
Diagnosis-Based Risk Adjustment for Medicare Capitation Payments

Randall P. Ellis, Ph.D., Gregory C. Pope, M.S., Lisa I. Iezzoni, M.D., M.Sc., John Z. Ayanian, M.D., M.P.P., David W. Bates, M.D., M.Sc., Helen Burstin, M.D., M.P.H., and Arlene S. Ash, Ph.D.

Using 1991-92 data for a 5-percent Medicare sample, we develop, estimate, and evaluate risk-adjustment models that utilize diagnostic information from both inpatient and ambulatory claims to adjust payments for aged and disabled Medicare enrollees. Hierarchical coexisting conditions (HCC) models achieve greater explanatory power than diagnostic cost group (DCG) models by taking account of multiple coexisting medical conditions. Prospective models predict average costs of individuals with chronic conditions nearly as well as concurrent models. All models predict medical costs far more accurately than the current health maintenance organization (HMO) payment formula.

INTRODUCTION

Since the early 1970s, Medicare has encouraged beneficiaries to enroll in HMOs, believing that they are a cost-saving alternative to the fee-for-service (FFS) sector. Initially reimbursed on an FFS basis, since the mid-1980s HMOs have been able to enter into at-risk contracts with the Health Care Financing Administration (HCFA). Premium payments to these at-risk HMOs are based on 95 percent of the adjusted average per capita cost (AAPCC) of Medicare beneficiaries participating in the traditional FFS Medicare program. The AAPCC, calculated annually by the Office

of the Actuary at HCFA, considers HMO enrollees' age, sex, welfare status, and whether or not they were in a nursing home. In addition, the relative cost weight calculated using these four demographic factors is adjusted using a geographic factor based on average costs of FFS beneficiaries in the counties served by the HMO.

Since its implementation, the AAPCC has prompted concerns about its fairness and accuracy (Eggers and Prihoda, 1982; Lubitz and Prihoda, 1984; Beebe, Lubitz, and Eggers, 1985). Numerous studies have demonstrated that the AAPCC explains only about one percent of total variability in annual costs across Medicare beneficiaries (Ash et al., 1989; Newhouse, 1986). Mathematica Policy Research, Inc. (Hill and Brown, 1990) found that all 98 HMOs studied experienced favorable selection: The costs of HMO enrollees were less than costs of non-HMO enrollees in the year prior to HMO enrollment. It has also been demonstrated (Brown et al., 1988; 1986; 1993; Brown and Langwell, 1987) that Medicare HMOs on average had lower mortality, and that HMO disenrollees had systematically higher costs than Medicare beneficiaries remaining in the FFS sector. Such findings have spurred interest in improving the AAPCC. The U.S. General Accounting Office (1994) concluded that major changes are needed in the program's methods, including implementation of a health status risk adjuster.

A variety of alternatives to the AAPCC have been proposed (Epstein and Cumella, 1988; Ash et al., 1989; Anderson et al., 1989). These alternatives differ in

Randall P. Ellis is with Boston University. Gregory C. Pope is with Health Economics Research, Inc. Lisa I. Iezzoni, John Z. Ayanian, David W. Bates, and Helen Burstin are with the Harvard Medical School. Arlene S. Ash is with the Boston University School of Medicine. This research was supported by Contract Number 500-92-0020 from the Health Care Financing Administration (HCFA). The views and opinions expressed in this article are the authors'. No endorsement by HCFA, Boston University, Health Economics Research, Inc., or Harvard Medical School is intended or should be inferred.

the type of information used to predict future costs. Epstein and Cumella classify potential adjusters for revising the AAPCC into six categories: perceived health status, functional health status, prior utilization, clinical descriptors, sociodemographic characteristics, and additional predictors.

Perceived health status and functional health status measures require expensive, ongoing surveys. In addition, this information has generally had only moderate predictive power, can be subjective, and requires substantial new data collection before introduction (e.g., Thomas and Lichtenstein, 1986; Whitmore et al., 1989; Schaffler, Howland, and Cobb, 1992). New sociodemographic characteristics and additional predictors are either unattractive conceptually (e.g., mortality rates) or have only limited predictive power (e.g., whether or not the beneficiary has a driver's license or lives alone).

Prior utilization measures include expenditures, number of outpatient visits, number of hospitalizations, or nursing home use. These have been shown to provide the highest predictive power of any risk adjusters (van Vliet and van de Ven, 1990; Thomas and Lichtenstein, 1986; Beebe, Lubitz, and Eggers, 1985; Anderson et al., 1986; 1989). However, prior utilization measures suffer from four weaknesses: (1) The necessary information (e.g., nursing home use) is generally not available; (2) the necessary data cannot be routinely measured within an HMO setting (e.g., levels of expenditure); (3) payments based on these measures (e.g., the number of admissions or visits) may create perverse incentives, inappropriately encouraging HMOs to hospitalize or provide outpatient treatment; (4) payments based on such measures may be unfair to HMOs that provide good care with less intensive utilization.

Risk-adjustment models based on diagnostic information appear best able to overcome the previously noted weaknesses. The information needed is available for large populations; diagnoses can potentially be measured (and in many cases already are) in HMOs; incentives and inequities can be mitigated.

One diagnosis-based model, DCGs, forms an important precursor to the work described here (Ash et al., 1986; 1989). The DCG approach uses diagnostic information from hospitalizations occurring during a base year to classify beneficiaries into one of eight (later increased to nine) DCGs. These eight DCGs, together with demographic characteristics, are then used to predict health costs in a subsequent year the prediction year. Since the original work by Ash et al. (1986), the DCG model has been further enhanced as described in Ellis and Ash (1988; 1989; 1995); Ash, Ellis, and Iezzoni (1990); and Ellis (1990). Ellis and Ash (1995) present findings from the DCG models that are the point of departure for this article.

This article develops, estimates, and evaluates risk-adjustment models which differ in the information used to predict 1992 Medicare payments. Comparisons are made with two models: an AAPCC-like model that classifies people using only demographic information, and a DCG model that also uses the principal inpatient diagnoses from the preceding year. Four extensions to these models are examined.

The first extension adds secondary diagnoses from hospital inpatient bills, diagnoses from hospital outpatient claims, and diagnoses from bills for ambulatory or inpatient physician services to principal hospital inpatient diagnoses. Using a DCG framework, individuals are classified based on the single highest cost diagnosis recorded for a person during the year.

The second extension expands the risk-adjustment framework to account for multiple medical conditions that persons may experience. We call this new framework the hierarchical coexisting conditions (HCC) model. The HCC model organizes closely related conditions into hierarchies. For conditions within a disease hierarchy, a person is characterized only by the most serious condition. Across such hierarchies, persons may be classified as having multiple conditions.

The third extension uses life-sustaining medical procedures to classify individuals. Relatively non-discretionary procedures used primarily to sustain the life of severely ill patients and associated with high future medical costs are utilized to predict costs in HCC model variants.

The fourth extension is that all models are estimated and evaluated both prospectively—using diagnoses (and other information) to predict subsequent year payments—and concurrently—using diagnoses to predict payments in the same year.¹ Both the DCG and HCC diagnostic classifications are redefined to reflect differences in the expenditures associated with a diagnosis in the year it is made versus the following year.

DATA

Our analysis uses a 5-percent sample of aged and disabled beneficiaries eligible for Medicare in 1991 or 1992, obtained from HCFA data files. The sample includes only people with a full 12 months of eligibility for both Part A and Part B coverage in 1991. We eliminated anyone dying during 1991, becoming eligible during 1991 or 1992, HMO enrollees, or beneficiaries in HCFA's End Stage Renal Disease Program. Appropriate statistical adjustments are made to account for partial year expenditures of those who died during 1992 (Ellis and Ash, 1995).

¹Concurrent models are sometimes called "retrospective" models.

We use a split sample design to avoid overfitting the data and biasing measures of goodness of fit. We randomly divided our 5-percent sample into 2.5-percent model development ($N=680,188$) and model validation ($N=680,438$) halves. These large sample sizes are crucial for accurately estimating the cost of expensive, but rare, medical conditions.

Dependent Variable

Our dependent variable was total 1992 Medicare program expenditures for each beneficiary, excluding beneficiary deductibles and copayments. Medicare-covered expenditures for hospital inpatient, hospital outpatient, physician, home health, hospice, skilled nursing facility, laboratory, durable medical equipment, and other services were all included. For inpatient services subject to Medicare's prospective payment system, diagnosis-related-group payments were aggregated with direct teaching, outlier, and organ transplant payments. To be consistent with future Medicare payment methods, 1992 physician payments from a fully-phased-in Medicare fee schedule (resource based relative value scale [RBRVS]) were simulated.² Actual reimbursement was used to capture other payments. Non-Medicare-covered services, including most nursing home care and outpatient drugs, are not included in our analysis. Deductibles, copayments, and non-covered services account for roughly one-half of the total health expenditures of the elderly.

Independent Variables

The independent variables used are of three types: demographic, diagnostic, and procedural. Demographic information is

²Assignment of fully phased-in RBRVS fees was conducted by researchers at Johns Hopkins University under the direction of Jonathan Weiner, Dr. P.H.

included as 12 age-sex cells, based on data obtained from Medicare enrollment files for January 1, 1992. Medicare beneficiaries eligible for coverage because of disability represent less than 9 percent of our sample. This sample size was too small for developing separate risk-adjustment models. We pooled disabled beneficiaries with aged beneficiaries rather than excluding them. Differential costs for the aged and disabled populations are implicitly incorporated by age because disabled beneficiaries are under age 65. Diagnoses were obtained from hospital inpatient, hospital outpatient, and physician claims, including both header and line item diagnoses. Diagnoses are coded according to the *International Classification of Diseases, Ninth Revision, Clinical Modification* (ICD-9-CM) (Public Health Service and Health Care Financing Administration, 1980). Diagnoses from Medicare Part B bills submitted by non-physicians, such as laboratories and medical equipment suppliers, were excluded. Selected procedures coded using the *Current Procedural Terminology, 4th Edition* (CPT-4) classification system were also extracted from Part B claims.

Grouping Diagnostic Codes

Given that there are more than 14,000 valid ICD-9-CM diagnostic codes, an important first step is to group ICD-9-CM codes into aggregates, building on the approach described in Ash et al. (1989) and Ellis and Ash (1995). Starting with the 104 groups of diagnoses used in this previous DCG work, the four physician authors of this article, in consultation with outside specialists, combined ICD-9-CM codes into diagnostic groups which we refer to as DXGROUPs. Two sets of DXGROUPs were formed: principal inpatient DXGROUPs, and all (hospital and physician) diagnoses, both inpatient and ambulatory, DXGROUPs. The 143 princi-

pal inpatient DXGROUPs are assigned from each beneficiary's principal inpatient diagnoses only, whereas the 432 all-diagnoses DXGROUPs are assigned from all hospital and physician diagnoses. Each reimbursable ICD-9-CM code is assigned to one and only one principal inpatient DXGROUP, and one and only one all-diagnosis DXGROUP.

The physicians formed DXGROUPs according to the following criteria:

- Groups should separate diagnoses by anticipated costliness.
- Groups should have a sample size of at least 500.
- Groups should be clinically homogeneous and meaningful.
- Alternative codes that can be used for the same medical condition should be grouped together.
- Each reimbursable ICD-9-CM code should belong to one and only one group.
- Each all-diagnoses group should be wholly contained within a single inpatient group.

The sample size goal of 500 corresponds to a relative standard error of mean expenditures of about 10 percent, which was seen as acceptably accurate. Where the second and third criteria conflicted—sample size versus clinical cogency—priority was given to clinical cogency. Thus, a separate DXGROUP for HIV/AIDS was formed even though it contains fewer than 500 beneficiaries. Special emphasis was placed on distinguishing (i.e., separately grouping) the very highest cost diagnoses; less emphasis was accorded to making fine clinical distinctions among lower cost diagnoses. The DXGROUPs necessarily reflect the frequency of medical conditions among Medicare's elderly and disabled population. Thus, for example, few distinctions were made among pregnancy, childbirth, and child-

hood disorders. In forming DXGROUPs, 1,835 ICD-9-CM diagnostic codes were employed: 1,021 (56 percent) are three-digit codes, 642 (35 percent) are four-digit splits, and 172 (9 percent) are five-digit splits. When a three-digit code is placed in a DXGROUP, all four- and five-digit codes that begin with the same three digits are assigned to the same DXGROUP; and similarly for the four-digit code assignments.

Groupings were reviewed by seven other physicians spanning a range of specialties from the Boston area and Indiana University, as well as by two medical coding experts, and were compared with other clinical groupings of diagnostic codes (e.g., Elixhauser, Andrews, and Fox, 1993). In most cases, DXGROUPs correspond to specific medical conditions. Examples of all-diagnoses DXGROUPs are lung cancer, diabetes without complications, Parkinson's disease, viral hepatitis, aortic aneurysm, and asthma.

DEMOGRAPHIC AND DCG MODELS

Table 1 summarizes our three classes of risk-adjustment models: demographic models, DCG models, and HCC models.

Demographic Model

We consider one model which uses only demographic information: an AAPCC-like model that predicts Medicare payments using twelve age-sex categorical variables and a dummy variable for Medicaid enrollment. Because disability eligibility coincides with being under age 65, we do not include a separate indicator for disability status.

Institutional status is used in the current AAPCCs but was unavailable to us. None of our models includes geographic factors similar to the county-level adjustments of the AAPCC.

DCG Models

The DCG modeling framework is described in detail in Ash et al. (1989) and Ellis and Ash (1995). Creation of a DCG model takes place in four stages. The first stage is to map ICD-9-CM diagnostic codes into DXGROUPs, as described. The second stage is to run the DXGROUPs through a sorting algorithm that places each DXGROUP in a relatively homogeneous cost group, or DCG. The sorting algorithm ranks the DXGROUPs by mean expenditures. The highest cost DXGROUPs are grouped into the highest numbered DCG. The DXGROUPs are then re-ranked by mean expenditures, excluding people with the costliest conditions, who have already been classified into the highest numbered DCG. Each successively lower-numbered DCG includes DXGROUPs not already classified into higher-numbered DCGs, with the lowest numbered DCG including the lowest cost set of conditions. A third stage is to use clinical judgment to reclassify poorly grouped DXGROUPs into DCGs. Poorly grouped DXGROUPs are modified for three reasons: to use judgment where sample sizes are small, to improve clinical plausibility, and to improve incentives. The final stage of development is to calibrate payment parameters through estimation of a multiple regression equation. For this regression, each person is uniquely assigned to the most expensive DCG to which any of their diagnoses belong.

For this article, we present two DCG model variants that differ in the type of information used to predict payments.³ The principal inpatient DCG (PIPDCG) model classifies people based on their single highest cost principal inpatient diagnosis. The PIPDCG model has the simplest data

³Further DCG and HCC model variations are presented in the final report (Ellis et al., 1996).

Table 1
Risk-Adjustment Models

| Model | Description | Maximum Number of Diagnostic Categories for Individuals* |
|--|--|--|
| Demographic Model | | |
| Adjusted Average Per Capita Cost (AAPCC) | Includes age, sex, and Medicaid status, as used in Medicare's current method of paying HMOs. | 0 |
| DCG Models | | |
| Principal Inpatient Diagnostic Cost Group Model (PIPDCG) | Pays for the single highest-cost principal inpatient diagnosis in addition to AAPCC factors. | 1 |
| All-Diagnoses Diagnostic Cost Group Model (ADDCG) | Pays for the single highest cost hospital or physician diagnosis in addition to AAPCC factors. | 1 |
| HCC Models | | |
| Hierarchical Coexisting Conditions Model (HCC) | 34 (prospective) or 44 (concurrent) Hierarchical Coexisting Conditions, plus age and sex. | 23** |
| Hierarchical Coexisting Conditions and Procedures Model (HCCP) | 40 (prospective) or 44 (concurrent) HCCs, 11 Procedure-Based HCCs, plus age and sex. | 33 |
| Hierarchical Coexisting Conditions, Procedures, and Hospitalizations Model (HCCPH) | 40 (prospective) or 39 (concurrent) HCCs, 11 Procedure-Based HCCs, and 3 (prospective) or 5 (concurrent) Principal Inpatient HCCs, plus age and sex. | 36 |

*Prospective or concurrent version of the model.

** For concurrent HCC model, maximum number is 25.

SOURCE: Ellis, R.P., Pope, G.C., Iezzoni, L.I., et al., 1996.

requirement: It can be implemented with knowledge of each person's principal inpatient diagnoses only. This is an important advantage in circumstances where ambulatory or secondary inpatient diagnoses are unevenly available, or inaccurate. On the other hand, as a payment system, the PIPDCG model establishes incentives to hospitalize enrollees because only inpatient diagnoses are used to classify individuals. Also, because the order of inpatient diagnoses is often somewhat arbitrary, the model is open to gaming because providers could reorder diagnoses to maximize reimbursement. Finally, no diagnostic information is available to classify the nearly 80 percent of Medicare beneficiaries who are not hospitalized in a given year.

The second DCG model presented is the all-diagnoses DCG (ADDCG) model. The

ADDCG model adds secondary inpatient, hospital outpatient, and physician diagnoses (for either inpatients or outpatients) to the principal inpatient diagnosis, and classifies people based on their single highest predicted cost diagnosis, with no distinction made as to the source of the diagnosis. It classifies all but the approximately 12 percent of Medicare beneficiaries who have neither hospital nor physician medical claims in a year. Because it makes no distinction by source of diagnosis, it avoids incentives to hospitalize, and it does not reward coding proliferation because it pays only for the single highest cost diagnosis.

Both prospective and concurrent versions of the principal inpatient and all diagnosis DCGs were developed, for a total of four DCG models. Prospective and concur-

rent DCGs are defined analogously using the methods previously described. Concurrent DCG models differ from prospective DCG models in that people are assigned to DXGROUPs and thus to DCGs, based on current year rather than previous year diagnoses. In our case, concurrent models use 1992 diagnoses and prospective models use 1991 diagnoses, both to predict 1992 payments. In addition, the mapping of DXGROUPs into DCGs is redefined based on the sorting algorithm results for the concurrent DXGROUPs rather than the prospective groups. Acute conditions having particularly high costs in the current year (e.g., heart attack) are classified into the highest numbered concurrent DCGs, whereas chronic conditions with stable or rising costs over time (e.g., cancer) are classified into the highest numbered prospective DCGs.

HCC MODELS

Rationale for HCC Models

DCG models predict a person's costliness by identifying his or her single highest cost diagnosis. For example, if a person has lung cancer, diabetes without complications, and coronary artery disease, the DCG model would consider lung cancer only, because lung cancer is the diagnosis which predicts the highest future Medicare expenditures. Focusing on the single highest-cost diagnosis has several virtues: simplicity, less sensitivity to incomplete diagnostic coding, and not rewarding proliferation of diagnostic coding by health plans. However, a single diagnosis can describe a person's health status only partially. This is especially true among elderly and disabled Medicare beneficiaries, many of whom have multiple chronic health problems.

In contrast with DCG models, HCC models characterize health status by consider-

ing multiple coexisting medical conditions. Rather than focusing on the highest cost condition, HCC models sum the incremental predicted cost (payment) for each condition to arrive at the total predicted cost (payment). HCC models will predict a different level of expenditures for a person with lung cancer versus a person with lung cancer, diabetes, and coronary artery disease. Like the DRGs used by Medicare for hospital payment, DCGs are mutually exclusive and exhaustive: A person belongs to one and only one DCG. In contrast, a person may be characterized by no HCCs, one HCC, or multiple HCCs.

Defining Coexisting Conditions

The narrowly defined DXGROUP categories are inappropriate for additive, multiple condition models. The large number of categories creates incentives for coding proliferation, i.e., coding as many ICD-9-CM codes as possible to maximize reimbursement. It also raises the danger of some conditions being classified into two or more different categories, thus being paid for more than once. In addition, regression parameter estimates are often implausible (e.g., negative) or imprecise because of small sample sizes.

To overcome these limitations, the physician panel created more aggregated groupings of medical conditions. We call each such group of DXGROUPs a coexisting condition: coexisting because an individual may simultaneously have more than one, and condition because the group of diagnoses does not necessarily reflect comorbidity (which in clinical terms means a related condition). Coexisting condition groups combine DXGROUPs belonging to a major body system or disease type by costliness and clinical relation. The major body systems and disease types generally follow those established by

the ICD-9-CM coding system (e.g., infectious diseases, neoplasm, mental disorders). Grouping of DXGROUPs by costliness was informed by the DCG assignment of DXGROUPs, coefficients from a regression of Medicare expenditures on the 432 DXGROUPs, and mean Medicare expenditures by DXGROUP. Existing lists of comorbidities were also consulted for guidance (Charlson et al., 1987; Deyo et al., 1992; Keeler et al., 1990).

The coexisting condition groups used in two prospective HCC models are shown in Table 2. Coexisting condition groups were also defined for the concurrent HCC model (not shown in Table 2), using the same criteria and methods as for the prospective groups, except that all analysis was done on a concurrent basis. The most significant difference between the prospective and concurrent HCCs is that new groups were created for particularly high cost acute conditions, for example, heart attack, cerebral hemorrhage, and acute renal failure. Before making any exclusions, there are 81 concurrent HCCs compared with 66 prospective HCCs.

Creating Hierarchies

Hierarchies were created among subsets of the coexisting conditions based on clinical judgment. Hierarchies are defined among related medical conditions where some can be assigned precedence over others because they are a more costly or clinically significant disease process. Coexisting conditions redefined according to these hierarchical rules are called HCCs. The hierarchies specify that a person with multiple, clinically related coexisting conditions is assigned only to the highest ranked among these related coexisting conditions. For example, the diabetes hierarchy specifies that if a person is coded into HCC 12, Higher Cost Diabetes, he or

she may not also be assigned to HCC 13, Lower Cost Diabetes. Similarly, a person in HCC 4, Metastatic Cancer, is not allowed to be in HCC 5, High Cost Cancers, or in any of the other six HCCs in the neoplasm hierarchy. The hierarchical relationships among HCCs used in two prospective HCC models are indicated in Table 2.

The hierarchies serve three main functions. First, they improve clinical validity. For example, if a person has a more severe manifestation of diabetes, characterizing that person also with a less serious type of diabetes is not clinically useful. Second, the hierarchies limit incentives for coding proliferation. Without the hierarchies, a provider could be paid more for coding both more and less severe diabetes, and both metastatic cancer and anatomically specific cancer. With the hierarchies, payment is made only for the most costly HCC in a disease hierarchy. Third, the hierarchies improve the precision of the estimated payment weights (regression coefficients). With coding of HCCs limited to the highest-ranked HCC in a hierarchy, the expenditures associated with a disease type (e.g., neoplasm or diabetes) are loaded onto the highest-ranked conditions, rather than being diffused among higher and lower cost conditions. A more precise estimate of the relative payment weight for higher-cost diabetes versus lower-cost diabetes is obtained, for instance.

Procedure Groups

In addition to using diagnostic information, we explored the use of selected medical procedures for risk adjustment. In general, basing payments on procedures was considered undesirable because performance of many procedures is discretionary.⁴

⁴This principle distinguishes HCCs and DCGs from episode unit of payment classifications, such as DRGs, which pay more to providers who choose discretionary surgical treatments for many diagnoses.

Table 2
Prospective Hierarchical Coexisting Conditions Models, With Incremental Payment Weights¹

| HCC | Label | Example(s) | Hierarchy (Rank) | Percent of Medicare Beneficiaries ² | Incremental Payment | |
|-----|--|---|-------------------------------|--|---------------------|--------------------------|
| | | | | | HCC—Diagnosis Only | HCCP—Includes Procedures |
| 1 | High-Cost Infectious Diseases | Septicemia, HIV/AIDS | None | 1.1 | \$4,116 | \$3,045 |
| 2* | Moderate-Cost Infectious Diseases | Tuberculosis, meningitis | None | 2.1 | — | 1,411 |
| 4 | Metastatic Cancer | — | Neoplasm (1) | 1.3 | 6,298 | 4,332 |
| 5 | High-Cost Cancers | Lung cancer | Neoplasm (2) | 0.9 | 4,226 | 3,457 |
| 6 | Moderate-Cost Cancers | Kidney cancer, brain cancer | Neoplasm (3) | 2.0 | 2,168 | 1,680 |
| 7 | Lower-Cost Cancers | Prostate cancer, breast cancer | Neoplasm (4) | 4.8 | 910 | 576 |
| 8* | Carcinoma in Situ | — | Neoplasm (5) | 0.4 | — | 456 |
| 12 | High-Cost Diabetes | Hypoglycemic coma | Diabetes (1) | 1.5 | 3,939 | 3,871 |
| 13 | Lower-Cost Diabetes | Diabetes without complications | Diabetes (2) | 12.0 | 1,451 | 1,425 |
| 14 | Protein-Calorie Malnutrition | — | None | 0.6 | 3,961 | 2,451 |
| 17 | Liver Disease | Cirrhosis | None | 0.4 | 4,269 | 3,971 |
| 18 | High-Cost Gastrointestinal Disorders | Intestinal obstruction | Gastrointestinal (1) | 2.5 | 2,146 | 1,827 |
| 19* | Moderate-Cost Gastrointestinal Disorders | Ulcer without perforation | Gastrointestinal (2) | 7.8 | — | 751 |
| 21 | Bone Infections | Osteomyelitis | None | 0.6 | 2,111 | 1,770 |
| 22 | Rheumatoid Arthritis and Connective Tissue Disease | Systemic lupus erythematosus | None | 2.6 | 1,516 | 1,442 |
| 24 | Aplastic and Acquired Hemolytic Anemias | — | Hematological (1) | 0.3 | 5,505 | 4,778 |
| 25 | Blood/Immune Disorders | Hemophilia | Hematological (2) | 1.8 | 1,337 | 994 |
| 28 | Drug and Alcohol Dependence/Psychoses | — | Mental (1) | 0.7 | 2,442 | 2,318 |
| 29 | Higher-Cost Mental Disorders | Schizophrenia | Mental (2) | 4.4 | 1,635 | 1,603 |
| 31 | Quadriplegia/Paraplegia | — | Neurological (1) | 0.2 | 5,609 | 4,996 |
| 32 | Higher-Cost Nervous System Disorders | Parkinson's disease, multiple sclerosis | Neurological (2) | 6.3 | 1,556 | 1,436 |
| 34 | Respiratory Arrest | — | Cardio-respiratory arrest (1) | 0.3 | 9,282 | 6,561 |
| 35 | Cardiac Arrest/Shock | — | Cardio-respiratory arrest (2) | 0.3 | 1,759 | 1,271 |
| 36 | Respiratory Failure | — | Cardio-respiratory arrest (3) | 2.5 | 2,797 | 2,237 |
| 37 | Congestive Heart Failure | — | Heart (1) | 9.9 | 3,063 | 2,873 |
| 38 | Heart Arrhythmia | Ventric Tachycardia | Heart (2) | 3.2 | 1,333 | 1,212 |
| 39 | Valvular Heart Disease | Rheumatic Fever/Heart Disease | Heart (3) | 2.4 | 804 | 757 |
| 40 | Coronary Artery Disease | Myocardial infarction, angina pectoris | Heart (3) | 13.7 | 1,049 | 995 |
| 45 | Cerebrovascular Disease | Cerebrovascular accident | None | 8.4 | 1,253 | 1,174 |
| 46 | Vascular Disease | Atherosclerosis, aneurysm | None | 12.1 | 1,114 | 1,015 |
| 48 | Chronic Obstructive Pulmonary Disease | Emphysema, Asthma | Lung (1) | 11.9 | 1,555 | 1,448 |
| 49 | Higher-Cost Pneumonia | Pneumococcal pneumonia | Lung (1a) | 1.1 | 2,943 | 2,673 |
| 50* | Lower-Cost Pneumonia | Unspecified pneumonia | Lung (2a) | 4.4 | — | 1,104 |
| 51* | Pleurisy/Fibrosis of Lungs | Black lung disease | Lung (3) | 1.5 | — | 801 |
| 54 | Renal Failure | — | None | 1.4 | 3,454 | 2,907 |
| 58 | Chronic Ulcer of Skin | — | None | 2.5 | 2,633 | 2,461 |
| 60 | Hip and Vertebral Fractures | — | None | 2.4 | 1,109 | 998 |

See notes at end of table.

Table 2—Continued
Prospective Hierarchical Coexisting Conditions Models, With Incremental Payment Weights¹

| HCC | Label | Example(s) | Hierarchy (Rank) | Percent of Medicare Beneficiaries ² | Incremental Payment | |
|-----------------------------|--|---|--------------------------------------|--|---------------------|--------------------------|
| | | | | | HCC—Diagnosis Only | HCCP—Includes Procedures |
| 61 | Higher-Cost Injuries and Poisonings | Intracranial injury, third-degree burns | None | 2.2 | \$1,254 | \$1,052 |
| 63* | Complications of Medical and Surgical Care | Misadventure to patient in surgery | None | 3.4 | — | 709 |
| 64 | Coma | — | None | 0.3 | 1,694 | 1,361 |
| Procedure-Based HCCs | | | | | | |
| 67* | Major Organ Transplant | Heart transplant | Transplant (1) | 0.0 | — | 5,142 |
| 68* | Status/History of Major Organ Transplant | — | Transplant (2) | 0.1 | — | 1,156 |
| 70* | Tracheostomy | — | Artificial opening, tracheostomy (1) | 0.1 | — | 24,474 |
| 71* | Gastrostomy | — | Artificial opening (1) | 0.2 | — | 5,022 |
| 72* | Enterostomy | — | Artificial opening (1) | 0.1 | — | 5,119 |
| 73* | Artificial Opening Status/Attention | Attention to gastrostomy | Artificial opening (2) | 0.4 | — | 2,236 |
| 74* | Machine Dependence | Ventilator dependence | Tracheostomy (2) | 1.2 | — | 2,190 |
| 77* | Venous Access Port | — | None | 0.1 | — | 7,139 |
| 78* | Chemotherapy | — | None | 0.8 | — | 4,642 |
| 79* | Dialysis | — | None | 0.0 | — | 16,586 |
| 80* | Major Surgical Amputations | Amputation of leg | None | 0.1 | — | 2,607 |

*Not included in hierarchical coexisting conditions (HCC) payment model.

¹Payment models also include 12 age and sex cells. Age and sex weight must be added to sum of HCC weights to obtain total payment.

²Percent of sample in category after application of hierarchical restrictions.

SOURCE (for payment weights): 1992 (expenditures) and 1991 (diagnoses) Medicare claims.

A few procedures, however, are so invasive and unpleasant that physicians are extremely unlikely to be influenced by financial considerations. These procedures will be used only as a last resort to sustain life in severely ill patients. The four physician authors identified groups of life-sustaining procedures appropriate for inclusion in risk-adjustment models. Specific CPT-4 codes were selected for each type of procedure. The physicians selected procedures according to the following criteria:

- The procedure should indicate a severely ill patient.
- The procedure should be associated with high expected medical costs.
- Little discretion should apply to the decision to use the procedure.

Ten groups of procedure codes and five groups of related ICD-9-CM V-Codes were identified: major organ transplants, dial-

ysis, chemotherapy, radiation therapy, mechanical ventilation, major surgical amputations, and creation of artificial openings in the body (e.g., tracheostomy). The specific procedure groups included in our final HCC models are shown near the end of Table 2. Also shown in Table 2 are the hierarchies created among the procedure groups. These were established using criteria analogous to those for the diagnostic hierarchies. When procedure groups are included, we call the model the Hierarchical Coexisting Conditions and Procedures (HCCP) model. The same procedure groupings are used in both prospective and concurrent versions of the HCCP model.

Inpatient Diagnostic Groups

We also explored the incremental predictive ability of using principal inpatient diagnoses for beneficiaries who are hospital-

ized. In our prospective HCC models, we found that the predictive power of hospitalizations was concentrated in a relatively small number of diagnoses. Nearly one-half of all admissions were not associated with higher incremental cost in the subsequent year. Most of the incremental explanatory power of previous year hospitalization is concentrated in just a few principal inpatient conditions: chronic obstructive pulmonary disease, congestive heart failure, metastatic cancer, and high-cost mental disorders. We consolidated the diagnoses associated with hospitalization in the previous year into 5 groups based on similarity of their incremental costliness (i.e., their regression coefficients). With the addition of the inpatient groups, the HCCP model is known as the HCCPH model, with the H denoting hospitalizations.

Not surprisingly, hospitalization is a very strong predictor of total Medicare expenditures in the current year. However, there is a wide range in the costs of enrollees who are hospitalized according to diagnosis. We reclassified principal inpatient diagnoses into six groups based on current year incremental costliness. Together with the concurrent diagnostic and the procedure groups, these inpatient groups define the concurrent HCCPH model. Altogether, then, we define six HCC models: prospective and concurrent versions of the HCC, HCCP, and HCCPH models.

Creating Appropriate Incentives

High explanatory power of a risk-adjustment model is desirable in order to create incentives for HMOs to enroll and appropriately treat high-cost individuals. Yet diagnosis-based risk-adjustment models can also create undesirable incentives for providers. Provider incentives that are of concern are primarily of two types: (1) incentives for coding of diagnoses (and

procedures) on Medicare claims; and (2) incentives for the provision of appropriate and cost-effective medical care. Generally, there is a tradeoff between explanatory power and provider incentives. The incentive facing providers can often be improved by reducing the types of information used for payment, but improved incentives come at the expense of ability to predict expenditures accurately.

In addition to incentives, one is concerned with fairness to providers and health plans. Ideally, payments should be relatively insensitive to variations in coding practices and to treatment choices such as rates of hospitalization, institutionalization, and procedures. At the same time, to be fair, a payment system should accurately reflect actual differences in enrollee health status across health plans. Thus, fairness demands power in explaining exogenous health status differences across plans while minimizing use of endogenous information on factors that are affected by plan style of care and coding practices.

To improve model incentives and fairness, we selected only a subset of HCCs for the models presented in this article. The goals of these exclusions were to reduce:

- sensitivity to variations in provider coding practices and medical care utilization;
- sensitivity to imprecise coding;
- susceptibility to provider manipulation of coding practices to maximize reimbursement, such as upcoding and coding proliferation; and
- incentives for excessive diagnostic testing or screening to identify health plan enrollees with reimbursable diagnoses.

We excluded from the models categories of diagnosis, procedure, or hospitalization that were not predictive of significantly higher Medicare expenditures, medically ambiguous, have relatively ambiguous cri-

teria for coding on claims, or are difficult to audit or verify. These decisions were based on both clinical judgment and empirical evidence on the future costliness of diagnoses. Our final, most preferred prospective HCC model includes only 34 of the initial 66 HCC diagnostic categories considered. Our final concurrent HCC model includes 44 of 81 initial categories. Similarly, the HCCP and HCCPH models reflect considerable exclusion of HCCs to improve incentives.⁵ Many of the most common diagnoses (osteoarthritis, high cholesterol, hypertension, symptoms) are eliminated, with minor loss in explanatory power, to focus on the less frequent high-cost diagnoses (Table 2). Forty-three percent of our sample of Medicare beneficiaries had no diagnoses remaining in the final prospective HCC model, and are classified only by age and sex.

Parameter Estimation

Multiple linear regression was used to estimate the parameters of each of the risk model variants described in Table 1. Annualized 1992 Medicare program expenditures were regressed against dummy variables that reflect the diagnostic, procedure, and hospitalization categories plus the 12 age-sex cells used in the current AAPCC methodology. For the prospective models, diagnoses, procedures, and hospitalizations were derived from 1991 Medicare claims; for the concurrent models, they were derived from 1992 claims. Regressions are weighted by the portion of the year each beneficiary is alive and eligible for Medicare. Because of the large number of parameters and alternative specifications, we present parameters from only two regression models in Table 2, the prospective HCC and HCCP models.

⁵For an expanded discussion of alternative specifications, see Ellis et al. (1996).

The coefficients of both prospective HCC models have good face validity. For example, metastatic cancer has a larger incremental cost than high-cost cancer, which has a larger incremental cost than moderate-cost cancer. Clinically more significant disorders have larger incremental costs than less significant disorders. Quadriplegia and paraplegia, metastatic cancer, liver disease, and respiratory arrest have some of the largest coefficients, for example. The life-sustaining procedures identify very costly patients, especially tracheostomy and dialysis. Payment weights are measured accurately: relative standard errors of coefficients in the two payment models are small, 10 percent or less, with only a few exceptions. Coefficients from the HCCPH and concurrent HCC models are presented in Ellis et al. (1996). Concurrent model parameters display similar patterns, also with good face validity, but tend to be considerably larger, especially for certain acute conditions.

Explanatory Power

In this section we evaluate the predictive ability of our risk-adjustment models. To avoid overstating predictive power because of overfitting, all of our predictive power statistics are calculated using the validation half of our sample.

Percentage of Variation Explained

Table 3 summarizes the explanatory power of the eleven risk adjustment models as measured by the R^2 statistic. The R^2 measures the proportion of the total variance in the dependent variable that is explained by the explanatory variables.

Prospective Models

The left-hand column of Table 3 presents R^2 from the prospective risk-adjustment

Table 3
Percentage of Variance (R^2) Explained by Selected Models: Validation Sample

| Label | Prospective Models | Concurrent Models |
|--|--------------------|-------------------|
| Percent | | |
| Demographic Model | | |
| Adjusted Average per Capita Cost (AAPCC) | 1.02 | 1.02 |
| DCG Models | | |
| Principal Inpatient Diagnostic Cost Groups (PIPDCG) | 5.53 | 41.95 |
| All-Diagnoses Diagnostic Cost Groups (ADDCG) | 6.34 | 33.04 |
| HCC Models | | |
| Hierarchical Coexisting Conditions Model (HCC) | 8.08 | 40.74 |
| Hierarchical Coexisting Conditions and Procedures Model (HCCP) | 8.73 | 46.59 |
| Hierarchical Coexisting Conditions, Procedures, and Hospitalizations Model (HCCPH) | 9.01 | 54.74 |

NOTES: All models include 12 age-sex cells. The dependent variable for all models is annualized 1992 Medicare payments. Prospective models use diagnoses, procedures, and hospitalizations on 1991 claims whereas concurrent models use 1992 diagnoses, procedures, and hospitalizations.

SOURCE: 1991 and 1992 Medicare Claims.

models. Four points are salient. First, all models incorporating diagnostic information have vastly greater explanatory power than our AAPCC model. The R^2 for our AAPCC model is 1.02 percent, whereas the lowest R^2 among the models incorporating diagnosis (the PIPDCG model) is 5.53 percent, more than a five-fold improvement. Second, the all-diagnoses DCG model demonstrates a surprisingly modest improvement over the PIPDCG model, with an increase of only 0.81 percentage points in the R^2 . Knowing the most serious inpatient diagnosis achieves 87 percent of the predictive power of knowing all (inpatient and outpatient) diagnoses in a DCG framework.⁶

Third, prospective HCC models have greater explanatory power than prospective DCG models that use equivalent inform-

ation. For example, the HCC model ($R^2 = 8.08$ percent) uses essentially the same information as the ADDCG model ($R^2 = 6.34$ percent). Fourth, only a modest amount of explanatory power is lost by excluding hospitalizations and procedures entirely from a prospective HCC payment model. Excluding both hospitalizations and procedures from the payment model lowers the R^2 by about 1 percentage point, from 9.01 percent to 8.08 percent, a 10-percent drop.

Explaining at best only 9 percent of the variation in Medicare payments, as the prospective risk-adjustment models do, may seem disappointing, leaving a full 91 percent unexplained. Yet much of medical expenditures are associated with inherently random events that are unpredictable even by a hypothetically perfect prospective model. The maximum explainable portion of medical expenditure variation is estimated at only 20-25 percent (Newhouse, 1995; van Vliet and van de Ven, 1993). Thus, the models presented in this article may explain nearly one-half of the explainable variance. Moreover, it is precisely this explainable portion of the dispersion in medical expenditures that is important to predict. It is the observable aspects of health and other characteristics predictably associated with future medical expenditures (e.g., chronic medical conditions) that health plans and beneficiaries can use as a basis for selection behavior. Random medical occurrences are unpredictable, and thus are true insurable events. The risk from them can be minimized by averaging through sufficiently large enrollee pools.

Concurrent Models

As previously discussed, we also estimated DCG and HCC model variants in which the classification of DXGROUPs into DCGs and HCCs was optimized to predict

⁶We also tested a specification in which we separately distinguish principal inpatient diagnosis from all other diagnoses in a DCG framework. This specification (not shown in Table 3) obtained an R^2 of 7.08, a further improvement of 0.74 over the ADDCG model.

1992 Medicare payments using 1992 (concurrent year) information instead of 1991 (prior year) information. The R^2 statistics from five concurrent risk adjustment models are shown in the far right column of Table 3. As before, these are calculated from the validation sample, and hence do not reflect possible overfitting.

The R^2 s from concurrent models are substantially higher than for the prospective models, ranging from 33.04 percent for the ADDCG model to 54.74 percent for the HCCPH model. The PIPDCG model does better than the ADDCG model, probably largely because PIPDCGs distinguish people who were hospitalized in 1992 from those who were not.⁷ The concurrent HCC model achieves an R^2 of 40.74 percent.

Incorporating procedural and hospitalization information into the concurrent HCC models results in a larger improvement in the R^2 than when it is included in the prospective models. Adding 11 procedure groups improves the R^2 by 5.85 percentage points, and adding hospitalizations improves the R^2 by a further 8.15 percentage points. This is not a surprising result, because procedural and hospitalization information is signaling what is done to a patient. If an expanded set of procedures and hospitalization dummies were used, even more of the variation could be explained, but at the cost of compromising incentives to avoid unnecessary treatments.

If the R^2 were the only measure of predictive power, then the concurrent models would be the clear favorites. Our concurrent risk-adjustment models explain much more of the variation in same year payments than prospective models, in part because they are adjusting payments for acute conditions (heart attack, pneumonia, stroke) which although expensive, are

difficult to predict prospectively. Yet this information is also difficult for enrollees or health plans to predict and use for selection, thus making it less important for risk adjustment. The advantage of concurrent over prospective models is less clear when relevant information potentially available to enrollees and plans is used to evaluate predictive power.

Predictive Ratios

Table 4 shows predictive ratios for our AAPCC model, five prospective risk-adjustment models, and five concurrent risk adjustment models for 51 non-random groups of beneficiaries from our validation sample. The predictive ratio is calculated as the total predicted 1992 payment for a group divided by the actual 1992 payment for that same group. A model performs well for a group when its predictive ratio is close to one; this indicates that aggregate payments under the risk-adjustment model will be very close to payments under the existing FFS system. The diagnostic codes of the chronic condition groups for validation were defined by a physician at HCFA without our input. Chronic condition validation groups are assigned from diagnoses on 1991 (prior year) claims.

All models do well when comparing across subgroups of the population that are defined purely by age and sex, factors used for risk adjustment. These predictive ratios are close to, but not exactly, one, because of sampling error resulting from the split sample design.

For chronic conditions, all diagnosis-based models do considerably better than our AAPCC model. Across a wide range of chronic conditions, the predictive ratios for our AAPCC range between 0.40-0.84, indicating that the AAPCC is underpaying for these people. However, under the prospective HCCP model, for example, this meas-

⁷A simple binary variable for hospitalization in 1992 (plus 12 age and sex cells) achieves an R^2 of 31.8 percent.

Table 4
Predictive Ratios for Alternative Risk Adjustment Models by Subgroup*

| Validation Group | Prospective Models | | | | | | | Concurrent Models | | | | | | | | |
|---------------------------------------|--------------------|--------|-------|------|------|-------|--------|-------------------|------|------|-------|--------|-------|------|------|-------|
| | AAPCC | PIPDCG | ADDCG | HCC | HCCP | HCCPH | PIPDCG | ADDCG | HCC | HCCP | HCCPH | PIPDCG | ADDCG | HCC | HCCP | HCCPH |
| All Enrollees | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| Aged | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| Disabled | 1.01 | 1.00 | 1.01 | 1.00 | 1.00 | 1.00 | 0.99 | 1.00 | 1.00 | 1.00 | 1.00 | 0.99 | 1.00 | 0.99 | 1.00 | 0.99 |
| Female, Under 65 Years of Age | 1.01 | 1.00 | 1.01 | 1.01 | 1.01 | 1.01 | 1.00 | 1.00 | 1.01 | 1.01 | 1.01 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| Female, 65-69 Years of Age | 1.02 | 1.02 | 1.02 | 1.02 | 1.02 | 1.02 | 1.01 | 1.02 | 1.02 | 1.02 | 1.02 | 1.01 | 1.02 | 1.01 | 1.01 | 1.01 |
| Female, 70-74 Years of Age | 0.99 | 0.98 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 1.00 | 0.99 |
| Female, 75-79 Years of Age | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| Female, 80-84 Years of Age | 1.02 | 1.02 | 1.01 | 1.02 | 1.02 | 1.02 | 1.02 | 1.02 | 1.02 | 1.02 | 1.02 | 1.02 | 1.01 | 1.02 | 1.01 | 1.01 |
| Female, 85 Years of Age or Older | 1.01 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.01 | 1.00 | 1.00 | 1.00 | 1.00 | 1.01 | 1.00 | 1.00 | 0.99 | 1.00 |
| Male, Under 65 Years of Age | 0.99 | 0.98 | 1.00 | 0.99 | 0.99 | 0.99 | 0.98 | 1.00 | 0.99 | 0.99 | 0.99 | 0.98 | 1.00 | 0.99 | 1.00 | 0.99 |
| Male 65-69 Years of Age | 1.01 | 1.01 | 1.01 | 1.01 | 1.01 | 1.01 | 1.01 | 1.01 | 1.01 | 1.01 | 1.01 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| Male, 70-74 Years of Age | 1.01 | 1.01 | 1.01 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| Male, 75-79 Years of Age | 1.00 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 1.00 | 0.99 | 0.99 | 0.99 | 1.00 | 1.00 | 1.00 | 0.99 | 0.99 |
| Male, 80-84 Years of Age | 0.99 | 0.98 | 0.99 | 0.98 | 0.98 | 0.98 | 0.98 | 0.99 | 0.98 | 0.98 | 0.98 | 0.99 | 0.99 | 0.97 | 0.98 | 0.98 |
| Male, 85 Years of Age or Older | 1.00 | 1.00 | 1.00 | 0.99 | 1.00 | 0.99 | 1.00 | 1.00 | 1.00 | 0.99 | 0.99 | 1.00 | 0.99 | 1.00 | 1.00 | 0.99 |
| Any 1991 Chronic Condition | 0.82 | 0.89 | 0.96 | 0.98 | 0.96 | 0.98 | 0.89 | 0.96 | 0.98 | 0.96 | 0.98 | 0.94 | 0.96 | 0.99 | 0.99 | 0.98 |
| Depression | 0.58 | 0.81 | 0.80 | 0.87 | 0.87 | 0.87 | 0.80 | 0.80 | 0.87 | 0.87 | 0.90 | 0.87 | 0.84 | 0.93 | 0.92 | 0.94 |
| Alcohol and Drug Dependence | 0.46 | 0.79 | 0.85 | 0.96 | 0.96 | 0.96 | 0.85 | 0.85 | 0.96 | 0.99 | 0.99 | 0.85 | 0.82 | 0.92 | 0.93 | 0.95 |
| Hypertensive Heart-Renal Disease | 0.69 | 0.83 | 0.89 | 0.93 | 0.87 | 0.93 | 0.89 | 0.87 | 0.94 | 0.94 | 0.94 | 0.87 | 0.90 | 0.97 | 0.97 | 0.95 |
| Benign/Unspecified Hypertension | 0.84 | 0.92 | 0.96 | 0.97 | 0.97 | 0.97 | 0.92 | 0.96 | 0.97 | 0.97 | 0.97 | 0.95 | 0.96 | 0.97 | 0.97 | 0.98 |
| Diabetes With Complications | 0.45 | 0.69 | 0.61 | 0.93 | 0.94 | 0.94 | 0.69 | 0.61 | 0.94 | 0.94 | 0.94 | 0.78 | 0.78 | 0.97 | 0.96 | 0.94 |
| Diabetes Without Complications | 0.63 | 0.75 | 0.85 | 1.02 | 1.02 | 1.02 | 0.75 | 0.85 | 1.02 | 1.02 | 1.02 | 0.87 | 0.88 | 0.99 | 0.99 | 0.97 |
| Heart Failure/Cardiomyopathy | 0.48 | 0.74 | 0.87 | 0.98 | 0.98 | 0.98 | 0.74 | 0.87 | 0.98 | 0.98 | 0.98 | 0.83 | 0.82 | 0.97 | 0.97 | 0.95 |
| Acute Myocardial Infarction | 0.45 | 0.79 | 0.83 | 0.87 | 0.89 | 0.87 | 0.79 | 0.83 | 0.87 | 0.89 | 0.90 | 0.80 | 0.81 | 0.92 | 0.92 | 0.92 |
| Other Heart Disease | 0.66 | 0.82 | 0.88 | 0.97 | 0.97 | 0.98 | 0.82 | 0.88 | 0.97 | 0.98 | 0.98 | 0.89 | 0.90 | 0.98 | 0.98 | 0.96 |
| Chronic Obstructive Pulmonary Disease | 0.61 | 0.80 | 0.82 | 0.98 | 0.98 | 0.98 | 0.80 | 0.82 | 0.98 | 0.98 | 0.98 | 0.88 | 0.87 | 0.99 | 0.99 | 0.96 |
| Colorectal Cancer | 0.54 | 0.76 | 0.87 | 0.93 | 1.00 | 1.00 | 0.76 | 0.87 | 0.93 | 1.00 | 1.00 | 0.82 | 0.91 | 1.01 | 1.02 | 0.96 |
| Breast Cancer | 0.68 | 0.78 | 0.93 | 1.08 | 1.06 | 1.06 | 0.78 | 0.93 | 1.08 | 1.06 | 1.06 | 0.87 | 1.02 | 1.18 | 1.13 | 1.02 |
| Lung or Pancreas Cancer | 0.32 | 0.64 | 0.89 | 0.92 | 0.94 | 0.94 | 0.64 | 0.89 | 0.92 | 0.94 | 0.96 | 0.74 | 0.80 | 0.97 | 0.97 | 0.93 |
| Other Stroke | 0.53 | 0.76 | 0.83 | 0.98 | 0.98 | 0.98 | 0.76 | 0.83 | 0.98 | 0.98 | 0.98 | 0.82 | 0.88 | 1.03 | 1.02 | 0.97 |
| Intracerebral Hemorrhage | 0.40 | 0.66 | 0.72 | 0.86 | 0.89 | 0.89 | 0.66 | 0.72 | 0.86 | 0.89 | 0.89 | 0.74 | 0.78 | 0.94 | 0.93 | 0.88 |
| Hip Fracture | 0.59 | 0.66 | 0.85 | 0.99 | 1.00 | 1.00 | 0.66 | 0.85 | 0.99 | 1.00 | 0.99 | 0.87 | 1.01 | 1.15 | 1.14 | 0.99 |
| Arthritis | 0.78 | 0.85 | 0.92 | 0.91 | 0.91 | 0.91 | 0.85 | 0.92 | 0.91 | 0.91 | 0.91 | 0.94 | 0.93 | 0.93 | 0.92 | 0.95 |

See notes at end of table.

Table 4—Continued
Predictive Ratios for Alternative Risk Adjustment Models by Subgroup*

| Validation Group | Prospective Models | | | | | | | Concurrent Models | | | | |
|---|--------------------|--------|--------|--------|--------|--------|--------|-------------------|-------|-------|-------|--|
| | AAPCC | PIPDCG | ADDCG | HCC | HCCP | HCCPH | PIPDCG | ADDCG | HCC | HCCP | HCCPH | |
| First (Lowest) Quintile, 1991 Expenditures | 2.49 | 1.92 | 1.22 | 1.30 | 1.26 | 1.27 | 1.51 | 1.23 | 1.09 | 1.11 | 1.18 | |
| Second Quintile, 1991 Expenditures | 1.78 | 1.37 | 1.37 | 1.24 | 1.21 | 1.20 | 1.30 | 1.28 | 1.12 | 1.13 | 1.13 | |
| Middle Quintile, 1991 Expenditures | 1.31 | 1.01 | 1.24 | 1.14 | 1.11 | 1.09 | 1.13 | 1.19 | 1.13 | 1.12 | 1.08 | |
| Fourth Quintile, 1991 Expenditures | 0.91 | 0.78 | 1.02 | 0.99 | 0.97 | 0.95 | 0.98 | 1.03 | 1.05 | 1.04 | 1.01 | |
| Fifth (Highest) Quintile, 1991 Expenditures | 0.48 | 0.85 | 0.78 | 0.85 | 0.88 | 0.90 | 0.80 | 0.80 | 0.88 | 0.89 | 0.90 | |
| First (Lowest) Quintile, 1992 Expenditures | 821.67 | 684.17 | 522.57 | 514.41 | 505.56 | 506.17 | 185.18 | 116.71 | 81.48 | 86.86 | 54.44 | |
| Second Quintile, 1992 Expenditures | 23.03 | 20.43 | 19.94 | 18.66 | 18.44 | 18.35 | 5.24 | 9.11 | 6.57 | 6.49 | 3.18 | |
| Middle Quintile, 1992 Expenditures | 6.92 | 6.56 | 7.05 | 6.78 | 6.72 | 6.67 | 1.63 | 4.10 | 3.61 | 3.49 | 1.64 | |
| Fourth Quintile 1992 Expenditures | 1.79 | 1.85 | 2.01 | 2.03 | 2.03 | 2.02 | 1.07 | 1.74 | 1.68 | 1.63 | 1.19 | |
| Fifth (Highest) Quintile, 1992 Expenditures | 0.24 | 0.31 | 0.32 | 0.34 | 0.35 | 0.35 | 0.88 | 0.68 | 0.74 | 0.75 | 0.91 | |
| 0 1991 Hospital Admissions | 1.30 | 1.00 | 1.10 | 1.05 | 1.04 | 1.02 | 1.09 | 1.10 | 1.05 | 1.05 | 1.03 | |
| 1 1991 Hospital Admissions | 0.64 | 1.16 | 0.97 | 0.99 | 1.01 | 1.03 | 0.92 | 0.91 | 0.96 | 0.96 | 0.98 | |
| 2 1991 Hospital Admissions | 0.46 | 0.97 | 0.82 | 0.91 | 0.94 | 0.97 | 0.84 | 0.83 | 0.92 | 0.92 | 0.94 | |
| 3+ 1991 Hospital Admissions | 0.29 | 0.73 | 0.62 | 0.77 | 0.82 | 0.87 | 0.73 | 0.69 | 0.82 | 0.82 | 0.86 | |
| 0 1992 Hospital Admissions | 4.51 | 4.21 | 4.18 | 4.09 | 4.06 | 4.04 | 1.00 | 2.35 | 2.07 | 2.07 | 0.99 | |
| 1 1992 Hospital Admissions | 0.39 | 0.45 | 0.46 | 0.47 | 0.48 | 0.48 | 1.38 | 0.94 | 0.91 | 0.90 | 1.22 | |
| 2 1992 Hospital Admissions | 0.20 | 0.26 | 0.27 | 0.29 | 0.29 | 0.30 | 0.97 | 0.70 | 0.75 | 0.75 | 0.97 | |
| 3+ 1992 Hospital Admissions | 0.12 | 0.18 | 0.18 | 0.21 | 0.21 | 0.22 | 0.64 | 0.50 | 0.65 | 0.66 | 0.80 | |

*Ratio of predicted to actual 1992 expenditures.

NOTE: Chronic conditions are defined based on 1991 diagnoses, Prospective risk-adjustment models use 1991 diagnoses, procedures, and hospitalizations. Concurrent models use 1992 information. Predicted and actual expenditures in the predictive ratios are for 1992.

SOURCE: 1991 and 1992 Medicare Claims, 2.5-Percent Validation Sample.

ure ranges from a low of 0.87 for depression to a high of 1.06 for breast cancer. For people with any of these chronic conditions in 1991, our AAPCC underpays by 18 percent on average, the prospective HCCP model by only 2 percent. For most of the chronic conditions, predicted payments from the HCC models are within \$500 of actual payments, compared with typical deviations of several thousand dollars under the AAPCC. Thus, the HCC models greatly reduce incentives for favorable selection.

The prospective models predict costs for previously-diagnosed chronic conditions nearly as well as the concurrent models, despite their much lower R^2 s. The R^2 advantage of concurrent models clearly lies in explaining expenditure variation associated with acute or newly diagnosed conditions, not with pre-existing chronic conditions that could be used by health plans to avoid high-cost enrollees.

Looking from left to right across the columns for the different prospective or concurrent models demonstrates a general improvement in these measures, consistent with their overall explanatory power as measured by R^2 . Multiple-condition HCC models generally do better than the single-diagnosis DCG models. The HCC model—which omits many of the more problematic and discretionary diagnoses, does not reward hospitalizations relative to outpatient treatment, and ignores procedure information—achieves predictive ratios that compare favorably with the other models.

Many of the diagnoses falling into the chronic conditions selected by HCFA are themselves used for risk adjustment under the DCG and HCC frameworks. A more stringent test is to look at 1991 expenditure quintiles and 1991 hospital admission groups. Again, all models do substantially better than our AAPCC, and there is a general improvement moving to the more highly predictive models. Prospective and

concurrent models do almost equally well. The HCC and HCCP models only underpay the 1991 highest-expenditure quintile by 10 to about 15 percent, whereas our AAPCC underpays by more than 50 percent. Enrollees with three or more hospitalizations in 1991 are underpaid by about 20 percent under the HCC and HCCP models, versus a 70-percent underpayment with our AAPCC.

Incorporating diagnosis substantially reduces the opportunities for risk selection based on prior utilization, but does not eliminate them. The average profit in 1992 from enrolling someone in the lowest quintile of 1991 expenditures risk adjusting by our AAPCC is \$2,134. Using the prospective HCC model to risk adjust lowers this potential profit to \$424. Similarly, the average loss from enrolling someone in the highest quintile of 1991 expenditures is -\$4,425 under the AAPCC, and -\$1,311 using the HCC model.

Concurrent models match payments to expenditures considerably better than prospective models or our AAPCC for concurrent (1992) expenditure quintiles and numbers of hospitalizations, consistent with their much higher R^2 in explaining 1992 expenditures. For example, the concurrent HCC model underpredicts expenditures for the highest 1992 quintile by only 26 percent, compared with 76 percent by our AAPCC and 66 percent by the prospective HCC model.

Distribution of Expenditures

Medical expenditures are highly skewed: In a given year, most people have relatively modest expenditures, but a few have very large expenditures. The far right column of Table 5 illustrates that in 1992, three-fourths of Medicare beneficiaries in our sample cost less than \$2,917, whereas the top one percent cost more than \$57,000 each. Very high expenditures may represent unpredictable acute

medical crises that no prospective risk-adjustment model can predict. However, good risk-adjustment models should, to some degree, reproduce the highly skewed nature of medical expenditures by predicting the upper tail of the distribution.

Table 5 shows that our AAPCC model cannot predict the tail. Its maximum predicted expenditures is only \$7,710, and its range from maximum to minimum is only \$5,324. Adding diagnostic information allows much higher-cost individuals to be identified. For example, the prospective ADDCG model has a maximum payment of \$21,543. The multiple condition HCC models predict greater maximum expenditures than the single-condition DCG models because the sickest individuals tend to suffer from multiple medical problems. The life-sustaining procedures are particularly useful in predicting the very high costs of a small number of individuals whose full expense cannot be ascertained from diagnoses alone. Incorporating hospitalizations raises predictions for the upper 5-10 percent of enrollees. Concurrent models, with their ability to capture expenditures associated with acute medical events, achieve roughly double the predicted amounts of the corresponding prospective models in the upper tail.

Because the mean predicted expenditure is the same for all models (at \$3,773), the extended upper tail of the diagnosis-based models is achieved by paying less at the lower and middle parts of the distribution. Our AAPCC model pays a minimum of \$2,386, whereas the prospective HCC and HCCP models pay only about \$1,100 for the youngest and healthiest beneficiaries (i.e., for a 65- to 69-year-old female with no diagnoses included in the payment models). The ADDCG model, which incorporates information from all diagnoses, not just the higher, cost conditions that are the focus of the HCCs, does a slightly better job at predicting the lower cost end of the distribution than the HCC models.

All concurrent models except the PIPDCG generate negative predicted costs for some individuals, at the very lowest percentiles of the distribution. This occurs because the coefficients on the oldest age groups are negative in all of the concurrent models that we estimated using all diagnoses. Our explanation for this is that the intensity of treatment for the most elderly individuals with a given condition is lower than for younger populations. The oldest individuals are also the most likely to have multiple conditions. The age dummies are attempting to adjust payments downward to offset this higher predicted payments, resulting in negative predictions for those in the oldest age groups not identified with any medical conditions included in the model. This problem could possibly be eliminated by exploring the use of interactions between age and the HCCs, which was beyond the scope of this project, or by omitting age-sex variables from the concurrent risk-adjustment models.

ADDITIONAL SPECIFICATIONS

We also investigated the usefulness of several alternative specifications. These variants differ in their dependent variables, samples, or in their explanatory variables (Table 6). The HCC model is the baseline for evaluating the explanatory power of additional variables or models. All models presented in Table 6 are prospective, and all R^2 statistics are calculated from the model development sample half, and thus, are not directly comparable to the validation sample R^2 s reported in Table 3. For comparison, the prospective HCC model's R^2 calculated on the development sample (8.62 percent) is shown at the top of Table 3.

Aged Versus Disabled Subsamples

Medicare currently has separate AAPCC risk-adjustment factors for aged and dis-

Table 5
Distribution of Predicted 1992 Expenditures From Alternative Risk Adjustment Models and Actual 1992 Expenditures

| Percentile | AAPCC* | Prospective Models | | | | | Actual Expenditures |
|------------|---------|--------------------|----------|----------|----------|----------|---------------------|
| | | PIPDCG | ADDCG | HCC | HCCP | HCCPH | |
| Maximum | \$7,710 | \$26,324 | \$21,543 | \$44,137 | \$78,176 | \$75,363 | |
| 99 | 6,642 | 14,768 | 14,264 | 16,364 | 16,724 | 17,491 | |
| 95 | 5,687 | 9,330 | 9,495 | 10,202 | 10,201 | 10,397 | |
| 90 | 5,085 | 7,236 | 7,324 | 7,818 | 7,762 | 7,688 | |
| 75 | 4,571 | 3,935 | 5,031 | 4,853 | 4,826 | 4,726 | |
| 50 | 3,614 | 2,899 | 2,919 | 2,845 | 2,803 | 2,790 | |
| 25 | 2,902 | 2,258 | 1,965 | 1,796 | 1,751 | 1,782 | |
| 10 | 2,386 | 1,859 | 1,239 | 1,411 | 1,349 | 1,371 | |
| 5 | 2,386 | 1,859 | 892 | 1,162 | 1,098 | 1,113 | |
| 1 | 2,386 | 1,859 | 735 | 1,162 | 1,098 | 1,113 | |
| Minimum | 2,386 | 1,859 | 735 | 1,162 | 1,098 | 1,113 | |

| Percentile | Concurrent Models | | | | | Actual Expenditures |
|------------|-------------------|----------|----------|-----------|-----------|---------------------|
| | PIPDCG | ADDCG | HCC | HCCP | HCCPH | |
| Maximum | \$38,356 | \$40,448 | \$97,955 | \$200,253 | \$205,806 | \$1,533,060 |
| 99 | 31,207 | 27,548 | 34,570 | 35,089 | 36,695 | 57,423 |
| 95 | 23,519 | 17,801 | 18,730 | 18,114 | 20,283 | 22,810 |
| 90 | 14,060 | 13,345 | 12,175 | 11,790 | 13,767 | 12,227 |
| 75 | 1,413 | 4,639 | 4,442 | 4,369 | 3,279 | 2,917 |
| 50 | 851 | 1,045 | 1,086 | 1,006 | 284 | 516 |
| 25 | 682 | 619 | 204 | 261 | 195 | 95 |
| 10 | 546 | 336 | 96 | 193 | 160 | 0 |
| 5 | 546 | 22 | -64 | 76 | -74 | 0 |
| 1 | 393 | -227 | -407 | -179 | -101 | 0 |
| Minimum | 393 | -227 | -407 | -179 | -101 | 0 |

* Represented by 12 age-sex cells and Medicaid eligibility.

NOTE: Predicted values are for validation sample half. All expenditures (including "actual expenditures") represent annualized amounts.

SOURCE: 1991 and 1992 Medicare claims

abled beneficiaries. We tested whether substantial differences exist between the estimated parameters of the HCC payment model for the aged and disabled sub-populations. Although the percentage of variance explained is higher among the disabled than the elderly (12.2 versus 8.4 percent), the estimated parameters are remarkably similar on the whole. Thus, allowing different coefficients for the aged and disabled only raises the combined sample R^2 from 8.62 percent to 8.69 percent (Table 6). Therefore, a combined risk-adjustment model for the aged and disabled is appropriate. Real differences do exist for substance abuse and high-cost psychiatric diagnoses, with the disabled considerably more expensive than the elderly. These differences could be recognized in a combined model by paying extra for disabled beneficiaries with these diagnoses.

Medicaid Status

Medicare's current AAPCC methodology uses Medicaid enrollment status as a risk-adjustment factor. Medicaid status adds a modest amount of predictive power to the HCC model, raising the R^2 from 8.62 to 8.71 percent. Medicaid enrollees are nearly \$1,000 more expensive than non-Medicaid enrollees holding constant age, sex, and diagnosis, providing a basis for risk selection by health plans. Including Medicaid status as a risk adjuster could improve access to care of dual Medicare-Medicaid eligibles by eliminating incentives for health plans to avoid them. On the other hand, Medicaid eligibility rules vary across the States, and it is not clear that Medicaid status is a proxy for exogenous differences in health stat-

Table 6
Explanatory Power of Alternative Prospective Models, Model Development Sample*

| Model or Factor | R ² | Comments |
|--|----------------|---|
| Base Case | Percent | |
| HCC Model | 8.62 | Base case for comparison |
| Sub-Samples | | |
| Separate Models for Aged and Disabled | 8.69 | With few exceptions, parameter estimates are not substantially different. Substance abuse and mental health expenditures are higher for the disabled. |
| Risk Adjusters Added Individually to HCC Model | | |
| Medicaid Eligibility | 8.71 | Coefficient = \$958, Standard Error = \$39. Eligibility rules vary by State. |
| Linear Age Plus Sex Dummy (Replaces 12 Age and Sex Cells) | 8.62 | Negligible gain for either aged or disabled subsamples. |
| 1991 Expenditures | 9.79 | Coefficient = \$0.21, Standard Error < \$0.01. |
| Cancer, Heart Disease, Stroke, Diabetes, COPD Interactions | 8.64 | COPD is chronic obstructive pulmonary disease. |
| Alternative Risk Adjusters | | |
| Cancer, Heart Disease, and Stroke Plus Age and Sex | 3.82 | — |
| Cancer, Heart Disease, Stroke, Diabetes, COPD Plus Age and Sex | 4.93 | — |
| Cancer, Heart Disease, Stroke, Diabetes, COPD Plus Interactions and Age and Sex | 5.02 | — |
| 1991 Expenditures Plus Age and Sex | 7.04 | Coefficient = \$0.38, Standard Error < \$0.01. |
| 1991 Hospitalization Dummy Plus Age and Sex | 3.94 | PIPDCG model R ² = 5.88 Percent |
| Transformed Dependent Variable | | |
| Expenditures Deflated by Geographic Input Price Index (GIP _I) | 8.88 | GIP _I measures area variation in wages and other prices. |
| Top-Coded at \$50,000 | 13.93 | Simulates outlier pool with \$50,000 threshold. |
| Top-Coded at \$25,000 | 14.83 | Simulates outlier pool with \$25,000 threshold. |
| Logged (1 + \$) | 18.75 | Medical expenditures are highly skewed. |
| Continuous Update Model | | |
| HCC Model | 24.08 | One month's expenditures are predicted using the preceding 12 months' diagnoses. |

*Because these R²s are computed on the model development sample, they are not directly comparable to the R²s in Table 3 computed on the validation sample. For example, the HCC model has an R² of 8.62 percent on the model development sample and an R² of 8.08 percent on the validation sample. In general, the validation sample R²s will be lower.

SOURCE: 1991 and 1992 Medicare claims.

us.⁸ Whether to include Medicaid status in a Medicare risk adjustment model is a decision for policymakers.

Simplified Lists of Conditions

Excluding many diagnoses from payment models reduces their explanatory

⁸For example, the higher costs of Medicaid eligibles could be because of their use of higher price sources of care, such as hospital emergency rooms, rather than physicians' offices, an inefficiency that should be eliminated by a well run managed care plan.

power only slightly, raising the question of how far exclusions and aggregations can proceed without substantially reducing explanatory power. We investigated this by estimating two prospective models with highly simplified lists of coexisting conditions. The first includes the three leading killers of Americans: heart disease, cancer, and stroke. These three conditions plus age and sex explain nearly four times as much of the variance in expenditures as

the AAPCC model, but less than one-half as much as the HCC model. Adding diabetes and chronic lung disease to these three raises the R^2 by another percentage point to nearly 5 percent. Although better than the AAPCC, it still falls well short of the HCC model. If complete and accurate diagnostic information is available, we recommend use of the HCC model. The simplified models may be useful in situations of incomplete information, possibly for new Medicare enrollees or when only self-reported medical conditions are available.

Interactions Among Conditions

We also investigated whether accounting for interactions among medical conditions would add substantially to the explanatory power of diagnosis-based risk adjustment models. We first added the 10 first-order interaction terms among the five conditions in the "simplified conditions" model previously described to that model. Adding the interactions increased the R^2 only slightly, from 4.93 percent to 5.02 percent. Next, we added the same set of 10 aggregated interactions to the HCC model, which increased its explanatory power only from 8.62 to 8.64 percent. We also tried weighting a person's conditions more or less heavily based on the total number of conditions he or she has, but found no meaningful improvement in explanatory power (Ellis et al., 1996). We conclude that a simple linear, additive relationship among multiple diagnostic categories provides a good fit to the data. Because non-linear interactions among conditions add complexity to risk-adjustment models with little apparent gain in explanatory power, we recommend the simple linear form.

Prior Utilization Measures

In previous research, prior utilization measures have been found to be the most highly predictive risk-adjustment variables

(Thomas and Lichtenstein, 1986; van Vliet and van de Ven, 1990). For comparison with our diagnosis-based models, we examined two measures of prior utilization: expenditures and hospitalization. Prior year expenditures has a fairly high explanatory power of 7.04 percent, less than the predictive power of the HCC models. When expenditures is added to the HCC payment model, the R^2 rises from 8.62 to 9.79 percent. Consistent with the predictive ratio results, there remains some possibility of risk selection within the HCC model using prior year expenditures, but much smaller opportunities than with the AAPCC. A dummy variable for prior year hospitalization achieves an R^2 of nearly 4 percent. Our diagnosis-based models thus have greater explanatory power than simple measures of prior utilization and avoid their undesirable incentives for over-provision of care.

Geographic Adjustments

Geographic adjustments and outlier pools are potential additional elements of a complete capitated payment system. Geographic adjustments account for the differential costs of providing medical care in different regions. The AAPCC's geographic adjustment is based on FFS costs in the beneficiary's county of residence. This adjustment is criticized for leading to huge inter-area variations in payments (as much as four-fold differences across counties) and unstable payment rates over time (Rossiter and Adamache, 1989; Newhouse, 1986; Welch, 1992). In addition, geographic differences reflect possibly inappropriate variations in medical practice styles. We developed an alternative geographic adjustment, the Geographic Input Price Index (GIPI) using input prices (wages, building rental rates, etc.) measured by Medicare's prospective payment system area hospital

wage index and the Medicare fee schedule Geographic Adjustment Factor for physician payment (Welch, 1992). The GIPI is computed for Metropolitan Statistical Areas and state non-metropolitan areas. Excluding Puerto Rico, the GIPI varies only from 0.785 in rural Mississippi to 1.272 in Oakland, California, where 1.000 represents the national average price level. Deflating Medicare expenditures by the GIPI adds only modest explanatory power to the HCC model: the R^2 increases only from 8.62 to 8.88 percent. Thus, little expenditure variation unexplained by diagnosis, age, and sex is accounted for by geographic differences in wages and other input prices. Nevertheless, we believe that these exogenous cost factors should be incorporated in Medicare capitation reimbursements.

Capitation Outlier Pools

Medicare does not currently use an outlier pool in HMO reimbursement, although outlier pools have been studied (Beebe, 1992; Keeler, Carter, and Trude, 1988; Ellis and McGuire, 1988) and proposed for payment demonstrations. Outlier pools offer reduced financial risk to providers, improved payment equity, and greater incentives for providers to enroll and treat very sick and expensive patients (Beebe, 1992; Keeler, Carter, and Trude, 1988). In a simple outlier pool, Medicare would reimburse an HMO for all of its enrollee's expenditures above some annual threshold amount. We simulated the effect of outlier thresholds of \$25,000 and \$50,000 on the explanatory power of the HCC model, using non-annualized expenditures to determine which cases exceeded these thresholds. High-expenditure observations were not dropped, they were simply top-coded at the capped amount. The HCC model explains a

higher proportion of capped expenditure variation than of total expenditure variation, but the difference is not dramatic. A conceptually preferable, but administratively more complex, outlier pool would be based on a variable threshold that is a fixed deductible (e.g., \$25,000) above the expenditure predicted by the HCC model (Keeler, Carter, and Trude, 1988). Thus, the threshold triggering outlier payments would be greater for beneficiaries diagnosed with cancer than for those with no diagnosed illnesses.

Continuous Update Models

The Continuous Update Model (Ellis and Ash, 1989) represents a compromise between the better incentives of the prospective model and the greater explanatory power of the concurrent model. It predicts each month's expenditures using diagnoses from the immediately preceding 12-month period. We estimated a Continuous Update Model using the HCCs defined for the prospective model and achieved an R^2 of 24.08 percent. This is better than any of the prospective models, but well short of the concurrent models. The Continuous Update Model is substantially more complex (administratively and computationally) than annual models because diagnoses and expenditures must be tracked by month.

CONCLUSIONS

Risk adjustment is increasingly recognized as a critical element of reforming Medicare's capitation payments to HMOs and other managed care entities (U.S. General Accounting Office, 1994). Risk adjustment can reduce the financial risk to HMOs of participating in Medicare and thus further the policy goal of increasing

Medicare beneficiaries' enrollment in managed care plans. It can also increase the equity of Medicare capitation payments. Risk adjustment encourages HMOs to compete on the quality and efficiency of their care rather than on attracting the healthiest enrollees, thereby improving access to HMOs of the sick and disabled.

Claims-Based Versus Other Risk Adjustment

Risk adjustment that uses diagnostic information on medical claims to adjust payments, such as the HCC model, appears to provide the best combination of ability to predict enrollee costs, incentives for appropriate care, resistance to manipulation by providers, cost effectiveness, and administrative feasibility. Surveys of self-reported enrollee health, chronic conditions, and functional status are expensive, prone to manipulation, difficult to validate, potentially unreliable, and have less predictive power (Fowles, Lawthers, and Weiner, 1995). Collecting direct clinical descriptors of health, such as blood pressure and cholesterol level through clinical examinations, is expensive, intrusive, and less powerful in explaining future utilization (Newhouse et al., 1989). Prior utilization of medical services has relatively high predictive power, but sets up inappropriate incentives for providing services (Thomas and Lichtenstein, 1986; van Vliet and van de Ven, 1990).

Purely financial risk sharing, without explicit measurement of health status, has also been proposed as a method of reforming Medicare's HMO reimbursement methodology (Ellis and McGuire, 1986; Beebe, 1992; Newhouse, 1995). The government could absorb part of the cost of caring for expensive Medicare enrollees (an outlier policy) or of all enrollees (payer cost sharing). The limitation of an outlier policy is that very high-cost cases are essentially random, and outlier payments do not direct

extra reimbursement to health plans that enroll a systematically higher-cost (i.e., less healthy) population (Ellis and McGuire, 1988). Also, the HMO has less incentive to manage the cost of very expensive cases or to avoid choosing expensive treatment modalities. Still, an outlier policy may have a role in conjunction with a risk adjuster, such as the HCC model, that accounts for systematic health status variation among enrolled populations.

An expanded outlier policy would have the government absorb some share of the total actual cost of providing care to Medicare beneficiaries, say one-half, in addition to paying a reduced, predetermined capitated amount. This hybrid of FFS and capitated payment systems, it is argued, balances the incentives for over-provision of services inherent in FFS against the incentives for under-provision of services inherent in capitation (Ellis and McGuire, 1993; Newhouse, 1995).⁹ A partial capitation system is consistent with a risk-adjustment model because the capitated portion of payment could be adjusted for beneficiary health status. However, such a system clearly reduces incentives for cost containment (this is by design) and may be unfair to efficient plans. Also, a partial capitation system (or an outlier system) would be much more difficult to implement than the HCC model because, to measure costs, it requires comprehensive service utilization data from HMOs, plus agreement on an algorithm to assign costs to utilization. In contrast, the HCC model requires only demographic and diagnostic (and, perhaps, some procedural) information.

Incentives in Risk Adjustment

One goal of risk adjustment for payment purposes is to accurately predict expenditures, but another is to establish incentives

⁹Incentives for underservice in Medicare capitation may be limited by the ability of Medicare enrollees to return to the FFS sector and competition among health plans for enrollees.

for appropriate, cost effective medical care. We considered diagnosis, certain medical procedures, and hospitalization as risk adjusters in addition to demographics. Risk adjustment based on diagnosis alone establishes the strongest incentives to avoid excessive medical care. Providers are not paid more for what they do, only for the diagnostic health status of their enrollees. Using diagnosis only to risk adjust is also fairer. Efficient plans that avoid hospitalizations and eschew aggressive, procedure-oriented styles of care are not penalized. One of our key findings is that the medical procedures we considered and hospitalizations add relatively little predictive power to diagnoses, especially in prospective models. Thus, acceptably high explanatory power can be achieved in a payment model without sacrificing strong incentives to avoid unnecessary care or rewarding aggressive, intensive styles of medical practice. In addition, we found that many common, low-cost, ambiguous or discretionary diagnoses can be excluded from payment models with limited reduction in predictive power. This exclusion greatly reduces the incentives and ability of health plans to manipulate coding practices to increase reimbursement. In short, a simple yet powerful risk-adjustment model with strong incentives to avoid excessive medical care and resistance to coding manipulation can be built from a parsimonious set of high-cost diagnoses. This is the HCC model.

Prospective Versus Concurrent Risk Adjustment

Diagnosis-based risk adjustment can be done either prospectively or concurrently. Our results show that either prospective or concurrent methods predict costs equally well, on average, for people diagnosed with chronic conditions or hospitalized in the prior year. Also, the models are equally

powerful in predicting expenditures for particularly high- or low-cost groups in the previous year. Thus, either model should attenuate incentives for risk selection by health plans about equally well.

Concurrent models explain costs in the current year much better than prospective models, but much of current year expenditure variation results from random acute medical events. By definition, these events are unpredictable and cannot be used for risk selection. Acute conditions are true insurable events that average out in relatively small random panels of enrollees (Ellis et al., 1996). Concurrent models thus have an advantage over prospective models in reducing unsystematic risk only in small enrollee groups.

Concurrent models establish poorer incentives for diagnostic coding and appropriate medical care than prospective models. Payment weights are generally larger in concurrent models, providing greater incentives for inappropriate coding of diagnoses. Moreover, the higher payment weights are attached to acute medical conditions, which, because of their transitory nature, could potentially be harder to audit and verify than chronic conditions. For example, multiple organ system failures (acute renal failure, respiratory failure) often could be coded or not for dying individuals. We excluded the diagnoses respiratory arrest and cardiac arrest from our concurrent models because they could be coded for anyone who dies. Also, certain potentially avoidable, but very high-cost, acute diagnoses (gangrene, peritonitis) that are sometimes indicators of poor quality of care (Weissman, Gatsonis, and Epstein, 1992) are paid more in a concurrent model. In short, concurrent models may be less appropriate as payment models, but particularly useful where payment incentives

are of less concern, such as for physician profiling. They also may be useful as a risk adjuster in situations where patients are triaged to providers on an acute care basis.

Diagnostic Coding Accuracy

The validity and reliability of our risk adjustment models depends on the accuracy of diagnoses coded on Medicare claims. In a preliminary study (Pope et al., 1994), we examined the internal consistency of diagnostic and other information coded on the 1991 Medicare claims in our sample. Our sense from this and other analysis (Fowles et al., 1995; Weiner et al., 1995) is that the diagnoses coded on FFS claims are probably generally accurate (i.e., actually present), but that coding of comorbidities is incomplete. The completeness of coded diagnoses would improve if they were the basis for capitated payment, but the veracity of the diagnoses would be more open to question. Rebased of payment weights will be necessary as coding practices evolve.

The diagnoses we used for model development are coded on FFS reimbursement claims. For capitated payment, HMOs and other managed care organizations would have to supply diagnostic information. Many of these organizations have not historically maintained detailed encounter-level data with diagnosis and procedure, especially for ambulatory encounters. This lack of information could prove an impediment to widespread early adoption of a risk-adjustment system incorporating ambulatory diagnoses. A phased introduction may make sense, with risk adjustment first based on widely available, accurate, and auditable diagnoses such as principal inpatient diagnoses, then proceeding to incorporate other hospital and physician diagnoses as the necessary data systems

are developed. HCFA could spur the necessary data collection by paying only the rate for a healthy person unless complete and accurate diagnostic data is supplied by a health plan. Any claims-based risk-adjustment model implemented for payment purposes will require careful monitoring to ensure that health plans do not behave undesirably.

Future Research

Several directions for future work on risk adjustment are particularly important. HCFA-funded research is currently ongoing (Arlene Ash, Principal Investigator) calibrating risk adjustment models to other samples, such as the under 65 years of age, employed, and Medicaid populations. That project is adapting the DXGROUP classification system to better reflect pregnancy-related conditions and infant, childhood, and young adult disorders that are rare among the aged Medicare population. The age profile of expenditures for certain diagnostic conditions deserves further consideration. Carve-outs for particular groups of conditions, such as mental health and substance abuse, is another useful direction for research. Calibrating the model on HMO encounter data would indicate if payment weights are affected by HMO versus FFS practice patterns.¹⁰ It would also be informative to expand the cost and expenditure measure to encompass all medical expenditures, including deductibles, copayments, Medicaid-covered expenses, drugs, dental, eye, long-term care, and other services not covered by Medicare. Finally, more work on concurrent and continuous update risk adjustment models and combinations of concurrent and prospective models is warranted.

¹⁰Dun et al. (1995) found that the performance of alternative risk adjustment models was insensitive to type of health plan (indemnity, IPA, PPO, HMO), region, etc., although the study was limited to private employed group data.

ACKNOWLEDGMENTS

We would like to acknowledge contributions to this project, and thank the following individuals: Mel Ingber, HCFA Project Officer; Debra Dayhoff, Health Economic Research, Inc. (HER) Senior Economist; Robert Baker and Fizza Gillani, HER Programmer/Analysts; and Tim Dawes, Angela Merrill, Amy Rensko, and Monika Reti, HER Junior Analysts.

REFERENCES

- Anderson, G.F., Cantor, J.C., Steinberg, E.P., and Holloway, J.: Capitation Pricing: Adjusting for Prior Utilization and Physician Discretion. *Health Care Financing Review* 8(2):27-34, Winter 1986.
- Anderson, G.F., Lupu, D., et al.: *Payment Amount for Capitated Systems*. Contract Number 17-C-98990/3-01. Prepared for the Health Care Financing Administration. Center for Hospital Finance and Management, Johns Hopkins University. Baltimore. December 1989.
- Ash, A., Ellis, R., and Iezzoni, L.: *Clinical Refinements to the Diagnostic Cost Group Model*. Cooperative Agreement Number 18-C-98526/1-03. Prepared for the Health Care Financing Administration. Baltimore, MD. June 1990.
- Ash, A., Porell, F., Gruenberg, L., et al.: Adjusting Medicare Capitation Payments Using Prior Hospitalization. *Health Care Financing Review* 10(4):17-29, 1989.
- Ash, A., Porell, F., Gruenberg, L., et al.: *An Analysis of Alternative AAPCC Models Using Data from the Continuous Medicare History Sample*. Contract Number 99-C-98526/1. Prepared for the Health Care Financing Administration. Health Policy Research Consortium, Brandeis University and Boston University, September, 1986.
- Beebe, J., Lubitz, J., and Eggers, P.: Using Prior Utilization to Determine Payments for Medicare Enrollees in Health Maintenance Organizations. *Health Care Financing Review* 6(3): 27-38, Spring 1985.
- Beebe, J.C.: An Outlier Pool for Medicare HMO Payments. *Health Care Financing Review* 14(1):59-63, Fall 1992.
- Brown, R., and Langwell, K.: Enrollment Patterns in Medicare HMOs: Implications for Access to Care. In Scheffler, R.M., and Rossiter, L.S., eds.: *Advances in Health Economics and Health Services Research*. JAI Press. Greenwich, CT. 1987.
- Brown, R., Langwell, K., Berman, K., et al.: *Enrollment and Disenrollment in Medicare Competition Demonstration Plans: A Descriptive Analysis*. Contract Number 500-83-0047. Prepared for the Health Care Financing Administration. Mathematica Policy Research, Inc., Princeton, NJ. September 1986.
- Brown, R., Clement, D.C., Hill, J.W., et al.: Do Health Maintenance Organizations Work for Medicare? *Health Care Financing Review* 15(1):7-23, 1993.
- Brown, R., Bergeron, J.W., Clement, D.G., et al.: *Biased Selection in the Medicare Competition Demonstrations*. Contract Number 500-83-0047. Prepared for the Health Care Financing Administration. Mathematica Policy Research, Inc., Princeton, NJ. February 1988.
- Charlson, M.E., Pompei, P., Ales, K.L., and MacKensie, C.R.: A New Method of Classifying Prognostic Comorbidity in Longitudinal Studies: Development and Validation. *Journal of Chronic Disease* 40:373-383, 1987.
- Deyo, R., Cherkin, D., and Ciol, M.: Adapting a Clinical Comorbidity Index for Use With ICD-9-CM Administrative Databases. *Journal of Clinical Epidemiology* 45(6):613-619, 1992.
- Dun, D.L., Rosenblatt, A., Jaira, D., et al.: *A Comparative Analysis of Methods of Health Risk Assessment*. Final Report to the Society of Actuaries. Harvard University School of Public Health. October 12, 1995.
- Eggers, P., and Prihoda, R.: Pre-Enrollment Reimbursement Patterns of Medicare Beneficiaries Enrolled in 'At Risk' HMOs. *Health Care Financing Review* 4(1):55-73, September 1982.
- Elixhauser, A., Andrews, R., and Fox, S.: *Clinical Classifications for Health Policy Research: Discharge Statistics by Principal Diagnosis and Procedure*. Research Note 17, AHCPR Publication Number 93-0043. Public Health Service, Agency for Health Care Policy and Research, U.S. Department of Health and Human Services. Washington. U.S. Government Printing Office, August 1993.
- Ellis, R.P.: *Time Dependent DCG Model*. Cooperative Agreement Number 18-C-98526/1-03. Prepared for the Health Care Financing Administration. Health Policy Research Consortium, Boston University, June 1990.
- Ellis, R.P., and Ash, A.: Refinements to the Diagnostic Cost Group Model. *Inquiry*, 32: 1-12, Winter 1995.
- Ellis, R.P., and Ash, A.: *The Continuous Update Diagnostic Cost Group Model*. Cooperative Agreement Number 18-C-98526/1-03. Health Care Financing Administration. Health Policy Research Consortium, Boston University, June 1989.

- Ellis, R.P., and Ash, A.: *Refining the Diagnostic Cost Group Model: A Proposed Modification to the AAPCC for HMO Reimbursement*. Health Care Financing Administration. Health Policy Research Consortium, Boston University, February, 1988.
- Ellis, R.P., and McGuire, T.G.: Supply-Side Cost Sharing in Health Care. *Journal of Economic Perspectives* 7(4):135-151, Fall 1993.
- Ellis, R.P., and McGuire, T.G.: Insurance Principles and the Design of Prospective Payment Systems. *Journal of Health Economics* 7(3):215-238, September 1988.
- Ellis, R.P., and McGuire, T.G.: Provider Behavior under Prospective Payment. *Journal of Health Economics* 5(2):129-152, Summer 1986.
- Ellis, R.P., Pope, G.C., Iezzoni, L., et al.: *Diagnostic Cost Group (DCG) and Hierarchical Coexisting Conditions (HCC) Models for Medicare Risk Adjustment*. Contract Number 500-92-0020. Prepared for the Health Care Financing Administration by Health Economics Research, Inc., April 1996.
- Epstein, A.M., and Cumella, E.: Capitation Payment: Using Predictors of Medical Utilization to Adjust Rates. *Health Care Financing Review* 10(1): 51-70, Fall 1988.
- Fowles, J.B., Lawthers, A., and Weiner, J.: Agreement Between Physicians' Office Records and Medicare Part B Claims Data. *Health Care Financing Review* 16(4):189-200, Summer 1995.
- Hill, J.W., and Brown, R.S.: *Biased Selection in the TEFRA HMO/CMP Program*. Contract Number 500-88-0006. Prepared for the Health Care Financing Administration. Mathematica Policy Research, Inc., Washington, D.C., September 1990.
- Iezzoni, L.: *Risk Adjustment for Measuring Health Care Outcomes*. Health Administration Press, Ann Arbor, MI. 1994.
- Keeler, E., Kahn, K.L., Draper, D., et al.: Changes in Sickness at Admission Following the Introduction of the Prospective Payment System. *Journal of the American Medical Association* 264(15):1962-1968, October 17, 1990.
- Keeler, E.B., Carter, G.M., and Trude, S.: Insurance Aspects of DRG Outlier Payments. *Journal of Health Economics* 7(3):193-214, September 1988.
- Lubitz, J., and Prihoda, R.: Use and Cost of Medicare Services in the Last Years of Life. *Health Care Financing Review* 5(3): 117-131, Spring 1984.
- Newhouse, J.P.: Reimbursing Health Plans and Health Providers: Efficiency in Production Versus Selection. *Journal of Economic Literature* Winter 1995.
- Newhouse, J.P.: Rate Adjustors for Medicare Under Capitation. *Health Care Financing Review Annual Supplement*, 45-56, 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.
- Pope, G., Merrill, A.R., Gillani, F.S., and Ellis R.P.: *Internal Consistency of Diagnostic Coding on Medicare Provider Claims*. Contract Number 500-92-0020. Prepared for the Health Care Financing Administration by Health Economics Research, Inc., May 1994.
- Public Health Service and Health Care Financing Administration: *International Classification of Diseases, 9th Revision, Clinical Modification*. DHHS Publication Number 80-1260. Public Health Service. Washington. U.S. Government Printing Office, September 1980.
- Rositter, L.F., and Adamache, K.: *HMO Competitive Behavior and Risk-Based Payments Under Medicare*. Cooperative Agreement Number 18-C-98737/3-02. Prepared for the Health Care Financing Administration, February 1989.
- Schauffler, H., Howland, J., and Cobb, J.: Using Chronic Disease Risk Factors to Adjust Medicare Capitation Payments. *Health Care Financing Review* 14(1):79-90, Fall 1992.
- Thomas, J.W., and Lichtenstein, R.: Including Health Status in Medicare's Adjusted Average per Capita Cost Capitation Formula. *Medical Care* 24(3): 259-275, 1986.
- U. S. General Accounting Office: *Medicare: Changes in HMO Rate Setting Method Are Needed to Reduce Program Costs*. GAO/HEHS-94-119, September 1994.
- van Vliet, R., and van de Ven, W.: Toward a Budget Formula for Competing Health Insurers. Paper presented at the Second World Congress on Health Economics, Zurich. September 1990.
- van Vliet, R., and van de Ven, W.: Capitation Payments Based on Prior Hospitalizations. *Econometrics and Health Economics* 2:177-188, 1993.
- Weiner, J., Parente, S., Garnick, D., et al.: Variation in Office-Based Quality: A Claims-Based Profile of Care Provided to Medicare Patients With Diabetes. *Journal of the American Medical Association* 273(19):1503-1508, May 17, 1995.
- Weissman, J., Gatsonis, C., and Epstein, A.: Rates of Avoidable Hospitalization by Insurance in Massachusetts and Maryland. *Journal of the American Medical Association* 268(17):2388-2394, November 4, 1992.
- Welch, W.P.: Alternative Geographic Adjustments in Medicare Payment to Health Maintenance Organizations. *Health Care Financing Review* 13(3): 97-110, Spring 1992.

Whitmore, R., Paul, J., Gibbs, D., and Beebe, J.:
Using Health Indicators in Calculating the AAPCC.
*Advances in Health Economics and Health Services
Research* 10: 75-109, 1989.

Reprint Requests: Gregory C. Pope, M.S., Health Economics
Research, Inc., 300 Fifth Avenue, 6th Floor, Waltham,
Massachusetts 02154.