
Risk-Adjusted Medicare Capitation Rates Using Ambulatory and Inpatient Diagnoses

Jonathan P. Weiner, Dr.P.H., Allen Dobson, Ph.D., Stephanie L. Maxwell, Kevin Coleman, M.A., Barbara H. Starfield, M.D., M.P.H., and Gerard F. Anderson, Ph.D.

Researchers at The Johns Hopkins University (JHU) developed two new diagnosis-oriented methodologies for setting risk-adjusted capitation rates for managed care plans contracting with Medicare. These adjusters predict the future medical expenditures of aged Medicare enrollees based on demographic factors and diagnostic information. The models use the Ambulatory Care Group (ACG) algorithm to categorize ambulatory diagnoses. Two alternative approaches for categorizing inpatient diagnoses were used. Lewin-VHI, Inc. evaluated the models using data from 624,000 randomly selected aged Medicare beneficiaries. The models predict expenditures far better than the Adjusted Average per Capita Cost (AAPCC) payment method. It is possible that risk-adjusted capitation payments could encourage health plans to compete on the basis of efficiency and quality and not risk selection.

INTRODUCTION AND OVERVIEW

HCFA uses a demographic-based system, the AAPCC method, to reimburse health maintenance organizations (HMOs) and other managed care organizations (MCOs) that enroll Medicare beneficiaries. The limitations of the AAPCC model provide both a context and a rationale for the development of improved risk-adjust-

ment methodologies (Hornbrook, 1991; Beebe, Lubitz, and Eggers, 1985; Lubitz, Beebe, and Riley, 1985; Ellis, Pope, and Iezzoni, 1996).

If capitated payment systems were based on risk-adjuster models that incorporate diagnostic data, the functioning of markets for Medicare MCOs would likely improve. Accurate risk adjustment would benefit plans by more fairly paying them if they serve high-risk persons; it would benefit HCFA by reducing the possibility of overpayment to plans for low-risk beneficiaries who on average use fewer services. As the number of Medicare beneficiaries enrolled in managed care plans rises, so too does the need for improved capitation payment methodologies.

This article describes two new diagnosis-based risk-adjuster models that could be used to set Medicare capitation rates. The remainder of this section discusses "risk assessment" and "risk adjustment" and briefly reviews the current AAPCC method. The second section describes the two new diagnostic-based Medicare risk-adjuster models developed by researchers at The JHU School of Public Health. After the two models were developed, they were evaluated by a separate collaborating team at Lewin-VHI Inc., which assessed the risk adjusters' predictive accuracy, resistance to gaming, and administrative feasibility. Results of this evaluation are described in section three. The article concludes with a discussion of some strengths and weaknesses of the JHU Medicare risk-adjuster models.

The research in this article was supported in part by the Health Care Financing Administration (HCFA) under ORD Contract Number 93-026/EE. Jonathan P. Weiner, Stephanie L. Maxwell, Barbara H. Starfield, and Gerard F. Anderson are with The JHU School of Public Health. Allen Dobson and Kevin Coleman are with Lewin/VHI Inc. The views and opinions expressed are those of the authors and do not necessarily reflect the views of JHU, Lewin/VHI Inc., or HCFA.

Risk Assessment and Risk Adjustment

Differences in the use of medical services across individuals are in part predictable. The use of medical services depends on health status, which largely can be described by a patient's sociodemographic characteristics, clinical history and previous use of medical services (Hornbrook, 1991). "Risk assessment" methods use some combination of these data to classify individuals and thus to assess their expected use of medical services. "Risk adjustment" converts these risk assessment classes into insurance premium dollars for an insured group (e.g., per member per month capitation payments to a contracting HMO). While a risk-adjustment methodology is the foundation of a risk-adjusted capitation payment system, other technical and policy issues must be addressed during the design of the overall reimbursement system. These issues include, for example, the administrative feasibility of implementation, system monitoring and regulation, and approaches for updating or revising payments reflecting changes in medical inputs or technology.

The Current AAPCC

Medicare currently uses a demographic risk-adjustment method, the AAPCC system, to help establish capitated payments to MCOs for Medicare enrollees. The AAPCC uses age, sex, welfare status, and institutional (nursing home) status to create a series of mutually exclusive rate cells. These rate cells are designed to reflect the costs of providing care to Medicare enrollees treated in the fee-for-service (FFS) setting. HCFA sets capitated payments at 95 percent of Medicare expenditures predicted by the AAPCC, which is

adjusted for differences in local prices and practice patterns (Health Care Financing Administration, 1988).

The AAPCC has been criticized on technical and conceptual grounds. Technical criticisms center on the methods used to calculate payment amounts, as well as on the specific sociodemographic and geographic adjustments applied to the payments. Conceptual criticisms center on five concerns: (1) the payment system's poor ability to explain variations in individual medical expenditures, which results in plans having the latitude to select healthy enrollees or deselect sicker enrollees within each rating cell (Beebe, Lubitz, and Eggers, 1985); (2) the reliance on FFS costs to set payments for capitated settings; (3) the possible financial incentives (created by the institutional adjustment factor) to enroll "marginal" patients in institutional settings; (4) the possible financial incentives (created by the county-level geographic adjustment factor) to perpetuate inefficiencies in services delivery; and (5) the increasing divergence of capitated payments based on FFS costs relative to capitated plan costs in areas where capitated health plan market share is growing.

The aim of this project was to develop risk-adjuster models that would predict the medical expenditures of Medicare enrollees more accurately than the AAPCC payment system, and to more fairly pay capitated health plans for the mix of Medicare patients they attract and serve.

MODEL DEVELOPMENT

Conceptual Framework

Using diagnostic data to improve the precision of risk adjusters was a major focus of this project. As such, the objective was to integrate two diagnostic risk-assess-

ment/risk-adjustment systems previously developed by researchers at JHU. The first system is the Ambulatory Care Group (ACG) case-mix measure, designed primarily as an ambulatory case-mix measure for concurrent and retrospective use among the non-aged population (Weiner et al., 1991; Starfield et al., 1991).¹ The second original JHU risk-adjustment model was the Payment Amount for Capitated System (PACS), funded by HCFA and designed as a prospective, inpatient-oriented risk adjuster for the Medicare aged population (Anderson et al., 1989; Anderson et al., 1990). The use of the diagnostic-based ACG and PACS risk assessment technologies is a fundamental difference between the two new JHU Medicare risk-adjuster models and traditional demographic models such as the AAPCC.

Both JHU Medicare risk-adjuster models described in this article incorporate Ambulatory Diagnosis Groups (ADGs), which are the basic morbidity classification system of ACGs. This system groups *International Classification of Diseases, 9th Revision, Clinical Modification (ICD-9-CM)* diagnosis codes into 34 distinct ADGs, based on the codes' expected impact on medical service use and cost. Seven specific clinical and epidemiologic criteria were used to assign ICD-9-CM codes to ADGs.² A patient is assigned one or more (up to 34) ADGs based on the diagnoses documented by providers on their claims records during a predetermined (generally

1 year) time window.

Aspects of the new JHU risk-adjuster models are also based on the PACS model. The new models derive from PACS the use of three demographic risk assessors (age, sex, and prior disability status) and an inpatient measure based on the Major Diagnostic Category (MDC)³ associated with a patient's prior-year hospital admissions.

Sources of Data

This project made use of HCFA's Standard Analytical files (SAFs) created from the National Claims History file (NCHF). SAFs contain 100 percent of institutional bills and a 5-percent sample of all physician/supplier claims. Since 1991, the physician/supplier (Part B) claims include diagnosis information on almost all records. Data for the development and evaluation of the JHU models were drawn from five SAFs (inpatient, outpatient, physician/supplier, hospice, and home health) for the years 1991 and 1992. Sociodemographic information on the 5 percent of beneficiaries for whom data were retained in the physician/supplier SAF were obtained from the HCFA Hospital Insurance Skeleton Eligibility Write-Off (HISKEW) file. In total, data were obtained from a national 5 percent sample of approximately 1.5 million aged Medicare beneficiaries.

This project was one of the first to apply ambulatory diagnosis codes to the task of risk adjustment of Medicare beneficiaries. A major advantage of using this information to develop risk measures is that annually over 85 percent of beneficiaries receive

¹ In 1996, approximately 100 organizations (mainly HMOs) are using ACGs to manage, evaluate, and finance health care for their enrolled working-age populations. These uses include clinician profiling, withhold pool adjustment, and provider capitation payment.

² The specific criteria for the assignment of ICD-9-CM codes to ADG are: (1) likelihood of persistence or recurrence of the diagnosis; (2) likelihood of return visits and/or the need for continued treatment; (3) likelihood of the need for specialist services; (4) likelihood of decreased life expectancy; (5) likelihood of short-term or long-term patient disability; (6) expected need and cost of diagnostic and therapeutic procedures; and (7) likelihood of a required hospitalization.

³ MDCs group patients into one of 27 broad organ-system categories based on the patient's principal hospital discharge ICD-9-CM diagnosis. For example, all diseases of the nervous system are grouped into MDC No. 1. MDCs were developed as a base component of the diagnosis-related group (DRG) system.

ambulatory services and are diagnosed with one or more conditions in this setting. In contrast, less than 20 percent of beneficiaries annually receive one or more diagnoses in the inpatient setting.

Several beneficiary groups with incomplete data were excluded before creating the project's final data base. Excluded were beneficiaries who were: (1) under age 65 (i.e., non-aged disabled beneficiaries); (2) end stage renal disease (ESRD) beneficiaries; (3) enrolled in HMOs; (4) treated in Indian health service hospitals; (5) railroad Board retirees; (6) lacked Part B coverage; and (7) not continuously enrolled for 1991 and 1992 (those who died in 1991 were also excluded, but those who died in 1992 were retained in the study population). After these restrictions were applied, the final data base included 1.24 million aged individuals. The final data base was then split into two half-samples, a "development" half-sample and an "evaluation" half-sample.

Modifying ADGs and PACS for the Medicare Population

Before testing the applicability of the existing ADG classification system to the elderly, several enhancements to the ICD-to-ADG grouping algorithm were made.⁴ For example, based on an empirical analysis of actual diagnosis codes assigned in the inpatient and ambulatory settings to elderly patients and using assignment criteria described above, several hundred new ICD-9-CM codes were incorporated into the ADG system. In addition, a number of previous ICD-to-ADG mappings were modified.

An iterative process of model refinement and review was performed on the 34 ADGs

to develop the best sub-set of significant ($p = 0.05$), stable, and clinically acceptable ADGs that were predictive of future medical expenditures. We performed this iterative process on five random sub-populations constructed from the development-half of the data. Based on these activities, 13 of the 34 ADGs were included in the new JHU Medicare risk-adjuster models.⁵ This process resulted in the exclusion of ADGs that were relatively poor predictors (i.e., their coefficients were not significant across the data subsamples) of future medical costs in the elderly population. For example, ADG 2, which clusters diagnoses that are "time-limited minor, infections" was omitted from the models, whereas ADG 9, which clusters diagnoses that are "likely to recur, progressive" was included.

Based on a similar process of iterative model refinement and review of the original PACS risk-adjuster variables, new combinations and deletions of MDCs were ultimately incorporated into one of the final JHU models. For example, MDC 25 (HIV-AIDS) was added to the original PACS grouping of MDCs 16-17 (Blood, Immunological, and Myeloproliferative Diseases). In addition, five MDCs from the PACS model were excluded from the new JHU models. The first of the JHU Medicare models used this project's subset of MDCs to incorporate inpatient diagnoses, and the final subset of ADGs to capture ambulatory diagnoses.

For the second JHU model, we developed a new approach for incorporating inpatient diagnoses. Our motivation was to design a model that did not incorporate count variables and explicit prior-hospital admission variables, and thus would not reward a specific practice pattern. Thus we created a new risk measure, termed the

⁴ Readers are referred to the project's final report for the "map" of Medicare population ICD-9-CM code assignments to ADGs used in the JHU models (Weiner et al., 1996).

⁵ All 34 ADGs continue to be recommended for use in other applications, such as concurrent profiling; capitation adjustment for non-aged enrollees; and research.

“Hospital Dominant” (abbreviated as “Hosdom”) marker, which reflects the presence of diagnoses treated predominantly, but not always or necessarily, in the inpatient hospital setting.⁶

The Hosdom marker was developed through a multi-step empirical analysis of the 1.24 million beneficiaries in the combined development and evaluation files. For every ICD-9-CM diagnosis, we determined the likelihood that a beneficiary received at least some care during the year in either an inpatient or ambulatory setting for that condition. We then ranked the diagnosis codes based on the proportion of patients that had been hospitalized during the year for that diagnosis. Based on this list we ultimately defined the hospital dominant conditions to include 843 diagnoses for which at least 50 percent of patients had been hospitalized for that condition once or more during the year. The percentage of patients hospitalized for most of the marker’s diagnoses was much higher than the 50 percent level. (See Table 1 for examples of ICD-9-CM codes that are considered Hospital Dominant conditions).

RESULTS

Multiple Regression Modeling Technique

The JHU risk-adjuster models used the multiple regression method to identify the degree to which year-1 risk measures independently contribute to year-2 medical expenditures. This statistical method results in the calculation of a weight for each demographic and diagnostic risk characteristic used in the risk-adjustment model. The weights (regression coefficients) associated with each individual’s

risk measures were then summed to determine a unique risk-adjusted capitation rate for each person. In contrast to an actuarial cell-based rate setting method where only a few characteristics (for example, age and sex) can be used to develop an average expected rate for all persons, a multiple regression approach can calculate an individualized expected payment for each person based on several characteristics.

The risk measures (independent variables) included in the regression equations are presented on Table 2 for the first JHU model, termed the “ADG-MDC Model,” and Table 3 for the second JHU model, the “ADG-Hosdom Model.” The weights listed on the tables correspond to the year-2 expected medical expenditures associated with the presence of each independent variable.

Dependent Variable

The measure predicted by the JHU risk-adjuster models (the dependent variable) is “total annual Medicare expenditures.” These expenditures were defined as: (1) DRG payments with outlier and capital adjustments for inpatient hospital expenditures; (2) actual Medicare reimbursements for other institutional expenditures; and (3) resource-based relative value scale (RBRVS) payment estimates for physician/supplier services. In order to track current geographic-based payment differentials, HCFA’s geographic adjusters were applied to the RBRVS and DRG components of the dependent variable. Geographic Practice Cost Index (GPCI) weights were applied to RBRVS units, and the wage index was applied to DRGs.

To capture only the medical expenditures incurred by the Medicare program, expenditures were adjusted to exclude patient copayments and deductible amounts. Medical expenditures for which

⁶ Readers are referred to the final report for the list of Medicare population ICD-9-CM codes included in the Hosdom marker (Weiner et al., 1996).

Table 1
Example of Diagnoses in JHU Models' ADG and HOSDOM Variables

Variable		Example ICD-9-CM Diagnoses
Hospital Dominant (HOSDOM) marker	384.9	Other Septicemia Due to Gram-Negative Organisms
	157.1	Malignant Neoplasm of Body of Pancreas
	276.5	Volume Depletion Disorder
	410.01	Acute Myocardial Infarction, Anterolateral Wall, Initial Episode Care
	540.0	Acute Appendicitis With Generalized Peritonitis
Ambulatory Diagnostic Groups		
3 Time Limited: Major	361.0	Retinal Detachment With Retinal Defect
4 Time Limited: Major-Primary Infections	466.1	Acute Bronchiolitis
6 Asthma	493.0	Extrinsic Asthma
7 Likely to Recure: Discrete	531.9	Gastric Ulcer, Unspecified as Acute or Chronic
9 Likely to Recur: Progressive	250.10	Adult-Onset Type Diabetes Mellitus With Ketoacidosis
11 Chronic Medical: Unstable	424.1	Aortic Valve Disorders
16 Chronic Specialty: Unstable, Orthopedic	723.0	Spinal Stenosis in Cervical Region
22 Injuries/Adverse Effects: Major	820.8	Fracture of Unspecified Part of Neck of Femur, Closed
23 Psychosocial: Time Limited, Not Severe	309.01	Adjustment Reaction With Brief Depressive Reaction
25 Psychosocial: Recurrent or Persistent, Unstable	290.0	Senile Dementia, Uncomplicated
27 Signs/Symptoms: Uncertain	458.0	Orthostatic Hypertension
28 Signs/Symptoms: Major	429.3	Cardiomegaly
32 Malignancy	174.9	Malignant Neoplasm of Breast (Female)

NOTE: ADG is ambulatory diagnosis group morbidity classification method of the Ambulatory Care Group case-mix system. HOSDOM is diagnosis (843 total) that is usually (50 percent or more) treated in the inpatient setting.

SOURCE: Analyses on 1991-92 project data by authors during 1994-95.

Medicare was the secondary payer also were excluded. Finally, expenditures were adjusted to account for the partial year experience and expenditures of beneficiaries who died during year 2 of the project data.⁷ In the development data, the average annualized 1992 payment per beneficiary was \$4,266. The average actual (non-annualized) payment was \$3,214.

Independent Variables

The JHU models used several year 1 (1991) risk measures (or independent variables) to predict year 2 (1992) medical expenditures.

First, the "intercept" variable weight is equivalent to the total annual expected payment for a "baseline" enrollee—one whose characteristics do not trigger any of the models' other risk measurement variables. Given our structure of the variables, the baseline enrollee is a 65 year old female

who has not received any disability-based Medicare benefits in the past; was not eligible for Medicaid benefits during year 1; and had no health system encounters during year 1 where any ICD-9-CM codes were assigned to the diagnosis-based variables (described later) included in the JHU models. The total 1992 expected payments, or annual capitation amount, for a baseline enrollee using the ADG-MDC risk-adjuster model was \$608 (Table 2). Using the ADG-Hosdom model, payment for a base-line enrollee was \$434 (Table 3).

Both JHU models incorporate the same four sociodemographic variables: sex, age, prior disability status, and Medicaid eligibility. The weights corresponding to these risk measures slightly differ for each of the two models.

The first sociodemographic variable is "male." (Forty percent of beneficiaries were male in the development data). The weight corresponding to this independent variable indicates how much greater a capitated payment should be for males relative to females in 1992. This amount is \$604

⁷ Readers are referred to this project's final report for a description of the weighted adjustment method applied to expenditures of year 2 decedents, and for examples illustrating the effect of adjusting these expenditures (Weiner et al., 1996).

Table 2
JHU Medicare Capitation Adjustment Model Number 1: The ADG-MDC Model

Year 1 Variable	Year 2 Weight (SE)	Label	Population
	(Dollars)		(Percent)
Intercept	608 (28)	Base Expected Payment	
Male	604 (26)	Male/Female	39.2
Years Over 65	67 (2)	Number of Years Over Age 65	10.1(Mean)
Ever Disabled	1,119 (51)	Ever Received SI Disability	6.3
Medicaid	761 (43)	Currently Medicaid Eligible	9.6
MDC 1	1,533 (36)	Nervous System Inpatient Admission	1.9
MDC 3/4	3,237 (46)	Ears, Nose, Throat, Respiratory Systems	2.9
MDC 5	1,897 (79)	Circulatory System	5.5
MDC 6	1,759 (30)	Digestive System	2.7
MDC 7	1,030 (53)	Hepatobiliary System, Pancreas	0.8
MDC 8	1,117 (27)	Musculoskeletal, Connective Tissue	2.7
MDC 9	1,762 (77)	Skin, Subcutaneous Tissue and Breast	0.7
MDC 10	2,938 (43)	Endocrine, Nutritional, Metabolic Systems	0.8
MDC 11	2,526 (116)	Kidney, Urinary Tract	1.0
MDC 18	3,061 (79)	Infectious, Parasitic Diseases	0.5
MDC 19/20	1,957 (32)	Mental Disease, Alcohol, Drug Abuse	0.7
MDC 21	1,882 (29)	Injuries, Poisonings, Burns	0.2
MDC 23/24	1,481 (40)	Health Status Factors, Trauma	0.5
MDC 25/16/17	3,875 (79)	Blood, Immunological, Myeloproliferative Diseases, HIV, AIDS	0.4
MDC 26	3,944 (60)	Transplants	0.4
VADG 3	542 (36)	Time Limited, Major Diagnosis	15.7
VADG 4	734 (64)	Time Limited, Major, Primary Infections	8.4
VADG 6	818 (123)	Asthma	2.5
VADG 7	225 (65)	Likely to Recur, Discrete	23.6
VADG 9	965 (134)	Likely to Recur, Progressive	6.7
VADG 11	1,345 (126)	Chronic Medical, Unstable	42.0
VADG 16	650 (107)	Chronic Specialty, Unstable, Orthopedic	2.7
VADG 22	525 (177)	Injuries/Adverse Effects, Major	10.0
VADG 23	698 (110)	Psychiatric, Time Limited, Minor	1.2
VADG 25	804 (245)	Psychiatric, Persistent or Recurrent, Unstable	2.8
VADG 27	460 (163)	Signs/Symptoms, Uncertain	20.2
VADG 28	551 (97)	Signs/Symptoms, Major	30.7
VADG 32	1,347 (206)	Malignancy	11.1

NOTES: JHU is Johns Hopkins University. ADG is ambulatory diagnosis group morbidity classification method of the Ambulatory Care Group case-mix system. MDCs are major diagnostic categories (clusters of diagnosis-related groups). VADGs are "visit" ambulatory diagnostic group categories (of ACG system) derived from all available diagnoses on face-to-face ambulatory visit claims. (See Table 1 for examples of ICD-9-CMs grouped into each ADG.) Independent variables are derived from 1991 claims data of a sample of approximately 620,000 Medicare beneficiaries. SEs are standard errors of the coefficients. Population is the percent of patients flagged by each model variable. As the MDCs are count variables, their percentages reflect the percent of patients who had one or more admissions per MDC. However, of those patients hospitalized in 1991 within an MDC, an average of 92 percent were admitted only once in that MDC. The dependent variable is 1992 Medicare total expenditures.

SOURCE: Analyses on 1991-92 project data by authors during 1994-95.

based on the ADG-MDC model, and \$613 based on the ADG-Hosdom model. The second sociodemographic variable is age, defined by the number of years over age 65. The weight corresponding to this risk assessor indicates the payment amount allowed for each year over the age of 65. The third sociodemographic variable is labeled "ever disabled." This indicates past eligibility (i.e., before beneficiaries turned 65) for Social Security Disability Insurance. (When disabled Medicare eligibles reach 65 their status in the system changes to "aged"). The final sociodemographic vari-

able is year 1 Medicaid eligibility status. This variable is triggered if an individual is eligible for Medicaid benefits during at least 1 month in year 1 (1991).

The ADG-MDC model captures year 1 inpatient data using 15 selected MDC categories. (Twenty percent of the development-half beneficiaries were admitted at least once during 1991). The MDCs are "count" variables—the weight corresponding to each MDC variable is the expected year 2 expenditure associated with each hospital admission in that category during year 1. For example, if an enrollee is admit-

Table 3

JHU Medicare Capitation Adjustment Model Number 2: The ADG-HOSDOM Model

Year 1 Variable	Year 2 Weight (SE)	Label	Population
	(Dollars)		(Percent)
Intercept	434 (28)	Intercept	
Male	613 (26)	Male/Female	39.2
Years Over 65	64 (2)	Number of Years Over Age 65	10.1(Mean)
Ever Disabled	1,176 (52)	Ever Received SI Disability	6.3
Medicaid	802 (43)	Currently Medicaid Eligible	9.6
HOSDOM	1,749 (43)	Hospital Dominant Diagnosis	16.4
ALADG 3	663 (35)	Time Limited, Major Diagnosis	18.6
ALADG 4	1,503 (44)	Time Limited, Major, Primary Infections	9.9
ALADG 6	1,216 (76)	Asthma	2.7
ALADG 7	365 (30)	Likely to Recur, Discrete	25.3
ALADG 9	1,696 (49)	Likely to Recur, Progressive	8.3
ALADG 11	1,415 (27)	Chronic Medical, Unstable	44.2
ALADG 16	593 (74)	Chronic Specialty, Unstable, Orthopedic	2.9
ALADG 22	462 (40)	Injuries/Adverse Effects, Major	11.9
ALADG 23	1,222 (107)	Psychiatric, Time Limited, Minor	1.4
ALADG 25	1,088 (69)	Psychiatric, Persistent or Recurrent, Unstable	3.7
ALADG 27	568 (32)	Signs/Symptoms, Uncertain	21.5
ALADG 28	753 (30)	Signs/Symptoms, Major	33.5
ALADG 32	1,429 (40)	Malignancy	1.5

NOTES: JHU is Johns Hopkins University. ADG is ambulatory diagnosis group morbidity classification method of the Ambulatory Care Group case-mix system. HOSDOM is a "hospital dominant" diagnosis (presence of one or more diagnoses that usually are treated in the inpatient setting). ALADGs are "all" ADG categories (of ACG system) derived from all available ambulatory and inpatient diagnoses on face-to-face claims. (See Table 1 for examples of ICD-9-CMs grouped into HOSDOM and each ADG.) Independent variables are derived from 1991 claims data of a sample of approximately 620,000 Medicare beneficiaries. SEs are standard errors of the coefficients. Population is the percent of patients flagged by each model variable. The dependent variable is 1992 Medicare total expenditures.

SOURCE: Analyses on 1991-92 project data by authors during 1994-95.

ted to the hospital multiple times within the same MDC, one would multiply that MDC's weight by the number of year 1 admissions when calculating the individual's expected year 2 payments. The 1992 weights corresponding to the MDC variables of the ADG-MDC model range from roughly \$1,000 to \$4,000.

Both the ADG-MDC and ADG-Hosdom risk-adjuster models incorporate 13 selected ADG groupings. However, the claims source of diagnostic information in these 13 ADGs differs in the two JHU models. The ADG-MDC risk adjuster uses 13 "Visit ADGs" (VADGs). VADGs refer to ADGs that are assigned from diagnoses (either primary or secondary) noted by providers during "face-to-face"⁸ encounters in the ambulatory visit setting.⁹ (See Table 1 for examples of ICD-9-CM codes that fall into each of the ADG categories.) Seventy-one percent of the development data beneficiaries had one or more VADG variables.

Unlike the MDC "count" count variables, each VADG is a dummy variable (1=yes, 0=no) that can be triggered only once during the base year, regardless of the number of diagnoses an individual may have in each VADG. For example, VADG 3

⁸ "Face-to-face encounters" are defined as visits involving an evaluation and/or management service or a procedure performed by a physician (MD or DO) or a limited license professional (nurse practitioner, physician's assistant, dentist, podiatrist, social worker, chiropractor, or psychologist). A range of HCFA Common Procedure Coding System (HCPCS) procedure codes (which is an expansion of the CPT-4 system) were used in developing the JHU risk-adjuster models in order to identify and limit diagnoses to those during face-to-face encounters. The procedure code ranges are: (1) 00100-01999 for anesthesia; (2) 10160-69979 for surgery (excluding maternal care); (3) 77261-77799 for therapeutic radiology; (4) 78000-79999 for nuclear medicine; (5) 90701-99199 for medicine (includes 1991 evaluation and management codes); and (6) 99000-99499 for 1992 evaluation and management codes.

⁹ Diagnosis codes designated as "ambulatory visit" codes derive from all available line-items and header diagnoses from hospital outpatient facility claims, and all available line-item and header diagnoses (four maximum of each) from physician/supplier claims—that are associated with one or more of eight ambulatory-oriented places of service. These eight places, drawn from the HCFA provider data file, are the following: (1) office; (2) home; (3) outpatient hospital department; (4) hospital emergency room; (5) ambulatory surgical center; (6) State and local clinic; (7) outpatient rehabilitation clinic; and (8) intermediate care facility for the mentally retarded.

(which clusters diagnoses that are time limited, but major) is associated with an increase in year 2 individual capitation payments of \$542, regardless of the number of similar diagnoses or visits that a patient had during year 1. The 1992 weights corresponding to the VADG risk assessors of the ADG-MDC model range from roughly \$225 to \$1,350.

In contrast, the ADG-Hosdom risk-adjuster model uses 13 "All ADGs" (ALADGs.) ALADGs refer to ADGs that are assigned from all available (primary and secondary) diagnoses noted on inpatient and outpatient facility claims, as well as those noted by clinicians during face-to-face encounters in both the ambulatory and inpatient settings. The 1992 weights corresponding to the ALADG risk assessors of the ADG-Hosdom model range from roughly \$460 to \$1,700. Seventy-two percent of development data beneficiaries had one or more ALADG variables.

Finally, the ADG-Hosdom risk-adjuster model incorporates the new "Hospital Dominant" marker (in lieu of the prior admission-based MDCs). The Hosdom marker is a binary variable indicating the presence within an individual's claims records of one or more of 843 ICD-9-CM codes that are serious enough to usually be treated on an inpatient basis. If the marker is triggered, then a payment weight of \$1,749 is applied (only once) when summing scores to calculate an individual's annual capitation payment amount. The Hosdom amount is in addition to the weight of the ADG in which the Hosdom diagnosis may fall. About 16 percent of development data beneficiaries had one or more Hosdom diagnosis in 1991.

Table 4 summarizes and compares the percent of total variation in individual expenditures explained (using adjusted R-square statistics) by the two JHU risk-adjuster models and by a comparison

model similar to the AAPCC. For baseline comparison purposes throughout this project, JHU tested a multiple regression risk-adjuster model that approximates the components of HCFA's AAPCC payment system. (As described earlier, the AAPCC makes HMO-specific adjustments for the age, sex, welfare status, and nursing home residence status of risk contract enrollees). JHU's comparison model, hereinafter referred to as the "AAPCC" model, includes four sociodemographic components as constructed for JHU's two risk adjusters: sex, age, Medicaid eligibility status, and prior disability status. Medicaid and prior disability status were included as risk assessors in the comparison model to serve as rough proxies for the AAPCC welfare and nursing home residence status risk assessors. (Welfare and nursing home status were not available in the project's data.)

As seen on Table 4, the ADG-MDC model explains 6.3 percent of total variation at the individual level in year 2 medical expenditures based on year 1 variables. The ADG-Hosdom model explains 5.5 percent of total variation in individual year 2 medical expenditures based on year 1 variables. The "AAPCC" comparison model explains only 1.0 percent. When the data are truncated (i.e. medical expenditures above specific thresholds are capped at the thresholds), the explanatory power improves significantly. This indicates how reinsurance improves the explanatory power of risk adjusters, and how reinsurance thresholds may decrease the risk of very high cost (outlier) beneficiaries.

Table 4 also shows the adjusted R-square statistics when the three risk-adjuster models use year 1 variables to predict concurrent year 1 medical expenditures. Most capitation models incorporate risk variables in a predictive manner (i.e., using year 1 variables to predict year 2

Table 4
Percent of Variation in 1992 and 1991 Expenditures Explained by 1991 Models

Model (1991)	Medicare Expenditures					
	Not Truncated		Truncated at \$100,000		Truncated at \$50,000	
	1992	1991	1992	1991	1992	1991
ADG-MDC	6.3	64.4	8.0	66.3	9.0	69.2
ADG-Hosdom	5.5	40.9	7.0	42.3	8.0	45.6
"AAPCC"	1.0	1.2	1.3	1.3	1.6	1.4

NOTES: ADG-MDC is ambulatory diagnostic group-major diagnostic category. ADG-Hosdom is ambulatory diagnostic group-hospital dominant diagnosis. "AAPCC" is adjusted average per capita cost. HCFA's actual AAPCC system is approximated in this study by incorporating age, sex, Medicaid eligibility and prior disability status into a linear regression model. The percentages represent the adjusted R-square statistic of the individual level multivariate regression models as seen on Tables 1 and 2 for the approximately 620,000 beneficiaries in the development data base. Statistics are shown for the prospective model (1991 model predicting 1992 expenditures) and for a concurrent model (1991 model predicting 1991 expenditures).

SOURCE: Analyses on 1991-92 project data by authors during 1994-95.

medical expenditures). However some proposed risk-adjustment models are based partly on retrospective payments (payments made when services are actually delivered) that are linked to the presence of certain high-cost conditions. In addition, other applications of risk-adjustment models, such as provider profiling, use independent variables to explain resource use in the concurrent year.

Using JHU Models to Set Capitation Rates

As described, the two JHU risk-adjuster models use the multivariate regression method to assign a unique risk score to each Medicare enrollee. While the multivariate regression approach used to determine the weights associated with each risk measure is fairly complex, the process needed to calculate an individual enrollee's score and an overall group capitation rate involves straight-forward addition. Table 5 illustrates the arithmetic necessary to determine annual capitation rates for five hypothetical health plan enrollees. To illustrate the process, this table: (1) presents five beneficiaries with varying morbidity levels and health system encounters in year 1; (2) calculates the year 2 capitation rate of each patient based on his sociode-

mographic and diagnostic assignments; and (3) compares the capitation rates determined from the two JHU models and the baseline comparison "AAPCC" model. The ADG-MDC and ADG-Hosdom Tables 1 and 2 serve as "look-up" tables for identifying the payment weights associated with each risk assessor in the models and applying them as necessary to the enrollees presented on Table 5.

For example, "Enrollee 1" on Table 5 represents an 85 year old male who had no ambulatory visits or inpatient admissions during year 1 (as defined by the risk-adjuster models.) Table 5 sums the risk scores calculated on Table 2 that are associated with the relevant three characteristics (male, 20 years over age 65, and no health system encounters) of the individual. The ADG-MDC model predicts this individual's year 2 expenditures (equivalent to his adjusted capitation rate) to be \$2,552; the ADG-Hosdom model predicts this individual's year 2 expenditures to be \$2,327. The baseline comparison "AAPCC" model predicts this individual's year 2 expenditures to be \$4,785.

Two important features become apparent upon comparing the individual risk scores, or payment amounts, across risk adjusters of the five hypothetical enrollees on Table 5. First, regardless of the year 1

Table 5
Determining Year 2 Capitation Rates For Five
Health Plan Enrollees Using Year 1 Risk Measures

Enrollee	Risk-Adjustment Models		
	ADG-MDC	ADG-Hosdom	"AAPCC"
Enrollee 1 (No Health System Encounters)			
Male	\$604	\$613	\$732
85 Years (20 Years * Payment Weight)	1,340	1,280	2,160
Base Cost (Model Intercept)	<u>608</u>	<u>434</u>	<u>1,893</u>
Capitation Rate	2,552	2,327	4,785
Enrollee 2 (No MDCs, 3 ADGs)			
Male	604	613	732
85 Years	1,340	1,280	2,160
Base Cost (Model Intercept)	<u>608</u>	<u>434</u>	<u>1,893</u>
Depression (ADG 23)	698	1,222	
Gastric Ulcer (ADG 7)	225	365	
Coronary Atherosclerosis (ADG 11)	<u>1,345</u>	<u>1,415</u>	
Capitation Rate	4,820	5,329	4,785
Enrollee 3 (No MDCs, 3 ADGs, Hosdom Marker)			
Male	604	613	732
85 Years	1,340	1,280	2,160
Base Cost (Model Intercept)	<u>608</u>	<u>434</u>	<u>1,893</u>
Depression (ADG 23)	698	1,222	
Gastric Ulcer (ADG 7)	225	356	
Coronary Atherosclerosis (ADG 11)	1,345	1,415	
Hosdom Diagnosis Marker	<u>0</u>	<u>1,749</u>	
Capitation Rate	4,820	7,078	4,785
Enrollee 4 (2 MDCs, 6 ADGs, Hosdom Marker)			
Male	604	613	732
85 Years	1,340	1,280	2,160
Base Cost (Model Intercept)	<u>608</u>	<u>434</u>	<u>1,893</u>
Depression (ADG 23)	698	1,222	
Gastric Ulcer (ADG 7)	225	365	
Coronary Atherosclerosis (ADG 11)	1,345	1,415	
Corneal Edema (ADG 3)	542	663	
Diabetes (ADG 9)	965	1,696	
Heart Palpitations (ADG 27)	460	568	
2 Circulatory Admissions (MDC 5 x 2) or 1 Hosdom Diagnosis Marker	<u>3,794</u> <u>0</u>	<u>0</u> <u>1,749</u>	
Capitation Rate	10,581	10,005	4,785
Enrollee 5 (4 MDCs, 6 ADGs, Hosdom Marker)			
Male	604	613	732
85 Years	1,340	1,280	2,160
Base Cost (Intercept)	<u>608</u>	<u>434</u>	<u>1,893</u>
Depression (ADG 23)	698	1,222	
Gastric Ulcer (ADG 7)	225	365	
Coronary Atherosclerosis (ADG 11)	1,345	1,415	
Corneal Edema (ADG 3)	542	663	
Diabetes (ADG 9)	965	1,696	
Heart Palpitations (ADG 27)	460	568	
2 Circulatory Admissions (MDC 5 x 2) or 1 Hosdom Diagnosis Marker	<u>3,794</u> <u>0</u>	<u>0</u> <u>1,749</u>	
2 Respiratory Admissions (MDC 3 x 2)	<u>6,474</u>	<u>0</u>	
Capitation Rate	17,055	10,005	4,785

NOTES: ADG-MDC is ambulatory diagnostic group-major diagnostic category. ADG-Hosdom is ambulatory diagnostic group-hospital dominant diagnosis. AAPCC is adjusted average per capita cost. HCFA's actual AAPCC system is approximated in this study by incorporating age, sex, Medicaid eligibility, and prior disability status into a linear regression model. See Tables 2 and 3 for weights associated with each risk measure.

SOURCE: Analyses on 1991-92 project data by authors during 1994-95.

diagnostic history of the enrollees, the "AAPCC" system predicts the same year 2 adjusted capitation amount—\$4785. On an individual level, the "AAPCC" comparison model generally overpays health plans for healthy enrollees, and underpays health plans for less healthy enrollees. Second, the ADG-Hosdom risk-adjuster model generally results in higher year 2 capitation rates for enrollees who were not hospitalized in year 1 (enrollees two and three). For enrollees who were hospitalized (enrollees four and five), the ADG-MDC risk-adjuster model results in higher year 2 capitation rates.

INDEPENDENT EVALUATION METHODS

The JHU Medicare risk-adjuster models were evaluated by Lewin-VHI, Inc. along two dimensions. First, the ability of each model to predict the future medical expenditures of Medicare beneficiaries, particularly of non-random enrollee groups, was assessed. Second, the potential that either JHU risk-adjuster model could be "gamed" by health plans and providers, and the administrative feasibility of implementing each model, were considered.

The evaluation was conducted using the "evaluation" half (approximately 620,000 beneficiaries) of the data base JHU developed from the random sample of Medicare enrollees in 1991 and 1992. The payment amounts (i.e., the regression coefficients) of the models derived from the JHU development data were applied to the evaluation data in order to develop expected capitation rates for the population in the evaluation data.

Statistical Measures

Two statistical measures were used to evaluate the accuracy of the JHU models in

explaining and predicting year 2 expenditures based on year 1 data. The first measure, the adjusted *R*-square statistic, indicates the fraction of the total variance in year 2 (1992) medical expenditures at an individual level accounted for by a risk-adjuster model based on 1991 variables. The model with the higher adjusted *R*-square statistic for a given population explains the higher fraction of variance in the individual-level expenditures for that group, and is viewed as the best model according to these criteria.

The second empirical measure assessed the ability of each model to predict year 2 expenditures of groups of individuals. This measure, the predictive ratio, was calculated using the following equation:

$$\frac{\rho \text{Expected Expenditures}_i}{\rho \text{Actual Expenditures}_i}$$

for all members "i" of a beneficiary group. The predictive ratio for a particular group is the ratio of the sum of expected year 2 expenditures for all individuals in that group as predicted by the risk-adjuster models using year 1 diagnostic data, divided by the sum of actual year 2 expenditures of all individuals in that group. A predictive ratio of 1.00 indicates that a risk-adjuster model predicts the expenditures of a group perfectly. Predictive ratios less than 1.00 indicate that a model under-predicts the expenditures of the group; predictive ratios greater than 1.00 represent over-prediction.

The adjusted *R*-square statistics and predictive ratios of the JHU Medicare risk-adjuster models were compared with those of the comparison "AAPCC" model for several groups of Medicare enrollees. Comparisons with the latter model indicate the magnitude of the improved performance of the JHU models relative to current demographic payment systems.

Random and Non-Random Enrollee Groups

The first set of Medicare enrollee groups used for the empirical evaluation of the risk-adjuster models was of several differently-sized random enrollee groups. Three sets of 100 groups, each set consisting of 500, 5,000, and 50,000 individuals, were selected at random from the evaluation data base. Predictive ratios were calculated for each set of 100 groups; adjusted *R*-square statistics were not calculated for these groups.

The predictive ratios for the repeated random groups indicate the underlying risks of losses or gains that health plans would face under each risk-adjuster model, assuming they enroll beneficiaries purely at random. Capitated health plans do not enroll beneficiaries at random, but instead seek to have "positive" enrollment relative to their capitated payment system. Thus, it is possible that results from random groups may overstate the actual risk faced by capitated health plans.

The remaining sets of patient groups used to test the JHU risk-adjuster models were non-random groups of individuals. Evaluating risk adjusters using non-random groups provides information on the relative ability of the JHU (and comparison "AAPCC") risk-adjuster models to limit health plan selection bias. For example, one risk-adjuster model may better predict the expenditures of enrollees grouped by age and sex, but may more poorly predict the expenditures of other types of enrollee groups. Or, a risk-adjuster model may consistently over- or underestimate the medical expenditures of particular groups. It is important to note that these non-random groups represent extreme cases—i.e., they represent a health plan's enrollment profile if the plan enrolled only individuals with a specific medical condition or individuals

with uniformly low (or high) medical expenditures.

The non-random groups were constructed as follows:

- **Age-Sex Cells**—individuals of each sex were grouped into 5-year bands (based on their age in year 1): 65-69; 70-74; 75-79; 80-84; and 85+.
- **Medical Conditions**—individuals with one or more of 17 mainly chronic medical conditions, as found in the 1991 data, were grouped into: (1) depression; (2) alcohol and drug abuse; (3-4) hypertension; (5-6) diabetes; (7-9) cardiac conditions; (10) pulmonary conditions; (11-13) cancers; (14-15) stroke; (16) hip fracture; and (17) arthritis.
- **Expenditure Groups**—individuals were placed in one of five groups, based on their 1991 Medicare total expenditures. Quintile one includes the least expensive 20 percent of the evaluation group population.

EVALUATION RESULTS

Random Groups

Ideally, predictive ratios should cluster around 1.00, particularly for random groups. In addition, based on the Law of Large Numbers, distributions of predictive ratios should cluster more tightly around 1.00 as the size of a random group increases. The range and distribution of predictive ratios for the 100 random groups of three sizes are shown on Table 6.

The results on Table 6 support both hypotheses. First, the median predictive ratios for the three sizes of random groups are close to 1.00 for all three models (the JHU models and the "AAPCC" comparison model). The median predictive ratio ranges from 1.03-1.04 for groups of 500 enrollees; 0.98-0.99 for groups of 5,000

Table 6
Distribution of Predictive Ratios for Repeated Random Samples to Compare Three Risk-Adjustment Models

Model	5th Percentile	25th Percentile	Median	75th Percentile	95th Percentile
Group Size = 500 Enrollees					
ADG-MDC	0.8355	0.9092	1.0312	1.1047	1.2968
ADG-Hosdom	0.8304	0.9169	1.0370	1.0977	1.2879
"AAPCC"	0.8115	0.9073	1.0367	1.1140	1.3352
Group Size = 5,000 Enrollees					
ADG-MDC	0.8602	0.9297	0.9839	1.0593	1.1477
ADG-Hosdom	0.8683	0.9391	0.9892	1.0582	1.1444
"AAPCC"	0.8590	0.9175	0.9776	1.0521	1.1698
Group Size = 50,000 Enrollees					
ADG-MDC	0.9063	0.9344	0.9972	1.0545	1.1127
ADG-Hosdom	0.9129	0.9363	1.0002	1.0410	1.1139
"AAPCC"	0.8901	0.9040	1.0040	1.0461	1.1325

NOTES: ADG-MDC is ambulatory diagnostic group-major diagnostic category. ADG-Hosdom is ambulatory diagnostic group-hospital dominant diagnosis. AAPC is adjusted average per capita cost. See text for description of three risk-adjuster models. Results are based on 100 randomly selected groups of 500, 5,000, and 50,000 Medicare beneficiaries. Predictive ratios = expected expenditures / actual expenditures.

SOURCE: Analyses on 1991-92 project data by authors during 1994-95.

enrollees; and 1.00 for groups of 50,000 enrollees. Second, the distribution of predictive ratios for all three models clusters more tightly around 1.00 as the size of random groups increases.

The actual range of the predictive ratios, however, was relatively large for each of the models, even for random groups of 50,000 individuals. For example, the 5th and 95th percentile values on Table 6 show that 5 percent of random groups of 50,000 enrollees have predictive ratios of less than 0.90 or more than 1.10 for the JHU models and the comparison model. These results suggest that health plans with 50,000 random enrollees face a 5 percent likelihood, due solely to chance of incurring losses or gains of 10 percent or more if payments were based on the three models evaluated.

Reinsurance Simulations

The predictive ratios for the random groups discussed above were estimated using non-truncated expenditure data. By truncating 1992 medical expenditures at \$50,000, and re-estimating each risk adjuster model regression equation, one can simulate the effects of stop-loss rein-

surance.¹⁰ Reinsurance would provide plans with protection against the losses associated with patients who incur catastrophic expenditures. The presence (or absence) of an unusually large number of high-cost enrollees in the random groups of 50,000 could account for the Table 6 finding that 5 percent of these groups incur losses (or gains) of 10 percent or more.

The stop-loss reinsurance results on Table 7 reflect reinsurance set at 80 percent of individual expenses above the \$50,000 threshold (with health plans at-risk for the remaining 20 percent). The results show that reinsurance has a small effect on the range of predictive ratios for random groups of 50,000 individuals. For example, on Table 7 the range of predictive ratios between the 5th and 95th percentiles narrows only slightly when stop-loss reinsurance is introduced for the ADG-MDC model (from 0.91-1.11 to 0.92-1.09); for the ADG-Hosdom model (from 0.91-1.11 to 0.92-1.09); and for the "AAPCC" comparison model (from 0.89-1.13 to 0.90-1.11).

There are two possible explanations for why the reinsurance analysis indicates that

¹⁰ Readers are referred to the project's final report for a discussion and tables on reinsurance analyses for random groups smaller than 50,000 (Weiner et al., 1996).

Table 7
Distribution of Predictive Ratios for
Repeated Random Groups of 50,000 Individuals With Reinsurance

Stop/Loss Threshold	5th Percentile	25th Percentile	Median	75th Percentile	95th Percentile
ADG-MDC Model					
None	0.9063	0.9344	0.9972	1.0545	1.1127
\$50,000	0.9177	0.9357	0.9838	1.0350	1.1047
ADG-HOSDOM Model					
None	0.9144	0.9363	1.0002	1.0410	1.1139
\$50,000	0.9224	0.9369	0.9882	1.0235	1.0891
"AAPCC" Comparison Model					
None	0.8923	0.9040	1.0040	1.0461	1.1325
\$50,000	0.9017	0.9124	0.9867	1.0273	1.1061

NOTE: ADG-MDC is ambulatory diagnostic group-major diagnostic category. ADG-Hosdom is ambulatory diagnostic group-hospital dominant diagnosis. AAPCC is adjusted average per capita cost. See text for description of three risk-adjuster models. Results are based on 100 randomly selected groups of 50,000 Medicare beneficiaries. Predictive ratios = expected expenditures / actual expenditures. The reinsurance system tested included a 20-percent plan coinsurance rate over \$50,000, with Medicare being responsible for 80 percent.

SOURCE: Analyses on 1991-92 project data by authors during 1994-95.

stop-loss reinsurance does not significantly affect the predictive ratios of random groups of 50,000 individuals, even though reinsurance can be expected to improve the predictive ratios of smaller enrollee groups. First, the Law of Large Numbers predicts that as the size of a random group increases, the probability of that group (including an unusually high proportion of individuals with catastrophic medical expenditures) declines. Thus, stop-loss reinsurance did not affect the predictive ratios of the 100 random groups of 50,000 likely because none of these groups included a large enough proportion of individuals with medical expenses above the \$50,000 threshold.

A second reason reinsurance does not affect the predictive ratios of groups of 50,000 individuals may be due to the limited protection afforded health plans by reinsurance schemes. This point is best understood through a numeric example. Suppose an individual with \$80,000 of medical expenses is included in a random group. Next, suppose a risk adjuster, truncated at the stop-loss threshold of \$50,000, predicts the expenditures of this individual to be \$10,000. The total expenditures paid to a reinsured health plan for this individual equals the predicted \$10,000 plus

80 percent of the individual's expenses above \$50,000—0.80 (\$80,000-\$50,000) = \$24,000—or a total of \$34,000. Even with reinsurance, the plan thus receives payments of only 42.5 percent of that individual's medical expenses (\$34,000/\$80,000 = 42.5). This example indicates that while reinsurance does limit a plan's losses on its high cost enrollees, plans still incur significant losses for many enrollees with expenses above the stop-loss threshold. In turn, these losses imply that reinsurance may not significantly alter predictive ratios of large random groups of enrollees.

Non-Random Groups

Age and Sex Cohorts

Adjusted *R*-square statistics and predictive ratios for the first set of non-random groups, 5 year age-sex cohorts, are presented on Table 8. The adjusted *R*-square statistics from all of the age-sex cohorts groups are considerably higher for the two JHU models than for the "AAPCC" comparison model. These statistics range from 3.9-7.9 percent for the ADG-MDC model; 3.6-6.6 percent for the ADG-Hosdom model; and 0.1-1.2 percent for the compari-

Table 8
Adjusted R-Square Statistics and Predictive Ratios of
Three Risk-Adjustment Models for Age-Sex Groups

Group	ADG-MDC Model		ADG-HOSDOM Model		"AAPCC" Model	
	Adjusted R-Square	PR	Adjusted R-Square	PR	Adjusted R-Square	PR
	(Percent)		(Percent)		(Percent)	
Female, Age 65 to 69	7.89	1.0117	6.62	1.0150	1.22	1.0178
Female, Age 70 to 74	6.12	0.9937	5.37	0.9959	0.60	0.9832
Female, Age 75 to 79	5.94	0.9928	5.17	0.9933	0.33	0.9728
Female, Age 80 to 84	5.39	1.0042	4.73	1.0055	0.23	0.9926
Female, Age 85+	4.46	1.0266	3.68	1.0247	0.18	1.0673
Male, Age 65 to 69	5.80	1.0357	5.11	1.0338	0.55	1.0937
Male, Age 70 to 74	5.10	1.0023	4.40	1.0023	0.35	1.0120
Male, Age 75 to 79	4.66	0.9516	4.07	0.9554	0.22	0.9307
Male, Age 80 to 84	4.71	0.9821	4.35	0.9842	0.12	0.9455
Male, Age 85+	3.93	1.0099	3.60	1.0075	0.08	0.9986

NOTE: ADG-MDC is ambulatory diagnostic group-major diagnostic category. ADG-Hosdom is ambulatory diagnostic group-hospital dominant diagnosis. AAPCC is adjusted average per capita cost. PR is predictive ratio. Predictive ratios = expected expenditures/actual expenditures. Groups were defined by age in year 1 (1991) and sex. See text for description of three risk-adjuster models.

SOURCE: Analyses on 1991-92 project data by authors during 1994-95.

son "AAPCC" model. Thus, both JHU models account for several times more individual variation in year 2 medical expenditures than does the comparison "AAPCC" model for each age/sex cohort.

In contrast, the predictive ratios of the two JHU and "AAPCC" models for these age-sex cohorts are fairly similar. For the "AAPCC" model, predictive ratios range from 0.93-1.09—a range that is only slightly larger than that for the ADG-MDC model (0.95-1.04) and ADG-Hosdom model (0.96-1.02). The similarities between the predictive ratios of the "AAPCC" model and the JHU models for the age-sex groups are not surprising, given that each of the models include age and sex as independent variables.

The results on Table 8 also show that adjusted R-square statistics for the age-sex cohorts are highest for the ADG-MDC model. However, for seven of the 10 age-sex cohorts the ADG-Hosdom model has predictive ratios closest to 1.00.

Expenditure Quintiles

A common criticism of HCFA's AAPCC system for setting capitated Medicare pay-

ments is the ability of some health plans to "cream-skim," or engage in biased, positive enrollee selection. However, beneficiaries also self-select into health plans. Younger or healthier elderly, for example, may be more likely to enroll in Medicare MCOs than elderly with multiple or severe illnesses (Brown, 1994). For whatever reason, health plans that enroll beneficiaries who have not been heavy users of medical services may be rewarded financially by the AAPCC payment system, while plans enrolling a disproportionate share of heavy users will be penalized. A well-functioning risk-adjustment system should adjust for differences in the expected use of medical services by enrollees, thus encouraging plans to compete on the basis of price and quality, and not through risk selection.

The extent of the potential rewards for cream-skimming under the AAPCC are apparent in the predictive ratio results of enrollees grouped into medical expenditure quintiles. Expenditure quintiles were defined according to each individual's medical expenditures in year 1 (1991). The 20 percent of individuals with the lowest annual medical expenditures in year 1 formed the first expenditure quintile; the

Table 9
Adjusted R-Square Statistics and Predictive Ratios of
Three Risk-Adjustment Models for Expenditure Quintiles

Quintile (Average Cost)	ADG-MDC Model		ADG-HOSDOM Model		"AAPCC" Model	
	Adjusted R-Square	PR	Adjusted R-Square	PR	Adjusted R-Square	PR
	(Percent)		(Percent)		(Percent)	
First Quintile (\$1,415)	0.65	1.1913	0.66	1.0777	0.53	2.3417
Second Quintile (\$2,007)	1.00	1.1788	1.00	1.1736	0.73	1.6819
Third Quintile (\$2,807)	1.05	1.0693	1.07	1.1258	0.75	1.2379
Fourth Quintile (\$4,132)	1.30	0.9297	1.35	1.0019	0.60	0.8667
Fifth Quintile (\$7,569)	3.45	0.9212	2.60	0.8759	0.45	0.5014

NOTES: ADG-MDC is ambulatory diagnostic group-major diagnostic category. ADG-Hosdom is ambulatory diagnostic group-hospital dominant diagnosis. AAPCC is adjusted average per capita cost. PR is predictive ratio=expected expenditures / actual expenditures. Groups were defined by expenditures in year 1 (1991). See text for description of three risk-adjuster models.

SOURCE: Analyses on 1991-92 project data by authors during 1994-95.

20 percent with the highest medical expenditures in year 1 form the fifth expenditure quintile. Table 9 shows the mean year 1 (1991) expenditures of each quintile, and the adjusted R-square statistics and predictive ratios for each model by quintile.

As seen on the table, the comparison "AAPCC" model predictive ratio is 2.34 for the first (lowest) quintile. This indicates that the "AAPCC" model over-predicts the year 2 (1992) medical expenditures of the first quintile by 134 percent. Thus, health plans receiving capitated payments adjusted only for demographic risk measures would likely receive substantial profits if they enrolled individuals whose use of medical services in the past has been low. Conversely, the predictive ratio of 0.50 for the fifth (highest) quintile for the "AAPCC" model indicates that this model under-predicts the year 2 medical expenditures of the heaviest year 1 users of medical services by 50 percent. Plans that avoid enrolling or disenroll heavy users of medical services avoid substantial losses if their capitated payments are based only on demographic factors. Since the mean per person expenditure is much greater in the higher quintiles, the financial impact of predictive inaccuracies in the higher quintiles is much greater than it would be in the lower quintiles.

In contrast, the rewards for cream-skimming were much lower for JHU's ADG-MDC and ADG-Hosdom models. The ADG-MDC model over-predicts the medical expenditures of the first quintile by 19 percent and the ADG-Hosdom model by 8 percent.¹¹ Conversely, for the fifth quintile, the ADG-MDC model under-predicts medical expenditures by 8 percent and ADG-Hosdom model under-predicts medical expenditures by 12 percent. The rewards for cream-skimming are greatly reduced by the ADG-MDC and ADG-Hosdom models, which would encourage plans to compete more strongly through price and quality.

Diagnosis-Based Categories

The performance of the two JHU models and the comparison model was also assessed for 17 mainly chronic disease categories. Presumably capitated health plans can detect the presence of many of these diseases in current or prospective enrollees. If a risk-adjuster model tends to over-predict the medical expenditures of individuals with one or more of these con-

¹¹ A curious result is that the ADG-Hosdom model predictive ratios are "inverted," i.e., the predictive ratio for the first quintile (1.08) is nearer to 1.00 than that of the second quintile (1.17); however the predictive ratios for the third (1.13) and fourth (1.00) quintiles become closer to 1.00. We have no particular explanation for this intriguing statistical artifact.

ditions, then plans would have incentives to enroll these beneficiaries. Conversely, if a risk adjuster consistently under-predicts the expenditures of a disease category, then individuals with that condition could experience problems in access to capitated health plans.

As seen on Table 10, in 15 out of 17 cases both the ADG-MDC and ADG-Hosdom models do a far better job of predicting the year 2 medical expenditures of patient groups with these chronic conditions in year 1 than does the "AAPCC" model. In two disease groups, the second diabetes and second cancer groups, the "AAPCC" comparison model has predictive ratios closer to 1.00 than the JHU risk-adjuster models. Of the two JHU models, the ADG-MDC model has predictive ratios closer to 1.00 for 10 diagnosis groups; the ADG-Hosdom model has predictive ratios closer to 1.00 for five diagnosis groups. All three risk-adjuster models over- or under-predict the year 2 medical expenditures of some disease groups, such as the first hypertension group. These disease category results suggest areas of concentration for improving the discriminatory powers of the models within the limits of the ICD-9-CM coding system.

Other Findings

One set of analyses conducted in the model evaluation (but not included on the tables) compares the predictive ratios of the three risk-adjuster models for patient groups defined by the number of hospital admissions (none, one, two, and three or more) experienced by each individual in year 1 (1991). These analyses indicated an important difference between the two JHU models. Compared with the ADG-Hosdom model, the ADG-MDC model had predictive ratios much closer to 1.00 for individuals with two admissions in year 1 (1.01 ver-

sus 0.91 for the ADG-MDC and ADG-Hosdom models respectively); and for three or more admissions in year 1 (0.97 versus 0.66). The ADG-MDC model has a greater predictive power for these groups because the model includes MDC variables to capture year 1 hospital admissions.

Gaming and Administrative Feasibility

Gaming

One potential concern with either JHU risk-adjuster model is the possibility for "upcoding" by health plans and providers. Upcoding occurs when plans engage in strategic behaviors, including recording additional diagnoses or reclassifying diagnoses, designed to increase risk-adjusted capitated payments to the plans. The main advantage of a demographic risk-adjuster model such as HCFA's AAPCC payment system is its resistance to this type of gaming—health plans cannot change the age and sex, nor likely the welfare and nursing home residence status of their enrollees. However diagnosis-based risk adjusters such as the ADG-MDC and ADG-Hosdom models are susceptible to gaming through excessive coding, upcoding, and reclassifying of diagnoses. This concern may be particularly relevant for the ADG and Hosdom variables, in which a single ambulatory code in year 1 results in higher payments in year 2.

In general, however, several factors limit the ability of health plans to engage in code gaming under either the ADG-MDC or ADG-Hosdom models. First, health plans would need time to identify the best options and model variables for gaming, and yet more time to acquire or purchase the data collection and manipulation skills necessary for successful upcoding. Second, HCFA could adopt auditing and enforcement procedures designed to iden-

Table 10
Adjusted R-Square Statistics and Predictive Ratios of
Three Risk-Adjustment Models for 17 Conditions

Condition	ADG-MDC Model		ADG-HOSDOM Model		"AAPCC" Model	
	Adjusted R-Square	PR	Adjusted R-Square	PR	Adjusted R-Square	PR
Depression	0.45	0.9921	4.69	1.0215	0.76	0.9437
Alcohol and Drug Abuse	¹ -6.12	1.1128	¹ -3.20	1.2096	¹ -0.13	0.7918
Hypertensive Heart/Renal Disease	3.54	1.1664	2.57	1.2091	0.77	1.1712
Benign/Unspecified Hypertension	2.51	1.0564	2.23	1.0643	0.62	1.3546
Diabetes with Complications	3.83	1.0301	2.65	1.0591	0.96	0.8854
Diabetes without Complications	3.44	0.8528	3.03	0.8621	0.63	0.9260
Heart Failure/Cardiomyopathy	3.80	0.8965	3.16	0.8810	0.17	0.7133
Acute Myocardial Infarction	1.92	0.8827	1.98	1.0071	0.09	0.6335
Other Heart Disease	3.37	1.0353	2.78	1.0354	0.53	0.7873
Chronic Obstructive Pulmonary Disease	5.83	0.9415	4.59	0.9238	0.86	0.6834
Colorectal Cancer	5.42	0.8734	3.95	0.8981	0.30	0.5383
Breast Cancer	6.22	1.4189	5.16	1.4223	0.54	0.9270
Lung/Pancreas Cancer	4.93	0.7150	3.97	0.6589	3.10	0.3360
Other Stroke	4.91	0.9355	4.19	0.9911	0.49	0.5638
Intracerebral Hemorrhage	¹ -1.54	0.8111	¹ -0.11	0.9203	¹ -0.53	0.4415
Hip Fracture	3.63	0.9704	2.68	1.0531	0.16	0.6525
Arthritis	5.15	0.9572	4.57	0.9773	0.80	0.8151

¹It is possible for the adjusted R-square statistic to be negative.

NOTES: ADG-MDC is ambulatory diagnostic group-major diagnostic category. ADG-Hosdom is ambulatory diagnostic group-hospital dominant diagnosis. AAPCC is adjusted average per capita cost. PR is predictive ratio. Predictive ratios = expected expenditures/actual expenditures. Groups were defined by ICD-9-CM codes noted in year 1 (1991) claims data. See text for description of three risk-adjuster models.

SOURCE: Analyses on 1991-92 project data by authors during 1994-95.

tify the most obvious examples of code gaming. For example, dramatic increases in the percentage of a health plan's enrollees with a Hosdom diagnosis, or in the percentage of a plan's enrollees with relatively expensive ADG diagnoses, may indicate to a monitoring agent upcoding activities by the health plan. In addition, if most health plans engaged in some level of code creep or excessive code documentation, HCFA could make adjustments to the payment system. For example, specific payment amount could be rebased, and overall payments could be lowered through a conversion factor. Finally, it is recommended that any new risk-adjustment system be used to allocate a predetermined dollar amount across participating plans, and not to determine the amount of overall funds allocated to the Medicare budget. As such, gaming across health plans would be budget neutral.

Another gaming concern is whether health plans can exploit informational

advantages to risk select against a risk-adjusted capitation payment system. Some argue that health plans have better access to expenditure and encounter data to predict future medical expenditures than would some risk-adjuster systems. If so, plans may be able to identify the best risks relative to the risk-adjuster model used for capitated-payment model—i.e., identify individuals that a health plan predicts will have lower medical expenditures than does the risk-adjuster payment model. Plans that can identify and enroll these better risks can thus cream-skin the risk-adjuster payment model. The results of selection in a risk-adjusted environment could differ markedly from that of the current environment. It may indeed be profitable to enroll people with poor health status if the system does not underpay, or even overpays, for such people.

It is not known how many health plans have the data capabilities and knowledge required to adopt such strategies to game

diagnosis-based risk-adjuster models such as the ADG-MDC or ADG-Hosdom models. Whether this is possible, it does appear that the JHU models provide much less scope for plans to cream-skim than do less powerful demographic risk-adjuster models such as HCFA's AAPCC payment system. In addition, as experience with more comprehensive methods of risk adjustment is gained, capitated payment systems will become more sophisticated and will develop better defenses to cream-skimming.

Finally, one concern regarding the ADG-MDC model is whether its use of prior-admission based variables would encourage inappropriate hospital admissions, or multiple admissions instead of a single hospitalization. However, the model's year 2 payments triggered by year 1 admissions are much less than the costs of actual year 1 admissions. In addition, it is possible that health plans would lose the year 2 membership of enrollees with year 1 admissions through disenrollment or death, and thus lose all year 2 payments for such enrollees. In theory, however, cases could exist where the expected increase in year 2 payments exceed the difference in year 1 costs of treating some beneficiaries in ambulatory versus inpatient settings. As such, this would provide incentives under the ADG-MDC model for health plans to increase admissions.

Administrative Feasibility

Several administrative issues would need to be addressed in order to develop a comprehensive, risk-adjusted capitated payment system using either of the JHU (or any other new) risk-adjuster models. First, although all of the data required by the two JHU models are available in most existing health plan claims, encounter and

enrollment data bases, a significant minority of health plans do not yet collect the diagnostic data required to assign ADGs and MDCs. Second, a method to annually update payment weights would be needed in order to account for inflation, advances in the state of care, and the aging of the population. Two options for updating the JHU models' weights are to employ the U.S. per capita cost (USPCC) for aged Medicare enrollees, and to rebase, or reestimate, the models' regression equations on more recent data. Reestimating the equations also would allow for modifications of the ADG, MDC and Hosdom grouping algorithms.

Third, a 2- to 3-year lag could exist between when a full year of claims data would be available for assigning ADGs and MDCs and the year in which capitated payments would be made to plans. Given this lag period, payment rates would need to be updated with, for example, an inflation factor. Fourth, a risk adjuster would also have to address the lack of prior data of enrollees who "age-in" to the Medicare program throughout each year, and of other partial-year enrollees. For example, HCFA could make interim payments for individuals who age-in based on the AAPCC, until sufficient diagnostic data are available to assign payments based on the full models. In addition, policy decisions regarding issues such as geographic adjustments to payments, stop-loss reinsurance, or high-cost disease carve-out mechanisms would require modifications of risk-adjuster models. Finally, a method would be needed to phase-in any new, more sophisticated risk-adjusted payment system, particularly so that health plans more familiar with actuarial rate-cell systems would have time to gear-up for the individualized scoring approach of the ADG-MDC and ADG-Hosdom models. Demonstra-

tions conducted by HCFA would give an indication of the degree to which health plans have difficulty adapting to a new payment system.

CONCLUSION

Diagnoses-based risk-adjuster models have considerable promise to improve current methods of calculating risk adjusted, capitated premiums for HMOs and other MCOs. One of the most important strengths of JHU's ADG-MDC and ADG-Hosdom models is their clinical foundation. The diagnosis-based variables (ADGs, MDCs and the Hosdom marker) incorporated into the JHU Medicare risk-adjuster models are based on epidemiology and the natural history of disease. In addition, the ADGs and MDCs used in the JHU models are already widely used and accepted by clinicians, HMOs, and researchers.

The ADG-MDC and ADG-Hosdom risk-adjuster models are regression-based models that are more complex conceptually than traditional rate-cell risk models. Despite the added complexity, calculating risk scores and capitated premiums using either JHU risk-adjuster model is a matter of simple arithmetic. In addition, the diagnostic data required by the ADG-MDC and ADG-Hosdom models are now being collected by HCFA and most MCOs.

By incorporating diagnostic information, JHU's ADG-MDC and ADG-Hosdom models are better able to predict future medical expenditures than demographic risk-adjuster models such as the AAPCC, for both randomly selected and non-randomly selected individuals and groups. In particular, however, the JHU ADG-MDC and ADG-Hosdom models better predicted the expenditures of non-random groups.

One of the most important findings of the evaluation was the ability of both JHU

models to predict the year 2 medical expenditures of groups that either used few medical services or a great deal of medical services in year 1 (the first and fifth expenditure quintiles). In contrast, the "AAPCC" model grossly over-predicted the year 2 medical expenditures of the first quintile group and under-predicted the year 2 medical expenditures of the fifth quintile group. This finding indicates that the rewards of cream-skimming plans are greatly reduced under either JHU Medicare risk-adjuster model relative to the AAPCC.

The incorporation of diagnostic information, however, comes at some cost. It is possible that plans would engage in code-creep activities. Health plans could recode diagnoses or code additional diagnoses to increase their enrollee's risk scores, and thus their premiums under either the ADG-MDC or ADG-Hosdom models. HCFA could respond to code-creep by implementing auditing and enforcement activities or by reducing the conversion factor used to set premium payments to plans.

Another possible limitation of the ADG-MDC and ADG-Hosdom models is that they do not explicitly address or exclude "discretionary" hospital admissions or diagnoses. While some conditions may be more discretionary than others, JHU clinicians could not identify any ICD-9-CM codes for which hospital admissions are always discretionary. In addition, claims data do not currently contain the detailed clinical information required to determine whether a patient's diagnosis or hospital admission is discretionary.

To limit discretionary hospital admissions, the ADG-Hosdom model uses the Hosdom marker variable, thus avoiding increases in capitation payment directly related to hospital admissions. This is perhaps its main advantage over our second

model. On the other hand, the ADG-MDC model does increase second-year capitation payments as the result of explicit hospital admissions during the prior year. This could induce some plans to admit beneficiaries to hospitals who otherwise might be treated in the ambulatory setting. However, the relatively low year 2 payments that plans would receive for a year 1 hospital admission under the ADG-MDC model, as well as the chance that beneficiaries admitted might not be enrolled in year 2, limits the incentives for plans to increase discretionary hospital admissions. The main advantage of the ADG-MDC model is that for some very high use patient groups, this model's predictive power may be somewhat higher than that of the ADG-Hosdom model.

While the ADG-MDC and ADG-Hosdom models build on over a decade of research, there are areas where additional research could further improve both models. First, the Hosdom and ADG variables could be modified to require more than a single code to trigger the variables. This could reduce the susceptibility of these variables to upcoding by plans or the inaccuracy associated with temporary diagnoses such as "rule-out" codes. Second, either JHU model could be enhanced by incorporating reinsurance and/or diagnostic-based carve-outs of high cost cases. JHU and Lewin-VHI currently are working on a HCFA-sponsored project to develop risk-adjuster models that include reinsurance and diagnosis-specific high-cost carve-outs for the under-65 population (Lewin-VHI and Johns Hopkins University, 1996).

Any new risk-adjuster model will require multiple demonstrations before it can be used to establish risk-adjusted capitated payments to MCOs that enroll Medicare beneficiaries. In particular, HCFA's planned demonstrations in this area (Vladeck, 1995) offer an excellent opportu-

nity to test and evaluate the JHU Medicare risk-adjuster models.

In conclusion, there is reason for some optimism that risk-adjusted payments can be made powerful enough to help create a more level playing field for capitated MCOs to compete for Medicare enrollees. In such an environment, competition based on premium price and quality, rather than on selection of "good risks," would be encouraged.

ACKNOWLEDGMENTS

The authors gratefully acknowledge Yifei Hu and Andrew Baker for their tireless data base development efforts and model programming, and Chad Abrams for his ACG systems work. The substantive input of Melvin J. Ingber, Ph.D., our HCFA project officer, is gratefully acknowledged. The constructive criticism of three anonymous reviewers is also acknowledged.

REFERENCES

- Anderson, G., Lupu, D., Powe, N., et al.: Payment Amounts for Capitated Systems. Report prepared for the Health Care Financing Administration under Contract Number 17-C-98990/3. Baltimore, MD. Johns Hopkins University, December 1989.
- Anderson, G., Steinberg, E.P., Powe, N.R., et al.: Setting Payment Rates for Capitated Systems: A Comparison of Various Alternatives. *Inquiry* 27:225-233, Fall 1990.
- Beebe, J., Lubitz, J., and Eggers, P.: Using Prior Utilization Information to Determine Payments for Medicare Enrollees in HMOs. *Health Care Financing Review* 6(3):27-38, 1985.
- Brown, R., in Luft, H. (ed.): *HMOs and the Elderly*. Ann Arbor, MI. Health Administration Press, 1994.
- Ellis, R.P., Pope, G.C., Iezzoni, L.I., et al.: Diagnosis-Based Risk Adjustment for Medicare Capitation Payments. *Health Care Financing Review* 17(3): XX-XX, Spring 1996.
- Health Care Financing Administration: Study and Recommendations to Congress on Ways to Refine the Adjusted Average Per Capita Cost (AAPCC) and the Adjusted Community Rate (ACR). Washington, DC, November 1988.

Hornbrook, M. (ed.): Risk-Based Contributions to Private Health Insurance. *Advances in Health Economics and Health Services Research* 12, 1991.

Lewin-VHI, Inc., and The Johns Hopkins University: Development of a Risk-Adjustment System Under Health Reform. Final report of Year 1, Contract Number 500-92-0021. April, 1996.

Lubitz, J., Beebe, J., and Riley, G.: Improving the Medicare HMO Payment Formula to Deal with Biased Selection. *Advances in Health Economics and Health Services Research* 6:101-122, 1985.

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

Vladeck, B.: The Medicare Choice Initiative. Proposed Solicitation. Baltimore, MD. Health Care Financing Administration, 1995.

Weiner, J.P., Dobson, A., Maxwell, S.L., et al.: *The Development and Testing of Risk Adjusters Using Medicare Inpatient and Ambulatory Data*. Final report prepared for the Health Care Financing Administration under ORD Contract Number 93-026/EE. Baltimore, MD. The Johns Hopkins School of Public Health, 1996.

Weiner, J.P., Starfield, B., Steinwachs, D., and Mumford, L.: Development and Application of a Population Oriented Measure of Ambulatory Care Case-Mix. *Medical Care* 29:452-472, 1991.

Weiner, J.P., Powe, N., Steinwachs, D., and Dent, G.: Applying Insurance Claims Data to Assess Quality of Care: A Compilation of Potential Indicators. *Quality Review Bulletin* 16(12):424-438, 1990.

Weiner, J.P.: Application of ACGs to Risk Adjustment. Paper presented at HCFA-ORD Conference on Risk Adjustment and Health Policy Reform. Baltimore, MD. September 1993.

Reprint Requests: Jonathan P. Weiner, Dr.P.H., Professor of Health Policy and Management, The Johns Hopkins School of Public Health, 624 N. Broadway, Room 605, Baltimore, Maryland 21205.