Medicare Calibration of the Clinically Detailed Risk Information System for Cost

Kanika Kapur, Ph.D., Chien-Wen Tseng, M.D., Afshin Rastegar, M.S., Grace M. Carter, Ph.D., and Emmett Keeler, Ph.D.

The clinically detailed risk information system for cost (CD-RISC) contains definitions for several hundred severity-adjusted conditions that can be used to predict future health care costs. We develop a prospective Medicare CD-RISC model using a 5-percent sample of Medicare beneficiaries and data that contain 1996 diagnostic information and 1997 annualized costs. The CD-RISC model has a hierarchical structure that implies that only the most expensive condition-severity variable within a body system affects payments. This minimizes incentives to game the system by entering multiple related codes for the same condition. The R^2 for the CD-RISC model is 11 percent.

INTRODUCTION

The Medicare+Choice (M+C) capitation provisions of the Balanced Budget Act (BBA) of 1997 set a deadline of January 2000 for the implementation of a new risk-adjustment method. The risk-adjustment method CMS used in 2000 was based on the primary inpatient diagnostic cost group (DCG) model (Pope et al., 2000a). However, because the model uses only inpatient data, (1) it creates an incentive for M+C plans to hospitalize enrollees to get them assigned to a higher DCG category; and (2) is not sensitive to costs for enrollees who are chronically ill, but are not hospitalized. Health maintenance organizations (HMOs) that achieve cost savings by substituting other forms of care for inpatient care are particularly vulnerable to this flaw in the model. Thus, in 2004, CMS is planning to move to a model that uses both inpatient and outpatient data to identify diagnostic codes for grouping beneficiaries into risk groups (Centers for Medicare & Medicaid Services, 2003).

Recently much effort has gone into developing risk-adjustment systems to modify capitation payments for expected use (Ellis et al., 1996; Pope et al., 2000b; Kronick et al., 2002; 2000; and 1996; Ash et al., 2000; Weiner et al., 1996). All of these risk-adjustment methods tie payments to expected resource use based on diagnostic codes. Two of the models are derived from the DCG family of models, one is based on the adjusted clinical groups case-mix system, and another uses the chronic illness and disability payment system (CDPS). Another model is based on CD-RISC. All of these models share a basic approach. First, the *International Classification of Diseases, Ninth Revision, Clinical Modification* (ICD-9-CM) (Centers for Disease Control and Prevention, 2003), diagnosis codes are grouped into categories based on clinical judgment and experience, yielding relatively small, clinically homogenous categories of beneficiaries. Then statistical methods combine that information with expenditure

data to predict costs and generate payment levels. However, the models differ in specific details.

In an earlier study, RAND researchers developed the CD-RISC model, and tested this model on Medicaid and private sector managed care enrollees (Carter et al., 1997). In this article, we describe the development and testing of the CD-RISC model for the Medicare population. The system describes the resources needed to care for the Medicare population based on their burden of disease. It contains prospective models that predict annual costs based only on information known at the start of the year. The diagnoses are grouped into conditions with an attached severity level. The condition-severity groups are then organized into body systems with only the most expensive group in each body system affecting prediction. However, other conditions in both the same and other body systems may affect severity level, and thus differences in the effects of severity levels of the same condition often measure the interaction of the condition with other conditions. The structure of the CD-RISC model is designed to minimize incentives to game the system. M+C plans may game a risk-adjusted payment system in several ways. For example, if vague condition codes, such as abdominal pain were included in the payment model, providers may find it relatively easy to over use such a code. Furthermore, if multiple related condition codes, such as abdominal pain and ulcers, were included in a payment model, providers may have incentive to code both conditions to increase their payment. A risk-adjustment system may also be gamed if unnecessary care is provided in order to obtain additional payment.

The CD-RISC model presented here is a prospective model that is based solely on diagnoses recorded before the start of the payment year. Using retrospective data, such as current year utilization or current year diagnostic information, to set payments has advantages and disadvantages. The advantages are that a model that uses retrospective variables increases incentives to provide needed care and increases the accuracy of prediction. Retrospective data increase the accuracy of the model by including acute episodes and chronic diseases that are diagnosed for the first time during the year. This greater accuracy should reduce incentives to select healthier patients into a plan. Since physicians value providing good care, episodes of illness and outlier payments can encourage an increase in the provision of care to vulnerable populations. Retrospective payments are also likely to ease providers' financial risk, and therefore, should increase incentives for providers to remain in the Medicare market.

In general, including retrospective data adds incentives to provide unnecessary services and transfers part of the risk back to the payer (Ellis et al., 1996). Payment schemes based on retrospective data may increase the data burden if the payment is based on actual charges. However, methods that do not increase the data burden can be designed. For example, payments based on diagnosis-related groups (DRGs) may not necessarily increase the data collection burden. Providers already collect and maintain DRG information, and therefore, the partial capitation would not add any additional burden.

If people stay with the same health care plans for a moderate length of time, adjusting payments based on prior utilization would still provide plans with some incentive to provide care. This is not a flaw, since for both fairness (unmeasured sickness) and efficiency (to give plans less incentive to stint on care) reasons, it may be desirable to have providers and payers share

the risks of care for expensive patients (Keeler, Carter, and Newhouse, 1998). If data on current costs are available, it is easy to modify risk-adjustment systems that yield predictions of prospective costs to pay a blend of those predicted costs and current costs.

DATA

The data files used in the analysis for the Medicare sample were created by Health Economics Research (HER) in the course of developing the hierarchical conditions model (HCC) for the Medicare population (Pope et al., 2000b). In this article, we will review HER's file creation and describe the variables to be used in the analysis.

Sample Selection

The analytic file was created by HER from the 1996 and 1997 Medicare 5-percent sample standard analytic files, by restricting the data to individuals who satisfied the following conditions:

- Continuously enrolled in both Medicare Part A and Part B from January 1, 1996 to December 31, 1997 or until death if deceased in 1997.
- At least 1 month in 1997 entitled by age or disability, not residing in a hospice, and not enrolled in a HMO.
- No months of HMO enrollment in 1996.
- U.S. residence throughout 1996 and 1997.
- No months of working aged status in either 1996 or 1997.

These sample restrictions aim to ensure that the Medicare enrollees in the final sample have complete diagnostic information in 1996 and complete claims data in 1997. Only months of fee-for-service eligibility in 1997 were included in the sample—all HMO months were excluded because of incomplete claims data.

The 1997 Medicare 5-percent standard analytic files contain 2,017,964 beneficiaries with any form of Medicare eligibility. The prospective sample, used in the analyses in this article, was created by dropping beneficiaries who did not satisfy the five conditions listed, and contains 1,394,701 beneficiaries (Pope et al., 2000b).

Medicare Expenditures

The dependent variable for the analysis is Medicare expenditures for 1997. This variable includes all Medicare payments other than hospice payments and indirect medical education payments. HMOs are not responsible for hospice care, and indirect medical education is not paid to HMOs; therefore, these categories are excluded from the constructed measure of Medicare expenditures.

Medicare payments are summed only for the months in 1997 that the beneficiary met the sample selection criteria. After payments are summed, they are annualized by dividing the payments by the fraction of months in 1997 that each beneficiary was eligible. This fraction is also used to weight all analyses.

Demographic and Other Beneficiary Information

The age and sex information used in the CD-RISC model consists of 24 age-sex cells. There are 12 age cells for males: 0-34; 35-44, and so on with the oldest category being 95 or over. Similarly, there are 12 age cells for females: 0-34; 35-44, and so on with the oldest category being 95 or over.[1] Individuals who are under age 65 are entitled to Medicare through disability. Originally disabled status is another beneficiary variable used in the CD-RISC

---

[1] When an individual spends part of 1997 in one age category and the rest in another category, the fraction of time spent in each category is used to prorate the age-sex cells.

model. Originally disabled beneficiaries were entitled to Medicare before age 65 due to their disability status, but in 1997 are entitled on the basis of their age.[2]

The CD-RISC model also uses information on Medicaid eligibility in 1996. A beneficiary is defined to have Medicaid eligibility if he or she had any months of Medicaid eligibility in 1996.

Diagnostic Information

In the prospective CD-RISC models developed in this article, diagnostic information from 1996 is used to predict 1997 Medicare expenditures. We constructed the diagnostic profile for Medicare beneficiaries using ICD-9-CM codes assigned to claims from the following sources: hospital inpatient principal diagnoses, hospital inpatient secondary diagnoses, hospital outpatient department, physician and clinically-trained non-physician. Examples of clinically-trained non-physicians are psychologists, therapists, and podiatrists. We did not include additional diagnoses from home health, durable medical equipment (DME), and other facility types since the clinical validity of these diagnoses is suspect and including these sources would increase the burden of data collection. Although our main analysis includes diagnoses that were obtained from radiologists, anesthesiologists and pathologists (RAP), in a specification check discussed later, we excluded diagnoses that were obtained from RAP sources due to concerns about their clinical validity.

Sample Summary Statistics

Table 1 contains summary statistics on the prospective sample used to construct the CD-RISC Medicare model. This table shows the percent of the sample in each of the 24 age-sex cells, the percent that are originally disabled (but now aged), and the percent that are enrolled in Medicaid. Average annualized Medicare expenditures for each of these groups is also provided.

MODEL DEVELOPMENT

We developed the Medicare CD-RISC by updating the ICD-9-CM codes used in the original CD-RISC model and by adjusting the definitions of condition and severity for the Medicare population. We describe the three elements used to construct the CD-RISC model: (1) these are the construction of conditions; (2) the assignment of severity levels to conditions; and (3) the assignment of condition-severity combinations to body systems.

Construction of Conditions

A condition is a clinical construct operationalized as a grouping of ICD-9-CM diagnostic codes. Codes were grouped together when the referenced conditions were similar on: pathophysiology, duration, and the tests and services required to diagnose, monitor, and treat. Each ICD-9-CM code is assigned to at most one condition (Carter et al., 1997). The condition mapping was based initially on the subjective judgment of physician panels organized by Value Health Sciences[3] as part of the practice review system. Examples of conditions include breast cancer, diabetes, urinary tract infection, and hypertension.

We focused on developing conditions that were likely to be specifically important for the disabled age 65 or over Medicare population. We subdivided existing condition codes into two or more conditions when we believed that costs for the Medicare population may be systematically different

---

[2] As with the age-sex variables, originally disabled status is prorated if a beneficiary spent part of the year entitled due to disability and the rest of the year entitled due to age.

[3] Value Health Science is a company that provides health care information services.

Table 1

Summary Statistics on the Prospective Sample Used to Construct the Clinically Detailed Risk Information System for Cost Medicare Model: 1996-1997

Variable	Percent of Sample	Annualized 1997 Medicare Payment
Male	41.4	$5,408
Male-Disabled	6.9	4,291
0-34 Years	0.8	3,298
35-44 Years	1.6	3,851
45-54 Years	2.0	4,064
55-59 Years	1.1	4,695
60-64 Years	1.3	5,396
Male-Aged	34.5	5,633
65-69 Years	7.9	4,070
70-74 Years	10.4	4,848
75-79 Years	8.0	6,040
80-84 Years	5.0	7,211
85-89 Years	2.3	8,386
90-94 Years	0.8	9,228
95 Years or Over	0.2	8,484
Female	58.6	5,247
Female-Disabled	4.9	5,007
0-34 Years	0.5	3,669
35-44 Years	1.0	4,236
45-54 Years	1.4	4,813
55-59 Years	0.9	5,316
60-64 Years	1.0	6,274
Female-Aged	53.7	5,269
65-69 Years	9.8	3,569
70-74 Years	14.0	4,249
75-79 Years	12.1	5,292
80-84 Years	9.0	6,385
85-89 Years	5.5	7,463
90-94 Years	2.5	8,106
95 Years or Over	0.9	7,444
Disabled	11.9	4,588
Originally Disabled (Now Aged)	6.6	8,224
Medicaid Eligible in 1996	15.0	7,277
Full Sample	100.0	5,314

NOTES: The variables represented are not mutually exclusive. Therefore, the numbers in the percent of sample may not add to totals because of rounding. Sample consisted of 1,394,701 observations.

SOURCE: Centers for Medicare & Medicaid Services: Data from the 5 Percent Sample Standard Analytical Files.

within a condition code. For example, the condition code other renal disease was subdivided into new condition codes urinary incontinence and renal dialysis, because dialysis is likely to cost significantly more than treatment for urinary incontinence.

In developing the CD-RISC model, we were careful to exclude conditions that may be vague. While several conditions are labeled as other conditions, these groups include clearly defined ICD-9-CM codes that were grouped together on the basis of clinical judgment.

Assignment of Severity Levels

Conditions can have up to three severity levels: usual or low, medium, and high. Conditions are suffixed by the letters "L" for low or usual severity, "M" for medium severity, and "H" for high severity. Severity

levels for condition codes were determined by physician panels using clinical criteria to judge the expected resource use for the condition. As with condition assignment, the original severity assignment was developed as part of the practice review system (Carter et al., 1997).

When clinical judgment suggests that a certain condition is relatively heterogeneous in expected resource use, the condition is divided into severity levels that capture the expected resource use. In general, if a condition is homogenous in expected resource use, the condition is not divided into severity levels, and is simply assigned one severity of L. As with condition assignment, the practice review system originally determined the condition interactions in the CD-RISC. Physician panels determined if the presence of conditions in other body systems was likely to increase the severity of a condition code. Condition-severity codes were grouped with other similar codes on the basis of pathophysiology, duration, and the tests and services required to diagnose, monitor, and treat the condition.

The clinical information used to assign the severity levels for any one condition code can come from both the ICD-9-CM codes that fall into that condition code, and from other ICD-9-CM codes outside of that condition code that represent complications or comorbidities that increase the resources required to care for a condition. Thus, one ICD-9-CM code can affect the severity level of more than one condition. After all conditions have been assigned, the set of all ICD-9-CM codes for the patient are searched a second time to determine severity. An individual is always assigned to the highest severity level for which he or she qualifies.

To clarify how we assign severity levels to condition codes, we provide examples for the condition of diabetes. Severity level can be assigned from the diabetes ICD-9-CM codes, for instance a diagnosis of Diabetes without Complications uncontrolled (ICD-9-CM 250.03) gives the condition diabetes a moderate severity, while Diabetes with Coma (ICD-9-CM 250.2), gives the condition diabetes a high severity. A severity level can also be assigned to the condition Diabetes from Non-Diabetic ICD-9-CM codes that represent complications or comorbidities of diabetes and give insight to the severity of the illness. For instance, one of the Complications of Diabetes is an increased risk for a Stroke, so a diabetes ICD-9-CM code in combination with an ICD-9-CM code for Cerebral Vascular Accident or Stroke (ICD-9-CM 436) would move the condition diabetes to a high severity level. The effect of an ICD-9-CM code on severity level is unilateral unless otherwise clinically indicated, i.e., it does not necessarily follow that a diabetes ICD-9-CM code would change the severity level assigned to the condition code Cerebral Vascular Disease.

Body Systems

The combinations of condition and severity are organized into hierarchies within 19 body systems so that the model prediction for each patient depends on at most one combination of condition and severity per body system. This reduces incentives for multiple coding of the same or related conditions. Additional coding within the same body system is likely to be more frequent than additional coding between body systems in cases where diagnoses are ambiguous and hence subject to gaming (Carter et al., 1997). The Medicare CD-RISC model contained the following body systems: infection, cancer, metabolic, blood, mental, neurological, cardiovascular, ear/nose/throat (ENT), respiratory, gastrointestinal, renal, reproductive,

skin, muscular, injury, and newborn. The pregnancy and other body systems were dropped due to statistical insignificance.

The body systems are based on the ICD-9-CM coding structure. Similar hierarchical systems were used in the original DCG risk-adjustment model within disease, but the DCG-HCC model allows multiple diseases within the same body system to affect costs. For example, in the DCG-HCC model, individuals with human immunodeficiency virus/acquired immunodeficiency syndrome (HIV/AIDS) may also be coded as having septicemia or other opportunistic infections, but may not be coded as having the HCC code other infectious diseases. In contrast, in the CD-RISC model, since HIV/AIDS, septicemia, opportunistic infections, and other infectious diseases all belong to the infection body system, only the most costly condition within this body system will affect individual payments. In addition, the hierarchy used in the DCG-HCC model is based on a mix of costs and clinical judgment, whereas the hierarchy in the CD-RISC model is almost entirely empirically determined (Pope et al., 2000b). However, there are some exceptions to the empirical determination of the CD-RISC hierarchy. For example, we constrained the CD-RISC payments so that the low severity level of a condition could not have a higher payment than the high severity level of the same condition.

Within each body system, an ordering of conditions is established and each patient is assigned to the most expensive relevant condition-severity combination. The ranking of condition-severity combinations within each body system is in order of their costs as determined from the CD-RISC model. The rankings were derived using an iterative procedure that will be explained later in this article.

In developing the Medicare CD-RISC, we amended the original CD-RISC 16 body systems (Carter et al., 1997) by dividing the genital/urinary body system into two body systems—genital and renal/urinary. Genital and renal/urinary were separated into two body systems because older people often have separate problems in both body systems, for instance, sexual dysfunction (genital) and urinary incontinence (renal/urinary). By creating two body systems, these common diagnoses would be both available for predicting costs. The body system respiratory was also divided into ENT and respiratory based on similar reasoning. We experimented with adding a new body system, long-term care, that we believed captured conditions prevalent in the age 65 or over Medicare population. Most of the conditions from the long-term care body system were dropped from the analysis during model development because of concerns about gameability. For example, the ICD-9-CM codes for physical or speech therapy that are part of the long-term care body system may encourage the provision of unneeded care.

In developing the Medicare CD-RISC system, we aimed to avoid placing big ticket unrelated ICD-9-CM codes within the same body system or condition category. The ICD-9-CM codes for both ischemic heart disease and stroke were originally assigned to the cardiovascular body system. This is technically valid since both heart disease and stroke involve the cardiovascular body system, but it forced the model to use cost information from only one of two distinct major illnesses common to older Medicare beneficiaries. Stroke was moved to the neurological body system since most of the consequences of stroke are neurological.

In our adjustments to the CD-RISC mapping, we were also careful to control the incentive to game the system by grouping codes for the same problem (e.g., stomach pain and peptic ulcer) into the same body system, and restricting the use of condition codes that were vague and hence, potentially gameable.

STATISTICAL METHODS AND RESULTS

Model Definitions

We used regression analysis to determine the amount of payment that would be made for each person in the sample. Specifically, we estimated weighted linear regression models with annualized 1997 costs as the dependent variable. The estimation sample consists of 1,394,701 observations on Medicare beneficiaries. We did not transform costs—models using untransformed costs are standard in the field of risk adjustment, primarily because the large sample size used in risk adjustment estimation reduces the influence of outliers in the distribution on the estimates.

The set of explanatory variables used in the models included 24 age-sex indicators, an indicator for originally disabled, Medicaid status, and the CD-RISC condition-severity variables. The models were weighted by the fraction of the year that the beneficiary met the sample eligibility criteria. Weighting prevents bias in estimating annual payments for conditions where a substantial number of individuals died in the prediction year. Since expenses tend to rise immediately before death, annualizing expenses of individuals that died over-estimates the average monthly expenses for patients with these conditions relative to patients with conditions where no one or few died. Weighting also improves the efficiency of the models by down-weighting observations that have a higher variance associated with their costs due to partial eligibility in 1997.

We report the predictive power for three regression models: (1) demographic, (2) basic CD-RISC, and (3) CD-RISC with disabled interactions. The explanatory variables in the demographic model consist of the 24 age-sex indicators, an indicator for originally disabled, and Medicaid status. In the second model, we add indicators for condition-severity variables. In the third model, we explored adding interaction variables between disabled status and the condition-severity variables. It is plausible that certain conditions for disabled beneficiaries may require different payments for the same conditions for beneficiaries age 65 or over. For example, older people may commonly have simple urinary infections that are easily treated with one antibiotic. However, younger people with disabilities may be more vulnerable to complicated urinary tract infections that require multiple antibiotics or longer treatment because of underlying renal failure or a history of repeated infections leading to antibiotic resistance. Therefore, we included disabled interactions in the CD-RISC model when coefficients were substantially different for disabled and age 65 or over subsamples. We used the following criteria to determine which condition-severity variables were to be interacted with disability:

- The t-statistic of the difference between the age 65 or over, and the disabled coefficient must be greater than 2.
- The cell sizes for the disabled and the age 65 or over groups for the condition-severity variable should both be greater than 100.
- Disabled or age 65 or over condition-severity coefficients with t-statistics that are less than 1.75 are dropped from the model unless they are to be combined with adjacent severity levels according to condition-severity pruning rules.

Establishing the Hierarchy

In each body system, patients are assigned to the single most expensive condition-severity combination for which diagnoses were recorded. The hierarchy for condition-severity variables within body systems was established using an iterative weighted linear regression procedure.

In order to select the highest condition-severity coefficient within a body system, we needed to obtain initial ranks for these condition-severity codes. Therefore, we estimated a weighted regression model of annualized 1997 costs with 473 indicators for each of the condition-severity variables for the full sample of 1,394,701 observations. The age-sex indicators, Medicaid and original disability status variables, were not included in this model. We used the estimates from this model to create initial ranks that sorted condition-severity variables by coefficient size within body system. Using these ranks, we selected the highest ranked condition-severity variable within each body system for each individual. We then ran the first iteration of the regression model using this ranking assignment. The condition-severity coefficients from the first iteration were reordered by sorting coefficients within each body system. This new ordering formed the ranks for the condition-severity variables for the second iteration. This process was repeated until the ranks for the condition-severity variables did not change in consecutive iterations.

This iterative algorithm may not necessarily converge. If there are two condition-severity groups that have similar average costs and, by chance, people in the sample with both conditions have lower costs than those with only one condition, the ranks for the two conditions will alternate. Such cycles occurred in our data with six sets of condition-severity variables. We solved the problem of alternating ranks by creating a frequency-weighted average of the coefficients associated with the two conditions and paying this same amount for the pair of alternating conditions. This method works well because the differences between the coefficients for each pair of conditions were small.

As a simplified example of establishing the hierarchy, suppose an ordinary least squares model is estimated for only the ENT body system, and we find that the conditions are ordered by cost in the following way: (1) tonsils/adenoids disease, (2) allergy/hayfever, (3) epistaxis/throat hemorrhage, and (4) larynx problem. Using this information, we would impose a hierarchy so that individuals are only coded with the most costly ENT condition that they have. We would then rerun the model. If we found that the ordering of these four conditions did not change, the coefficients are deemed to have converged.

Pruning the Condition-Severity Variables

While developing the CD-RISC model with the full set of 473 condition-severity variables, we discovered that several of these variables were statistically insignificant and a few were even negative. Negative variables could not be included in a payment model for reasons of face validity and because negative coefficients create incentives not to code diagnoses. To address these issues, we implemented the condition pruning rules laid out for the CD-RISC model (Carter et al., 1997). Specifically, all condition-severity variables with a t-statistic of less than 1.75 are dropped from the model, unless there was a lower severity condition with a t-statistic

that is greater than or equal to 1.75.[4] For example, during the model iterations, we found that high severity alcohol use disorder had a coefficient of 646 and a t value of 0.9. However, this condition severity was not dropped since the low- and medium-severity levels of alcohol use disorder were positive and had t-values that exceeded 1.75. After implementing these pruning rules, we combined condition-severity variables with t-statistics of less than 1.75 with adjacent severity levels. In cases where there was a choice of two adjacent severity levels (for example, medium severity could be combined with either low severity or high severity), the two severities with coefficients that were closest in magnitude were combined.

Severity levels added another facet to the pruning rules. We expected that higher severity conditions would predict higher costs than lower severities for the same condition. However, contrary to clinical judgment, the expected order was not observed for a few conditions. Allowing the higher severity level for a condition to have a lower payment than the low-severity level may create perverse coding incentives. Therefore, we enforced a monotonicity criterion on the condition-severity variables. Specifically, we combined adjacent severity levels for a condition if a higher severity had a lower coefficient than a lower severity level. In cases where severity levels for a condition were combined, the condition number for the new condition-severity variable was suffixed by the concatenated severity levels. For example, "LMH" denotes that low, medium, and high severities were combined for the referenced condition.

Setting Final Payments

Multiple regression can sometimes lead to clearly unreasonable predictions for unusual subsamples. After examining predictions from the CD-RISC regression model with disabled interactions, we found that 0.05 percent of the predictions were negative, 0.3 percent were below $100, and 0.8 percent were below $300. In most cases, the low payments apply to beneficiaries who have no conditions. In response, we estimated a simple demographic model on the sample of individuals who have no conditions in the payment model. The results were used in a two-step process to set the final payments that would not be unreasonably low. First, we calculate the maximum of the prediction from the CD-RISC model and the prediction from the no-condition demographic model. This no-condition demographic model has 24 age-sex cells, originally disabled status, and Medicaid eligibility as independent variables. The dependent variable in this model is 1997 annualized costs. The model is estimated on the sample of beneficiaries who have no conditions in the payment model.[5] Next, we adjust the maximum of the prediction from the CD-RISC and no-condition model by applying a proportionate tax to maintain budget neutrality. The tax, which is calculated as the sum of 1997 annualized Medicare payments for the full sample divided by the sum of maximum of the predictions from the two models, is approximately 99 percent.

By setting the payment for each beneficiary to the maximum of the predictions from these two models (less the proportionate

---

[4] In some cases, high-severity condition variables have small cell sizes, and are therefore, statistically insignificant.

[5] Model coefficients for the no-condition demographic model are not reported in this article, but are available on request from the author.

Table 2

Clinically Detailed Risk Information System for Cost (CD-RISC) Model Fit: 1996-1997

Model	R^2
Demographic	0.0169
CD-RISC	
Without Disabled Interactions	0.1108
With Disabled Interactions	0.1113
With Disabled Interactions and Minimum Payments	0.1114
With Disabled Interactions and Minimum Payments (Conservative R^2)	0.1108

NOTE: R^2 is a measure of the model's explanatory power.

SOURCE: Centers for Medicare & Medicaid Services: Data from the 5 Percent Sample Standard Analytical Files.

tax), we ensure the payment for each beneficiary equals or exceeds the average Medicare expenditure for beneficiaries with no conditions and the same demographic makeup.

RESULTS

Table 2 presents the R^2 values from the CD-RISC models. The R^2 from the demographic model is only 0.0169, but is comparable to the R^2 values obtained for other demographic models in the literature. Note that we include originally disabled status and Medicaid eligibility in this model; these variables are not included in DCG-HCC age-sex model, hence this model obtained a lower R^2 of 0.0097 (Pope et al., 2000b). The Medicare Current Beneficiary Survey yielded a demographic R^2 of 0.0045 (Pope et al., 1998). The demographic models for the CDPS had R^2 values that ranged from 0.004 to 0.015, across the States among the Medicaid disabled (Kronick et al., 1996).

The R^2 values based on the basic CD-RISC model that includes condition-severity indicators is 0.1108. Inclusion of disabled interactions has a small effect on the R^2, increasing it to 0.1113. Therefore, it appears that the inclusion of disabled interactions does not appreciably increase predictive power. However, including these interactions is likely to be important in a payment model to avoid perverse selection incentives. We find that the Efron (1978) R^2

from the model including minimum payments is 0.1114—this value is almost identical to the R^2 from the model without these payments.

We use the Efron R^2 in the calculation of predictive accuracy for the CD-RISC model with minimum payments. The Efron R^2 is merely:

$1-\text{sum}((\text{cost}_i-\text{prediction}_i)^2)/\text{sum}--((\text{cost}_i-\text{cbar})^2)$, where cbar is the average cost for the sample and the summations are over all members i of the sample (Efron, 1978).

The CD-RISC model R^2 is comparable to the R^2 values obtained from other risk-adjustment models. The DCG-HCC model obtained an R^2 of 11.2 percent for the same data. Not surprisingly, the R^2 for the primary inpatient DCG model that uses only inpatient diagnoses is much lower—only 6.2 percent (Pope et al., 2000a). Since the R^2 is driven by the predictability of the sample, it can be misleading to compare R^2 value statistics from different data sets (Hadorn et al., 1993). The R^2 for the chronic illness and disability payment system is 18 percent for the disabled Medicaid sample, but only 8 percent for Aid to Families with Dependent Children adults, and 4 percent for children (Kronick et al., 2000).

Table 3 presents selected condition-severity coefficients from the CD-RISC model that includes disabled interactions. [5] When an interaction is included for a condition-severity variable, the coefficient for disabled

_____

[5] A full set of coeffficents for the model is available on request from the author.

Table 3

Selected Clinically Detailed Risk Information System (CD-RISC) for Cost: Model Estimates Including Disabled Interactions' 1996-1997

Body System	Severity	Payment Estimate	t-Statistic
Infection			
Flu/Virus	High(a)	841	2.96
Flu/Virus	High(d)	3,878	5.72
HIV Infection	Medium or High	4,471	5.24
HIV Infection	Low(d)	3,750	9.55
Cancer			
Lung Cancer	High	10,846	40.38
Lung Cancer	Medium	3,737	11.54
Lung Cancer	Low	2,130	9.28
Metabolic			
Diabetes	High	4,208	55.98
Diabetes	Medium(a)	2,534	49.13
Diabetes	Medium(d)	1,661	11.32
Diabetes	Low	845	17.03
Blood			
Deficiency Anemia	Medium or High(a)	1,283	20.18
Deficiency Anemia	Medium or High(d)	2,972	13.02
Deficiency Anemia	Low	702	11.75
Mental Depression	Low-Medium-High	462	6.20
Cardiovascular			
Ischemic Heart Disease/Angina	Low-Medium-High	1,256	34.73
Respiratory			
Respirator/Aspirator Dependence	Low	13,091	20.89
Renal			
Urinary Tract Infection	High(d)	4,568	8.99
Urinary Tract Infection	Medium or High	895	9.32
Urinary Tract Infection	Low	632	15.18
Dialysis	Medium(a)	18,669	17.69
Dialysis	Medium(d)	9,261	6.76

NOTES: Sample consisted of 1,394,701 observations. Model parameters were 232. The full model includes 24 age-sex indicators, an indicator for originally disabled, Medicaid status, and 199 CD-RISC condition-severity variables. The suffix (a) to the condition severity code denotes an estimate for the aged sample and (d) denotes an estimate for the disabled sample. HIV is human immunodeficiency virus.

SOURCE: Centers for Medicare & Medicaid Services: Data from the 5 Percent Sample Standard Analytical Files.

beneficiaries is suffixed with "d" and the coefficient for aged beneficiaries is suffixed with "a."[6] The coefficients from Table 3 show that the most expensive conditions are dialysis (medium severity) in the renal body system at $18,669; respirator/aspirator dependence (usual/low severity) at $13,091; and lung cancer (high severity) at $10,846. Note that the disabled dialysis coefficient is $9,261 and the aged dialysis coefficient is $18,668, showing that payments can differ substantially between the disabled and the aged. Other examples of conditions that have disabled interactions include metastasized cancer (usual/low severity), paralysis (high severity), and diabetes (medium severity). In general, the conditions with disabled interactions have medium or high severity (in cases where multiple severity levels exist for the condition).

Payments are calculated from the model estimates by adding the relevant coefficients. For example, a male age 65 ($1,148) with a medium or high severity HIV infection ($4,471), high severity flu/virus ($0— same body system as HIV), and depression ($462) has a payment of $6,081.

SPECIFICATION AND VALIDITY CHECKS

We describe four model checks that we conducted in order to verify the robustness and validity of the CD-RISC model.

Overfitting and Conservative R^2

One concern with the CD-RISC model developed in this article is that it may be overfitted and may not perform well in other samples. This is unlikely, because of the large number of observations, but to address this concern, we calculated a conservative R^2 estimate based on the multiple cross-validation technique (Daley et al., 1988; Hadorn et al., 1993). To implement this technique, we randomly divided the sample into tenths. We then estimated coefficients for the CD-RISC model with disabled interactions for 90 percent of the data and applied them to the remaining 10 percent of the data to form model predictions. These predictions were corrected for payments that were too low by estimating the no-condition demographic model for the 90 percent subsample. This procedure involves estimating the CD-RISC model 10 times in order to get out-of-sample predictions for each of the 10 random subsamples of the data. After obtaining final predictions, in Table 2, we calculated an Efron R^2 of 0.1108. This R^2 is very similar to the R^2 values of 0.1114 obtained directly from the model, suggesting that the model does not suffer from an overfitting problem.

Hierarchical Model Structure

The CD-RISC model imposes a hierarchical structure on the condition-severity variables within the body systems. For each individual, only the condition-severity variable with the highest coefficient within each of the body systems is used in predictions. This restriction reduces incentives to game at some cost to predictive power. To test the impact of the hierarchical structure on predictive power, we re-estimated the CD-RISC model (with disabled interactions) without imposing the hierarchical structure within body systems.

Specifically, all the demographic and condition-severity variables used in the CD-RISC model with disabled interactions were used as explanatory variables in this model. Individuals who had multiple condition-severity variables within a single body system now had all these condition-severity indicators set to one.[7] The R^2 from this model is 0.1178, slightly higher than the R^2 values obtained from the hierarchical models reported in Table 2.

Influence of RAP Diagnoses

The Medicaid CD-RISC model excluded RAP diagnoses due to concerns that these diagnoses may often represent rule-out diagnoses. Since rule-out diagnoses do not measure true health status, their inclusion in the model may bias the model estimates toward zero. In developing the Medicare CD-RISC model reported in earlier in this article, we included RAP diagnoses, in order to be able to compare the CD-RISC model with the DCG-HCC model. In this specification check, we re-estimate the Medicare CD-RISC model after excluding all RAP diagnoses from the model construction to determine the importance of RAP diagnoses for predictive power. The R^2 from this model is 0.1103, slightly lower than the R^2 values of 0.1114 obtained from the CD-RISC model that included RAP

---

[7] Although we did not impose the hierarchical assumptions with body system, we did implement the condition-severity pruning rules. So, statistically insignificant and negative condition-severity variables were either dropped or combined with other severity levels.

Table 4

Predictive Ratios for the Sample, by Number of Body Systems: 1996-1997

Body System	Frequency	Mean in Data	Mean Prediction	Predictive Ratio
0	291,692	1,809	1,816	1.00
1	265,228	2,932	2,780	0.95
2	247,204	4,107	4,135	1.01
3	196,995	5,421	5,644	1.04
4	141,820	7,023	7,322	1.04
5	96,976	8,954	9,187	1.03
6	63,501	11,268	11,268	1.00
7	39,935	13,694	13,576	0.99
8	24,506	17,083	16,077	0.94
9	14,281	20,415	18,785	0.92
10	7,512	23,543	21,669	0.92
11	3,417	28,261	24,736	0.88
12	1,292	30,930	27,718	0.90
13 or More	342	38,694	31,610	0.82

SOURCE: Centers for Medicare & Medicaid Services: Data from the 5 Percent Sample Standard Analytical Files.

diagnoses. The correlation between predictions from the CD-RISC model that included RAP diagnoses and the model that excluded these diagnoses was 0.99.

Model Fit in Validation Subsamples

We examined the model fit for selected validation sample groups by calculating predictive ratios that are equal to the mean predicted payment for a group divided by the mean payment in the data. Groups for which the model underpredicts (predictive ratios that are less than 1) may be subject to risk selection and may be avoided by providers.

Table 4 shows the predictive ratios for the sample by the number of body systems. The model predicts well for individuals with two or fewer body systems. For individuals with three to five body systems, predictive ratios are greater than "1"—this implies that the model overpredicts for these groups. For individuals with eight or more body systems, the model underpredicts. Adding the square of the number of body systems as a predictive variable increases the strength of predictions, but we judged it gave too much incentive to coding multiple body systems, so did not

include it as part of the final model. Still, the predictive ratios for the majority of the sample are quite close to "1."

Table 5 shows the predictive ratios for subgroups in each quintile of selected 1996 payment categories. The ratios are calculated for the subgroup with no expenditure in the payment category, and also for subgroups in each quintile of positive expenditures in the payment category. Predictive ratios for 1996 total expenditure show some overprediction for the low-cost subgroups and underprediction for the subgroup in the 5th quintile.

However, predictive ratios for 1996 home health expenditures show large underprediction for the 5th quintile, suggesting that the CD-RISC model may benefit from the inclusion of more detailed home health information. Similarly, predictive ratios for 1996 durable medical equipment expenditure show considerable underprediction for the highest quintiles of expenditures; however, the underprediction is less severe than that for the home health payment category. These predictive ratios are very similar to those obtained by the DCG-HCC model suggesting that the predictive ratios for both the CD-RISC and DCG-HCC models could be improved by

Table 5

Predictive Ratio for Subgroups in Each Quantile of Selected Payment Categories: 1996-1997

1996 Group	Frequency	1997 Mean in Data	1997 Mean Prediction	Predictive Ratio
Total				
No Expenditure	142,888	1,672	1,879	1.12
1st Quintile of Positive Expenditure	246,455	2,232	2,629	1.18
2nd Quintile of Positive Expenditure	246,826	3,056	3,657	1.20
3rd Quintile of Positive Expenditure	248,084	4,206	4,823	1.15
4th Quintile of Positive Expenditure	249,569	5,984	6,220	1.04
5th Quintile of Positive Expenditure	260,879	13,196	11,225	0.85
Home Health				
No Expenditure	1,255,298	4,259	4,671	1.10
1st Quintile of Positive Expenditure	27,443	10,016	9,908	0.99
2nd Quintile of Positive Expenditure	27,436	10,473	10,465	1.00
3rd Quintile of Positive Expenditure	27,730	12,657	11,393	0.90
4th Quintile of Positive Expenditure	28,185	16,127	12,434	0.77
5th Quintile of Positive Expenditure	28,609	27,534	12,980	0.47
Durable Medical Equipment				
No Expenditure	1,172,747	4,212	4,580	1.09
1st Quintile of Positive Expenditure	43,271	7,559	7,353	0.97
2nd Quintile of Positive Expenditure	43,833	9,385	8,610	0.92
3rd Quintile of Positive Expenditure	43,932	9,659	8,626	0.89
4th Quintile of Positive Expenditure	44,545	11,850	9,871	0.83
5th Quintile of Positive Expenditure	46,373	18,342	12,234	0.67

SOURCE: Centers for Medicare & Medicaid Services: Data from the 5 Percent Sample Standard Analytical Files.

using diagnostic information on home health and DME (Pope et al., 2000b). The advantage in prediction from using these data must be weighed against the extra burden of collecting these data and their clinical reliability.

DISCUSSION

CMS currently uses a risk-adjustment method for Medicare capitation payments that is based on the primary inpatient diagnosis of Medicare beneficiaries from prior years. Relying solely on inpatient diagnoses to set costs creates perverse incentives to hospitalize, and in addition, ignores chronic outpatient conditions. Therefore, in 2004, CMS will implement a new risk-adjustment method that improves on the current system by incorporating inpatient and outpatient diagnoses from prior year(s).

A number of risk-adjustment models using diagnostic information were developed as candidates for Medicare risk adjustment. The Medicare CD-RISC model presented in this article is one of several recent efforts to develop such risk-adjustment systems (Pope et al., 2000b; Kronick et al., 2000; 2002).

The criteria for judging such systems are: strength of predictions of utilization; ease of collection; ease of audit and difficulty of gaming; and size of incentives for inefficient care (Newhouse, 1986). Another criterion is acceptability to all those with a stake in the outcome of the risk adjustment, including payers, plans, providers, and patients (Carter et al., 1997, 2000).

The strength of prediction, or the ability to predict variations in the expected cost of care due to observable patient characteristics is widely seen as important to deter selection of low-cost patients by plans (Newhouse et al., 1989). In addition, better predictions should provide plans with incentives to provide good care, and especially good care for ill patients who tend to be more expensive (Ellis and McGuire, 1986, 1993). Risk-adjusted payment can

increase efficiency by deterring socially wasted effort to select patients and by replacing competition for healthy patients with price competition (Newhouse, Buntin, and Chapman, 1997). Adequate risk adjustment would allow Medicare to pay for the extent to which illness differs across those in different plans, but not pay extra for more expensive practice styles or higher prices (Robinson et al., 1991).

All the systems based on prior year diagnoses have similar ease of collection and ease of audit. The CD-RISC concept that the severity of a particular condition depends on the patient's other conditions is consistent with much clinical thinking and found in clinical payment systems, such as DRGs that have found acceptability. We think this concept may increase the clinical acceptability of CD-RISC. However, the other systems have become clinically more coherent which may make them also acceptable to clinicians. All throw out certain diagnoses that were good predictors historically when no payments depended on the predictions, but might be susceptible to coding games for higher payment.

The CD-RISC model differs from most other models in the literature by imposing a hierarchical structure on condition variables so that only one condition within a body system contributes to individual payments. The hierarchical structure within body system reduces incentives to game the system by entering multiple related codes for the same condition. Imposing this structure reduces the predictive power of the selected condition variables from 11.8 to 11.1 percent, but that lower value is still comparable to other models in the literature.

Among the existing risk-adjustment models, the CDPS's hierarchical structure bears the most similarity to the CD-RISC (Kronick et al., 2002). However, there are several key differences between the CDPS and the CD-RISC model. First, the CD-RISC model contains 199 condition categories compared with 66 diagnostic categories in the CDPS. Second, the assignment of severity levels in the CD-RISC model incorporates clinical information from other body systems. The CDPS system's condition assignment does not incorporate such information—instead this system incorporates selected condition interactions as separate model variables. Third, the hierarchical structure in the CD-RISC system is determined by empirically iterating the model until the condition ranks within a body system converge. The CDPS does not have this feature. Despite these differences, both models have a virtually identical R^2 of 11 percent.

While the CD-RISC model does not include interactions directly in the model, CD-RISC incorporates the concept of interactions between conditions in the assignment of different severity levels. The Medicare CDPS model includes several interactions of diagnostic groups and finds that these coefficients are relatively small, and the inclusion of interactions has no effect on the R^2 (Kronick et al., 2002). Given that the Medicare CDPS was developed using the same data set as the CD-RISC, including interactions explicitly in the CD-RISC model is unlikely to change the results in this article.

Episode payments were an integral part of the Medicaid CD-RISC model. The model included episode payments for a newborn and tested numerous other episode-based variables. Like DRG payments, the amount of payment for a particular episode is set prospectively, but the payments are made only when the episode occurs. Including episode payments in the Medicare CD-RISC model was beyond the scope of this article; however, these payments should be considered before implementing a payment mechanism.

These payments are especially relevant when one desires incentives to avoid the underprovision of valuable care or if one believes that selection against the relevant small vulnerable population is likely. Episode payments should also be considered in dealing with payments to small organizations such as physician groups. In the context of the Medicare CD-RISC model, episode payments are likely to be the most appropriate payment method for conditions, such as new cancer treatment, which come in standard packages that could be paid by fixed preset amounts based on benchmarking. Diagnosis-based ICD-9-CM codes are limited in their ability to discriminate between new onset of cancer and cancer in remission, so procedure based codes may be more appropriate in determining the correct payment for new cancer treatment. At present the model includes condition codes for chemotherapy and other procedure oriented codes that pick up such treatment. These codes could potentially be replaced by episode payments in future work.

In light of CMS' recent move toward risk-adjustment models that are more parsimonious in their data requirements, future work with the CD-RISC model should focus on evaluating predictive power while reducing the number of conditions or sources of diagnostic information used to develop model estimates.

ACKNOWLEDGMENTS

REFERENCES

Ash, A. S., Ellis, R. P., Pope, G. C., et al.: Using Diagnoses to Describe Populations and Predict Costs. *Health Care Financing Review* 21(3):7-28, Spring 2000.

Carter, G. M., Bell, R. B., Dubois, R. W., et al.: *A Clinically Detailed Risk Information System for Cost. DRU-1731-1-HCFA*. RAND. Santa Monica, CA. 1997.

Carter, G. M., Bell, R. B., Dubois, R. W., et al.: A Clinically Detailed Risk Information System for Cost. *Health Care Financing Review* 21(3):65-85, Spring 2000.

Centers for Disease Control and Prevention: *International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM)*. Internet address: http://www.cdc/icd9/abticd9.htm. (Accessed 2003.)

Centers for Medicare & Medicaid Services: Risk-Adjustment Data Fact Sheet Issued March 29, 2002. Internet address: http://www.cms.hhs.gov/health-plans(Accessed 2003.)

Daley, J., Jencks, S., Draper, D., et al.: Predicting Hospital-Associated Mortality for Medicare Patients: A Method for Patients with Stroke, Pneumonia, Acute Myocardial Infarction, and Congestive Heart Failure. J*ournal of the American Medical Association* 260(24):3617-3624, December 23-30, 1988.

Efron, B.: Regression and ANOVA with Zero-One Data: Measures of Residual Variation. *Journal of the American Statistical Association* 73(1):113-121, 1978.

Ellis, R. P., and McGuire, T. G.: Provider Behavior Under Prospective Payment. *Journal of Health Economics* 5:129-151, June 1986.

Ellis, R. P. and McGuire, T. G.: Supply-Side and Demand-Side Cost Sharing in Health Care. *Journal of Economic Perspectives* 7(4):135-151, 1993.

Ellis, R. P., Pope, G. C., Iezzoni, L. I., et al.: Diagnosis-Based Risk-Adjustment for Medicare Capitation Payments. *Health Care Financing Review* 17(3):101-128, Spring 1996.

Hadorn, D. C., Keeler, E. B., Rogers, W. H. and Brook, R. H.: *Assessing the Performance of Mortality Prediction Models*. MR-181-HCFA. RAND. Santa Monica, CA. 1993.

Keeler, E.B., Carter, G.M., and Newhouse, J.P.: A Model of the Impact of Reimbursement Schemes on Health Plan Choice. *Journal of Health Economics* 17:297-320, June 1998.

Kronick, R. T., Gilmer, T., Dreyfus, T., and Ganiats, T.: *CDPS-Medicare: The Chronic Illness and Disability Payment System Modified to Predict Expenditures for Medicare Beneficiaries*. Final Report to the Centers for Medicare & Medicaid Services, Pub. No. 500-00-0008. Baltimore, MD. June 24, 2002.

Kronick, R.T., Gilmer, T., Dreyfus, T., and Lee, L.: Improving Health-Based Payment for Medicaid Beneficiaries: CDPS. *Health Care Financing Review* 21(3):29-63, Spring 2000.

Kronick, R.T., Dreyfus, T., Lee, L. and Zhou, Z.: Diagnostic Risk Adjustment for Medicaid: The Disability Payment System. *Health Care Financing Review* 17(3):7-34, Spring 1996.

Newhouse, J. P.: Rate-Adjusters for Medicare Capitation. *Health Care Financing Review, Statistical Supplement, 1986*. December 1986.

Newhouse, J. P., Manning, W. G., Keeler, E. B., and Sloss, E. M.: Adjusting Capitation Rates Using Objective Health Measures and Prior Utilization. *Health Care Financing Review* 10(3):41-54, Spring 1989.

Newhouse, J. P., Buntin, M. B., and Chapman, J. D.: Risk-Adjustment and Medicare: Taking a Closer Look. *Health Affairs* 16(5):26-43, September/October 1997.

Pope, G. C., Ellis, R. P., Ash, A. S., et al.: Principal Inpatient Diagnostic Cost Group Model for Medicare Risk-Adjustment. *Health Care Financing Review* 21(3):93-118, Spring 2000a.

Pope, G. C., Ellis, R. P., Ash, A. S., et al.: Diagnostic Cost Group Hierarchical Condition Category Models for Medicare Risk Adjustment. Draft Final Report Prepared for Health Care Financing Admistration. Baltimore, Maryland. July 2000b.

Pope, G. C., Adamache, K. W., Walsh, E. G., and Khandker, R. K.: Evaluating Alternative Adjusters for Medicare. *Health Care Financing Review* 20(2):109-129, Winter 1998.

Robinson, J. C., Luft, H. S., Gardner, L. B., and Morrison, E. M.: A Method for Risk-Adjusting Employer Contributions to Competing Health Insurance Plans. *Inquiry* 28:107-116, Summer 1991.

Schneeweiss, R., Rosenblatt, R. A., and Cherkin, D. C.: Diagnostic Clusters: A New Tool for Analyzing the Content of Ambulatory Medicare Care. *Medical Care* 21:105-122, January 1983.

Weiner, J. P., Dobson, A., Maxwell, S. L., et al.: Risk-Adjusted Medicare Capitation Rates Using Ambulatory and Inpatient Diagnoses. *Health Care Financing Review* 17(3):77-100, Spring 1996.