivenins
31	bacterial vaccines
32	colony stimulating factors
33	immune globulins
34	in vivo diagnostic biologicals
36	recombinant human erythropoietins
37	toxoids
38	viral vaccines
39	miscellaneous biologicals
41	agents for hypertensive emergencies
42	angiotensin converting enzyme inhibitors
43	antiadrenergic agents, peripherally acting
44	antiadrenergic agents, centrally acting
45	antianginal agents
46	antiarrhythmic agents
48	calcium channel blocking agents
50	inotropic agents
51	miscellaneous cardiovascular agents
52	peripheral vasodilators
53	vasodilators
54	vasopressors
55	antihypertensive combinations
56	angiotensin II inhibitors
59	miscellaneous analgesics
60	narcotic analgesics
61	nonsteroidal anti-inflammatory agents
62	salicylates
63	analgesic combinations
68	barbiturates
69	benzodiazepines
70	miscellaneous anxiolytics, sedatives and hypnotics

71	CNS stimulants
72	general anesthetics
74	neuromuscular blocking agents
76	miscellaneous antidepressants
77	miscellaneous antipsychotic agents
79	psychotherapeutic combinations
80	miscellaneous central nervous system agents
84	heparin antagonists
85	miscellaneous coagulation modifiers
86	thrombolytics
88	antacids
89	anticholinergics/antispasmodics
90	antidiarrheals
91	digestive enzymes
92	gallstone solubilizing agents
93	GI stimulants
94	H2 antagonists
95	laxatives
96	miscellaneous GI agents
98	adrenal cortical steroids
100	miscellaneous hormones
102	oral contraceptives
103	thyroid drugs
104	immunosuppressive agents
106	antidotes
107	chelating agents
108	cholinergic muscle stimulants
109	local injectable anesthetics
110	miscellaneous uncategorized agents
111	psoralens
112	radiocontrast agents
116	iron products
117	minerals and electrolytes
118	oral nutritional supplements
119	vitamins
120	vitamin and mineral combinations
121	intravenous nutritional products
123	antihistamines
124	antitussives
126	methylxanthines
127	decongestants
128	expectorants
129	miscellaneous respiratory agents
130	respiratory inhalant products
131	antiasthmatic combinations
132	upper respiratory combinations
134	anorectal preparations
135	antiseptic and germicides
137	topical anti-infectives
138	topical steroids
139	topical anesthetics

140	miscellaneous topical agents
141	topical steroids with anti-infectives
143	topical acne agents
144	topical antipsoriatics
146	mouth and throat products
149	spermicides
150	sterile irrigating solutions
153	plasma expanders
154	loop diuretics
155	potassium-sparing diuretics
156	thiazide diuretics
157	carbonic anhydrase inhibitors
158	miscellaneous diuretics
159	first generation cephalosporins
160	second generation cephalosporins
161	third generation cephalosporins
162	fourth generation cephalosporins
163	ophthalmic anti-infectives
164	ophthalmic glaucoma agents
165	ophthalmic steroids
166	ophthalmic steroids with anti-infectives
167	ophthalmic anti-inflammatory agents
168	ophthalmic lubricants and irrigations
169	miscellaneous ophthalmic agents
170	otic anti-infectives
171	otic steroids with anti-infectives
172	miscellaneous otic agents
173	HMG-CoA reductase inhibitors
174	miscellaneous antihyperlipidemic agents
175	protease inhibitors
176	NRTIs
177	miscellaneous antivirals
178	skeletal muscle relaxants
179	skeletal muscle relaxant combinations
180	adrenergic bronchodilators
181	bronchodilator combinations
182	androgens and anabolic steroids
183	estrogens
184	gonadotropins
185	progestins
186	sex hormone combinations
187	miscellaneous sex hormones
191	narcotic analgesic combinations
192	antirheumatics
193	antimigraine agents
194	antigout agents
195	5HT3 receptor antagonists
196	phenothiazine antiemetics
197	anticholinergic antiemetics
198	miscellaneous antiemetics
199	hydantoin anticonvulsants

200	succinimide anticonvulsants
201	barbiturate anticonvulsants
202	oxazolidinedione anticonvulsants
203	benzodiazepine anticonvulsants
204	miscellaneous anticonvulsants
205	anticholinergic antiparkinson agents
208	SSRI antidepressants
209	tricyclic antidepressants
210	phenothiazine antipsychotics
211	platelet aggregation inhibitors
212	glycoprotein platelet inhibitors
213	sulfonylureas
214	non-sulfonylureas
215	insulin
216	alpha-glucosidase inhibitors
217	bisphosphonates
219	nutraceutical products
220	herbal products
222	penicillinase resistant penicillins
223	antipseudomonal penicillins
224	aminopenicillins
225	beta-lactamase inhibitors
226	natural penicillins
227	NNRTIs
228	adamantane antivirals
229	purine nucleosides
230	aminosalicylates
231	nicotinic acid derivatives
232	rifamycin derivatives
233	streptomyces derivatives
234	miscellaneous antituberculosis agents
235	amphotericins
236	azole antifungals
237	miscellaneous antifungals
238	aminoquinolones
239	miscellaneous antimalarials
240	lincomycin derivatives
241	fibric acid derivatives
243	leukotriene modifiers
244	nasal lubricants and irrigations
245	nasal steroids
246	nasal antihistamines and decongestants
248	topical emollients
250	monoamine oxidase inhibitors
252	bile acid sequestrants
253	anorexiants
256	interferons
257	monoclonal antibodies
261	heparins
262	coumarins and indandiones
263	impotence agents

264	urinary antispasmodics
265	urinary pH modifiers
266	miscellaneous genitourinary tract agents
267	ophthalmic antihistamines and decongestants
268	vaginal anti-infectives
269	miscellaneous vaginal agents
270	antipsoriatics
271	thiazolidinediones
272	proton pump inhibitors
273	lung surfactants
274	cardioselective beta blockers
275	non-cardioselective beta blockers
276	dopaminergic antiparkinsonism agents
277	5-aminosalicylates
278	cox-2 inhibitors
279	gonadotropin releasing hormones
280	thioxanthenes
281	neuraminidase inhibitors
282	meglitinides
283	thrombin inhibitors
284	viscosupplementation agents

* Not all class numbers are sequential due to former classes being dropped or broken out into multiple classes