

Adjusting Medicare capitation payments using prior hospitalization data

by Arlene Ash, Frank Porell, Leonard Gruenberg, Eric Sawitz, and Alexa Beiser

The diagnostic cost group approach to a reimbursement model for health maintenance organizations is presented. Diagnostic information about previous hospitalizations is used to create empirically determined risk groups, using only diagnoses involving little or no discretion in the decision to hospitalize. Diagnostic cost group and other models (including Medicare's current formula

and other prior-use models) are tested for their ability to predict future costs, using R² values and new measures of predictive performance. The diagnostic cost group models perform relatively well with respect to a range of criteria, including administrative feasibility, resistance to provider manipulation, and statistical accuracy.

Introduction

In 1982, the Tax Equity and Fiscal Responsibility Act permitted risk contracts with full prospective payment to health maintenance organizations (HMO's) for enrollment of Medicare beneficiaries. Such contracts are intended to be priced on the basis of 95 percent of what HMO enrollees would have cost Medicare had they remained in the fee-for-service (FFS) sector. This requires a methodology for estimating the hypothetical cost to Medicare. The current HMO payment model, known as the adjusted average per capita cost (AAPCC), starts with projected Medicare FFS reimbursements per capita in the counties of residence of HMO enrollees as the basis of payment. The average cost projections are then adjusted for differences in the distribution of HMO enrollees by age, sex, welfare status, and institutional status relative to the distribution of beneficiaries in the same geographic area who receive their care in the FFS sector.

Although individuals exhibit a great deal of variation in yearly medical costs, purely random variability will tend to average out in the aggregate even in a moderate-sized HMO. However, a major shortcoming of the existing AAPCC formula is its inability to adequately adjust capitation levels for systematic differences in the health status of enrolled groups (Trapnell, McKusick, and Genuardi, 1982; Thomas, Lichtenstein, and Wyszewianski, 1986; Gruenberg, Wallack, and Tompkins, 1986; Newhouse, 1986). To the extent that the AAPCC underestimates the expected costs for certain individuals (such as the chronically ill), HMO's with disproportionately many such enrollees are unfairly penalized through adverse selection. Conversely, an HMO with favorable selection is the beneficiary of undeserved profits.

In several studies, it has been shown that substantial selection bias does occur in Medicare HMO's (Eggers and Prihoda, 1982; Brown, 1988; U.S. General Accounting Office, 1986). However, healthy people can change only by getting sicker, and the heaviest health utilizers die soonest. Therefore, the health status of an HMO's continuing enrolled population will tend, over time, to regress toward the mean. Although this tendency may lessen the urgency for a health status adjuster in the Health Care Financing Administration (HCFA) payment formulas (Welch, 1985), the effects of continuing selection bias among new recruits and of nonrandom disenrollment patterns can still lead to a perpetual situation of enrollment bias.

Although the Government is worried about HMO's attracting healthier enrollees and discouraging the sicker ones through targeted marketing and selective program development, some HMO's remain convinced that they experience adverse selection. HMO's may also be concerned by the finding that more recent enrollees tend to be sicker than their continuing HMO enrollee counterparts (Halvorson and Stix, 1988). However, as long as pricing is fair, enrollment bias is not a problem; many HMO's would welcome the opportunity to enroll sicker patients without fear of economic disaster. Thus, it is important to be able to modify the current AAPCC classifications to better reflect differences in health status risks among Medicare beneficiaries (Trapnell, McKusick, and Genuardi, 1982; Lubitz, Beebe, and Riley, 1985; Gruenberg, Wallack, and Tompkins, 1986).

In the work reported here, we studied the future cost implications of current hospital utilization, distinguishing among different reasons for prior hospitalizations through principal diagnoses, and developed a diagnostic cost classification that could serve as the basis for health status rate adjusters in an HMO payment formula.

Incorporating health status adjustments into a capitation payment formula through diagnostic information from hospitalizations has the potential to produce a payment model that is easier to administer, more sensitive, and less easily manipulated than models based on volume of previously used services or

This research was supported by Grant No. 99-C-98326/1 from the Health Care Financing Administration. The views and opinions expressed are the authors', and no endorsement by the Health Care Financing Administration or the Department of Health and Human Services is intended or should be inferred.

Reprint requests: Arlene Ash, Health Care Research Unit, Boston University Medical School, 720 Harrison Avenue, Suite 1102, Boston, Massachusetts 02118.

costs. In this article, we compare the predictive performance of payment models developed using diagnostic hospitalization information with several previously proposed prior-use models. The use of common data (one file for estimating each model and a second for testing) and a wide array of measures of predictive accuracy provides for direct and comprehensive comparisons of the predictive performance of a variety of alternative models.

Prior use and diagnostic refinements

The superior predictive performance of prior-use models relative to the AAPCC risk classifications is well established (McCall and Wai, 1983; Anderson and Knickman, 1984; Beebe, Lubitz, and Eggers, 1985; Anderson et al., 1982; Anderson, Resnick, and Gertman, 1983; Thomas and Lichtenstein, 1986; Lubitz, Beebe, and Riley, 1985). By paying more for sicker enrollees than for healthy ones, HCFA can defuse the debate about the amount and direction of biased selection and can encourage the enrollment of Medicare's most costly beneficiaries into the apparently more efficient setting of managed care. How should we choose among prior-use models?

Thomas et al. (1983) proposed the following criteria for alternative payment models: predictive accuracy, invulnerability to provider manipulation, and administrative simplicity. These are useful for making comparisons. However, a lack of agreement exists about their relative importance, and even for a single criterion it may be unclear how to achieve a definitive ranking.

A prior-use model developed by Beebe, Lubitz, and Eggers (1985) was tested and evaluated in an HMO demonstration project at Senior Health Plan in Minnesota. In addition to age, sex, and welfare status, the model uses an individual's number of inpatient hospital days and an indication as to whether Part B deductibles were met in the 2 years prior to HMO enrollment. This model yields more accurate predictions of future costs than the current AAPCC does. Further, it is simpler to administer than many other prior-use models because it relies only on data that are readily available for all Medicare beneficiaries in the FFS system. However, the model can be used to set payments only for individuals with such Medicare FFS experience just prior to their HMO enrollment. In particular, data would not be available for newly eligible Medicare beneficiaries (65-year-olds) and beneficiaries previously enrolled in an HMO. Also, the prior-use information required by the formula (namely, the number of days hospitalized and the meeting of Part B deductibles) are coarse indicators of ongoing health needs; no distinctions are made between self-limiting and continuing illnesses, nor are hospitalizations that might be avoided by an efficient provider distinguished from those that any provider would feel to be necessary.

A problem with any model that bases its payments on a person's utilization in the Medicare FFS sector

prior to HMO enrollment is the issue of updating. Over time, the health status of individuals will change, and the model will become less accurate as the prior experience becomes more remote. Updating risk classification on the basis of utilization within the HMO, however, raises questions. One question concerns the nature of the data that it is appropriate to ask HMO's to collect. Another concerns perverse incentives in the payment model that would pay more to an HMO that is a higher utilizer, irrespective of its enrollees' actual level of need.

The use of condition-specific prior-use information to classify enrollees with higher risks of significant future medical expenses was proposed quite early as a potential refinement to the AAPCC (Trapnell, McKusick, and Genuardi, 1982). In the works of Anderson et al. (1982) and Anderson, Resnick, and Gertman (1983), it was demonstrated that certain types of hospitalizations, irrespective of their current costs, could serve as predictors of high future costs. Using a priori clinical judgment, hospitalizations were classified as HCN (high cost next year) if the associated principal diagnosis suggested that the hospitalized individual had entered a state of chronic or long-term frailty requiring a continuing high level of care. It was clear from empirical analysis (even with the crudely coded clinical data available in the 1970's) that the subsequent costs of care for beneficiaries incurring an HCN hospitalization were, in fact, higher than the costs for other hospitalized beneficiaries (Ash et al., 1986).

Another important research contribution toward improving simple prior-use models by use of hospital diagnostic information was the more recent work of Anderson et al. (1986). They suggested that HMO payments should be influenced only by medically necessary prior hospital use, that is, use likely to have been judged necessary in any delivery system. The value of this approach is that actual HMO hospitalization experience could be used in rate adjustments without penalizing providers who reduce hospitalization rates by avoiding unnecessary episodes. A limitation of their methodology is that no pricing distinctions were made among hospitalizations based on the expected level of future need that was suggested by the reason for the hospital episode.

Both of these diagnostic payment models use a priori clinical judgment in defining diagnostic classifications for prior hospitalizations. In terms of explained variance, each approach yielded a model far superior to the current AAPCC payment model. In this article, we develop an alternative diagnostic approach, based on risk categories that we call diagnostic cost groups (DCG's). DCG's are formed using empirically determined similarities in the future costs of individuals hospitalized for different reasons. However, diagnoses thought to involve too much discretion in the decision to hospitalize are ignored in setting payments with the recommended model, irrespective of their future cost implications.

Although our primary objective was to extend and improve the HCN concept by empirically identifying

diagnoses associated with high future costs, clinical judgment was incorporated in the study in two ways. First, it was used to aggregate individual three-digit codes of the *International Classification of Diseases, 9th Revision, Clinical Modification*, or ICD-9-CM (Public Health Service and Health Care Financing Administration, 1980) into a manageable number of clinically meaningful subgroups for empirical analysis. Second, it was used to reclassify a number of diagnoses involving substantial ambiguity or discretion either in the definition of the diagnosis or in the medical necessity of hospitalizations for individuals with that diagnosis. Thus, although we did not consider physician discretion with the rigor reflected in the work of Anderson et al. (1986), clinical judgments about physician discretion were used in the model development.

Methodology

Data

The data for this study were drawn from HCFA's Continuous Medicare History Sample (CMHS), a file that contains demographic and Medicare usage information derived from claims forms for a systematic 5-percent sample of all Medicare beneficiaries. In turn, a 5-percent sample of the CMHS was randomly selected, yielding 38,705 beneficiaries, or 0.25 percent of the Medicare population. For each of the years 1974 through 1980, beneficiary records were obtained containing the following utilization data: total Part A (hospital) and Part B (physician) costs, number of hospital admissions and days in hospital, and a single code for principal diagnosis for each of up to five hospital admissions in each year.

An estimation file, based on 1975-77 data, and a test file, using the 3 years 1978-80, were extracted from this sample file for the purposes of model fitting and validation, respectively. In each of these two files, the first 2 years serve as a base period, and the final year serves as a target for prediction. In each file, eligible individuals were defined as those whose original Medicare entitlement was not for renal disease, who were at least 65 years of age on January 1 of the first base year, who were alive on January 1 of the target year, and who were eligible for both Parts A and B Medicare reimbursements during the entire 3-year period or until their death in the third year. There were 18,677 eligible individuals in the estimation file and 20,263 eligible individuals in the test file.

A third file, drawn directly from the CMHS (rather than from our 5-percent sample of it), contains data on a 1-percent sample of the eligible individuals who were hospitalized in 1979. (Here, "eligible" is defined as previously outlined, except that the base period is now the single year 1979 and the target year is 1980.) The file consists of one observation for each of the more than 71,000 hospitalizations experienced by the eligible enrollees in 1979, the principal diagnosis of

that hospitalization, and the total 1980 cost for the individual who experienced it. This episode file was used in establishing a classification of diagnoses into cost groups (the DCG's) based on different levels of 1980 costs for people who were hospitalized with different diagnoses in 1979.

The CMHS data have two shortcomings for the purposes of our study. The first is that the International Classification of Diseases system changed in 1978 from the Eighth Revision to the Ninth Revision, which is currently in use. Thus diagnostic groupings formed on 1979 hospital data can be only approximately replicated in the earlier time period. Also, a classification based on the current codes cannot be applied to more than 1 year of subsequent data (that is, 1980). Because the coding systems have a substantial amount of overlap, these problems were not considered crucial.

More importantly, however, major changes in coding and in other aspects of hospital behavior have been introduced since 1983 as a result of Medicare's change to prospective payments for hospital care under the diagnosis-related group system. This means that the current work needs to be validated with more recent data, both to see if the ideas are still applicable and, if they are, to find the proper prices in the new environment.

Development of diagnostic cost groups

The nine diagnostic cost groups used in this article were developed in two steps. First we aggregated approximately 800 three-digit ICD-9-CM diagnostic classifications into 78 diagnostic subgroups. This was done as follows. When an approximate minimum sample size of 100 was not met by an individual three-digit code, it was pooled with another code chosen on the basis of clinically judged similarity. Some codes with fewer than 100 episodes were retained as distinct if no clinically meaningful grouping could be identified; some codes with slightly more than 100 cases were pooled if a clinically close category was available. Otherwise we did not combine codes. The pooling was done to provide reliable estimates of expected 1980 costs for each group and yielded the 78 diagnostic subgroups shown in Table 1. Also shown are the number of individuals who incurred at least one hospitalization in 1979 with a principal diagnosis in each subgroup as well as the 1980 mean costs per person-year for the subgroups. In computing these means, the 1980 costs for a person who was hospitalized twice in 1979 with widely different principal diagnoses would contribute to establishing the expected future costs in two distinct subgroups. Costs were treated as costs per person-month of survival in the target year. Thus, for example, a person who died at midyear with \$10,000 in costs was counted as one-half of a person-year of experience, at a cost of \$20,000 per year.

In the second step of the analysis, the 78 diagnostic subgroups (based on 1979 utilization) were aggregated into nine diagnostic cost groups. The major criterion

Table 1

Original and revised diagnostic cost groups for Medicare enrollees hospitalized in 1979, by mean annualized 1980 cost, number of patients hospitalized, major diagnostic category name, and ICD-9-CM diagnostic subgroup

DCG	Mean annualized 1980 cost	Number of persons hospitalized in 1979	Major diagnostic category name ¹	Diagnostic subgroup and ICD-9-CM code
1	\$924	313	GENi	Genital prolapse (618)
1	1,064	17	PRGa	Pregnancy-related problems (630-676, 760-779)
1	1,134	169	GENh	Disorders of breast (610-611)
1	1,143	71	DIGd	Appendicitis, excluding other disease (540-542)
1	1,282	293	NRVf	Diseases of ear and mastoid processes (380-389)
1	1,343	179	GENg	Diseases of male genitalia, except prostate (603-609)
2	1,625	1,349	DIGi	Cholelithiasis and other disorders of gallbladder and biliary tract (574-576)
2	1,641	1,758	DIGe	Hernia of abdominal cavity (550-553)
2	1,771	184	RSPc	Other diseases of respiratory tract (470-478)
2	1,837	448	NEOe	Benign neoplasms (210-229)
2	1,848	1,602	GENf	Prostate disorders (600-602)
2	1,915	2,506	NRVb	Cataract (366)
3	2,055	330	NRVd	Disorders of the peripheral nervous system (350-359)
3	2,068	189	GENc	Urethral stricture (598)
3	2,070	329	GENj	Other diseases of female genital tract, excluding prolapse and abnormal bleeding (617, 619-629)
3	2,080	585	NRVe	Disorders of the eye and adnexa, except for NRVb and NRVg (360-364, 367-379)
3	2,154	141	NRVg	Glaucoma (365)
3	2,178	545	GENd	Hydrocephrosis, calculus of kidney and ureter, other disorders of kidney and ureter, calculus of lower urinary tract (591-594)
3	2,184	3,028	INJa	Injuries involving fractures and dislocations (800-839)
3 ²	2,201	456	MSKd	Rheumatism, excluding the back and polymyalgia rheumatica (726-729)
3	2,238	879	MSKc	Dorsopathies, except for inflammatory spondylopathies (721-724)
3	2,247	211	GENb	Kidney infections (590)
3	2,266	392	INJd	Superficial injury (910-919); contusions (920-924); effects of object entering through orifice (930-939)
3	2,306	624	SYMa	General symptoms (780)
3	2,329	589	INJc	Intracranial injury (excluding skull fracture), internal injury, open wounds, injury to blood vessels, late effects of injuries, poisonings, toxic effects and external causes (850-909); crushing injury (925-929); burns (940-949); injury to nerves, spinal cord, unspecified injuries (950-959)
3	2,404	161	DIGa	Diseases of oral cavity, salivary glands, and jaws (520-529)
3	2,441	175	RSPb	Acute respiratory infections, except bronchitis (460-465)
4 ²	2,470	820	DIGf	Noninfective enteritis and colitis (555-556, 558)
4	2,510	197	CNGa	Congenital anomalies (740-759)
4 ²	2,528	305	INJb	Sprains and strains of joints and muscles (840-848)
4	2,549	639	NEOc	Malignant neoplasm of bone, skin, cartilage, soft tissue, male and female breasts (170, 171, 173, 174, 175); skin melanoma and carcinoma (172)
4	2,571	1,903	DIGg	Intestinal obstruction (nonherniated), diverticula of intestine (560-562); peritonitis and other disorders of intestine or peritoneum (567-569)
4 ²	2,573	1,580	GENe	Disorders of urethra and urinary tract (595-597, 599); urinary symptoms and nonspecific findings on urine examination (788, 791)
4 ²	2,599	1,508	DIGb	Gastric, duodenal, peptic, and gastrojejunal ulcer, diseases of the esophagus (530-534); gastrointestinal hemorrhage (578)
4 ²	2,645	1,173	CRCh	Diseases of veins and lymphatics and diseases of circulatory system (451-459); anal fissures, fistulae and anal or rectal abscess (565-566)
4 ²	2,677	326	INFa	Infectious diseases (001-139) except those in INFb
4 ²	2,690	891	DIGc	Gastritis and duodenitis and other disorders of stomach and duodenum (535-537); functional digestive disorders (564); digestive symptoms (787)
4 ²	2,713	298	ENDe	Diseases of other endocrine glands (251-254, 256, 257, 259); disorders of the thyroid gland (240-246); nutritional and metabolic symptoms (783); nonspecific findings on blood examinations (790)
4	2,789	3,171	CRCe	Cerebrovascular disease (430-438)
4 ²	2,800	511	SYMc	Respiratory symptoms (786)
4 ²	2,823	1,628	RSPd	Pneumonia and influenza (480-487)
4 ²	2,832	911	SKNa	Diseases of the skin and subcutaneous tissue (680-709); symptoms involving skin (782)
4 ²	2,854	434	SYMd	Symptoms involving abdomen and pelvis (789)
5 ³	2,874	1,241	MSKa	Diseases of connective tissue (710); rheumatoid arthritis and inflammatory polyarthropathies (714); osteoarthritis and like disorders (715); rheumatic fever and rheumatic heart diseases (390-398); polyarthritis and like conditions (446); inflammatory spondylopathies (720); polymyalgia rheumatica (725); symptoms of nervous and musculoskeletal systems (781)

See footnotes at end of table.

Table 1—Continued

Original and revised diagnostic cost groups for Medicare enrollees hospitalized in 1979, by mean annualized 1980 cost, number of patients hospitalized, major diagnostic category name, and ICD-9-CM diagnostic subgroup

DCG	Mean annualized 1980 cost	Number of persons hospitalized in 1979	Major diagnostic category name ¹	Diagnostic subgroup and ICD-9-CM code
5 ³	2,908	905	MSKb	Various arthropathies (711-713, 716); disorder and derangement of joints (717-719); osteopathies, chondropathies, and acquired musculoskeletal deformities (730-739)
4 ²	2,915	33	GENk	Inflammatory disease of female pelvic organs (614-616)
5 ³	2,924	5,613	CRCb	Ischemic heart disease (410-414); diseases of pulmonary circulation (415-417); cardiovascular symptoms (785)
5	2,945	393	CRCg	Aortic aneurysm (441); other aneurysms (442); arterial embolism and thrombosis (444)
5 ³	2,953	928	MNTb	Neurotic, personality, and sexual disorders (300-302); alcohol and drug dependence and abuse (303-305); other personality disorders, nonpsychotic mental disorders, and mental retardation (306-319)
0	2,958	4,349	SYMe	Nonspecific abnormal findings (792-796); other ill-defined and unknown causes of morbidity and mortality (799)
5 ²	3,052	255	EVLa	Supplementary classification of factors influencing health status (V01-V82)
5 ²	3,082	436	RSPa	Acute bronchitis and bronchiolitis (466)
5 ³	3,068	453	MNTa	Organic psychotic conditions (290-294); other psychoses (295-299); senility without psychosis (797)
5	3,103	625	DIGj	Diseases of pancreas, intestinal malabsorption (577, 579)
5 ³	3,107	1,096	CRCa	Hypertensive disease (401-405)
5 ³	3,131	678	CRCf	Atherosclerosis (440); other vascular disease (443); disorders of arteries, arterioles, and capillaries (447-448); vascular insufficiency of intestine (557)
5	\$3,176	620	ENDb	Nutritional deficiencies (260-269); metabolic disorders (270-275, 277); obesity and immune disorders (278, 279); disorders of blood and blood-forming organs (280-289); adrenal and polyglandular disorders (255, 258)
5	3,178	430	NEOf	Carcinoma in situ in skin and on other unspecified sites (232, 234); neoplasms of uncertain behavior or unspecified nature (235-239); malignant neoplasm of the lip (140)
5	3,227	331	ENDc	Disorders of fluid, electrolyte, and acid base balance (276)
5 ²	3,228	2,187	CRCc	Other forms of heart disease, except heart failure (420-427, 429)
5 ²	3,267	144	INJf	Toxic effects of nonmedical substances (980-989); unspecified effects of external causes (990-995)
5	3,284	306	NRVa	Other disorders of the central nervous system (340-349)
5 ²	3,305	223	INJe	Poisoning by drugs, medicines, and biological substances (960-979)
5	3,312	276	NRVc	Hereditary and degenerative diseases of central nervous system (330-337)
5 ²	3,378	457	INJg	Complications of medical care not elsewhere classified (996-999)
5 ²	3,690	131	SYMb	Symptoms involving head and neck (784)
6	3,714	1,491	ENDa	Diabetes mellitus (250)
6	3,796	711	NEOa	Malignant neoplasm of digestive organs and peritoneum (150-159); carcinoma of digestive organs (230)
6	3,821	1,039	NEOd	Malignant neoplasm of genitourinary organs (179-189)
7	4,036	177	DIGH	Liver disorders and diseases (570-573)
7 ²	4,083	150	INFb	Various infectious and parasitic diseases (013, 038, 045-049, 070, 093-095, 112, 114-116, 135, 320-326)
7 ²	4,114	290	ENDd	Other and unspecified anemias (285)
7	4,139	1,806	RSPe	Chronic obstructive pulmonary disease and like conditions (490-496); pneumoconioses and other lung diseases due to external agents (500-508)
7	4,166	800	RSPf	Other diseases of the respiratory system (510-519)
8	4,438	1,750	CRCd	Heart failure (428)
8	5,018	837	NEOg	Malignant neoplasm of oral cavity and pharynx (141-149); malignant neoplasm of unspecified sites (190-199); malignant neoplasm of lymphatic and hematopoietic tissue (200-208)
8	5,771	349	NEOb	Malignant neoplasm of respiratory and intrathoracic organs (160-165)
9	13,151	182	GENa	Nephritis, nephrotic syndrome, and nephrosis (580-589)

¹First 3 letters indicate major diagnostic category to which included ICD-9-CM codes principally belong. Major diagnostic category names are useful for referencing Figure 1.

²Entire diagnostic subgroup reclassified into DCG O (null group) in revised classification.

³Subset of ICD-9-CM codes removed from diagnostic subgroup and reclassified into DCG O (null group). Changes are as follows: MSKa—781 dropped; MSKb—only 711 and 730 retained; CRCb—785 dropped; MNTa—797 dropped; CRCa—401 dropped; CRCf—only 447 and 557 retained; MNTb—only 303 and 304 retained.

NOTES: ICD-9-CM is *International Classification of Diseases, 9th Revision, Clinical Modification*. DCG is diagnostic cost group. Groups are ordered by increasing cost. Original DCG classification shown at left. Changes (explained in footnotes) aimed at removing highly discretionary hospitalizations from list of events that cause high future payments.

SOURCE: Health Care Research Unit, Boston University: Analysis of data from the Health Care Financing Administration's Continuous Medicare History Sample.

Figure 1
**Annual 1980 cost for Medicare enrollees, by presence of at least one 1979 hospitalization in each of
 78 diagnostic subgroups: United States**

1980 Cost	Major diagnostic category name ¹	Diagnostic cost group
\$800	GENi	
Whole population mean → \$1,000	DIGd, GENh, PRGa	1
\$1,200	NRVf, GENg	
\$1,400		
\$1,600	DIGe, DIGi	2
\$1,800	NEOe, RSPc, GENf	
\$2,000	NRVb	
Mean for hospital population → \$2,200	NRVd, NRVe, GENc, GENj	
\$2,400	NRVg, GENb, GENd, MSKc, MSKd, INJa	
\$2,600	SYMa, INJc, INJd	
\$2,800	RSPb, DIGa	
\$3,000	NEOc, DIGf, DIGg, CNGa, INJb	
\$3,200	CRCh, DIGb, GENe	
\$3,400	INFa, ENDe, DIGc	
\$3,600	CRCe, RSPd, SKNa, SYMc	
\$3,800	CRCb, CRCg, GENk, MSKa, MSKb, SYMd	
\$4,000	MNTb, SYMe	
\$4,200	MNTa, CRCa, CRCf, RSPA, DIGj, EVLa	
\$4,400	NEOf, ENDb, ENDc, CRCc	
\$4,600	NRVa, NRVc, INJe, INJf	
\$4,800	INJg	
\$5,000	ENDa, SYMb	6
\$5,800	NEOa, NEOd	
\$5,800	DIGH	
\$4,000	INFb, ENDd, RSPe	7
\$4,200	RSPf	
\$4,400	CRCd	
\$4,600		
\$4,800		
\$5,000	NEOg	
\$5,800	NEOb	
\$13,200	GENa	9

¹ Definitions can be found in Table 1.

SOURCE: Health Care Research Unit, Boston University: Analysis of data from the Health Care Financing Administration's Continuous Medicare History Sample.

for inclusion in the same cost group was similar average total costs in 1980. A histogram of these costs is shown in Figure 1. Two clusters of lower cost diagnostic subgroups and four with higher future-cost implications than the middle group (with costs of \$2,100-\$3,400) can be seen. These clusters eventually became DCG's 1 and 2 (lower cost) and 6-9 (higher cost). The middle cost range, which did not produce

distinct clusters, was split into three groups. Each of these three (DCG's 3, 4, and 5) has an expected cost range equal to approximately \$400, which is somewhat larger than the standard error for mean 1980 costs within most of the 78 subgroups. It is simpler to work with 9 groups with than 78, and little loss in predictive accuracy resulted from the aggregation.

A diagnosis at the low end of a break point in terms of cost was classified in the next lower DCG if, in the opinion of our clinicians, substantial ambiguity or discretion existed in either the definition of the diagnosis or the appropriateness of hospitalization for a person with that diagnosis. The reclassification eliminated the bonus an HMO would otherwise receive for using relatively well-paid, nonspecific categories of diagnosis or for hospitalizing individuals who might be properly treated otherwise. For example, the diagnostic category that comprises symptoms involving head and neck (code 784) was downgraded from DCG 6 to DCG 5 because of concerns about vagueness. In the stub of Table 1, we indicate the DCG for each of the 78 diagnostic subgroups. On a second pass, a more drastic downward reclassification was undertaken, placing many diagnoses in DCG 0, which is equivalent to no hospitalization. Footnotes in Table 1 indicate that all (footnote 2) or some (footnote 3) of the diagnoses that were originally classified in the DCG shown in the stub were eventually reclassified as DCG 0. Pneumonia and influenza (codes 480-487), for example, were downgraded to DCG 0 because the decision to hospitalize many individuals with these conditions is highly discretionary.

This completed the classification of all hospitalizations into 1 of 10 categories, from DCG 0 to DCG 9. Individuals were then classified on the basis of the highest numbered DCG into which any of their hospitalizations fell. Those with no hospital experience in the base year were included in DCG 0. The DCG approach is based on the following presumptions: first, that the rates of occurrence of the high-cost DCG's in a population are a reasonable proxy for the health status of the population; second, that a subsequent year's health care costs can be reasonably estimated through this proxy measure of need.

The 1980 costs of people with different 1979 utilization experiences span a wide range. These differences in costs are summarized using relative costliness indexes (RCI's). The RCI for a subgroup of enrollees is the ratio of the group's average cost to the average cost of the total Medicare study population. Although actual dollar costs change markedly from year to year, the relative costliness of individuals within DCG groups (as reflected in their RCI) should be quite stable. The 79.1 percent of Medicare enrollees who had not been hospitalized at all in 1979 experienced a 1980 RCI of 0.75 (computed as \$840, their 1980 average cost, divided by \$1,113, the mean 1980 cost for all Medicare enrollees). When the 7.0 percent classified in DCG 1, 2, or 3 were pooled with those never hospitalized, the RCI for this group (now 86.1 percent of the population) rose only to 0.77. Finally, when an additional 5 percent of the population was reclassified into DCG 0 in the revisions indicated by footnotes in Table 1, the RCI for this so-called BASE group became 0.83. Thus, in the final model, DCG 0 is a mix of those not hospitalized and those whose hospitalizations either

Table 2
Medicare enrollee groups based on 1979 diagnostic cost groups (DCG's), percent of Medicare population, and 1980 relative costliness indexes: United States

Group	DCG ¹	Percent of total Medicare population	1980 relative costliness index ²
Base (the nonhospitalized and those hospitalized with appendectomy, hernia, prostate, etc.)	0-3	91.1	0.83
High cost (those hospitalized with ischemic heart disease, internal obstruction, urinary tract infection, etc.)	4,5	5.6	2.22
Very high cost (those hospitalized with diabetes mellitus, cancer, obstructive pulmonary disease, etc.)	6-8	3.2	3.49
Renal failure (those hospitalized with nephritis, nephrotic syndrome, etc.)	9	0.07	11.50

¹Final (that is, revised) DCG categories, in which some higher cost episodes have been reclassified in DCG 0 because of concerns about the highly discretionary nature of these hospitalizations.

²Mean 1980 cost for group divided by mean 1980 cost for all eligible Medicare enrollees.

SOURCE: Health Care Research Unit, Boston University: Analysis of data from the Health Care Financing Administration's Continuous Medicare History Sample.

had relatively low expected future costs or were highly discretionary. The actuarially correct price for this mixed population is only about 10 percent higher than the price for the nonhospitalized people alone (0.83 versus 0.75).

Relative group sizes and their 1980 RCI's based on the 1979 DCG classification are shown in Table 2. There, we can see that RCI's for the more expensive DCG's range from more than 2 to as high as 11.50.

The DCG classifications were checked against the list of HCN diagnoses reported in the work of Anderson et al. (1982) and Anderson, Resnick, and Gertman (1983). Reassuringly, there was substantial agreement between the clinically judged high cost next year diagnoses and those that actually were followed by high costs in the following year.

A second check on the validity of the DCG classification as a way of identifying individuals who will experience a continuing need for high levels of health services in the future was to extend the prediction period from 1 to 3 years. Using 1976 as the base year, future costs for 1977 alone and for the 3-year period 1977-79 were regressed on age, sex, welfare status, and DCG classification. The reimbursement per year for each DCG was similar in both analyses. This suggests that the DCG designation captures some element of long-term chronicity.

Table 3
Selected Medicare enrollee subgroups and their costs: United States

Subgroup	Number of persons	Percent of total Medicare population	Mean 1980 cost	Relative costliness index ¹
ALL—All individuals eligible for the test file	20,263	100	\$1,113	1.00
ADM—Number of hospital admissions in 1979 is:				
0	16,091	79	840	0.75
1	2,930	14	1,727	1.55
2	834	4	2,448	2.20
3	248	1	3,976	3.57
4 or more	160	1	5,879	5.28
COST79—Total Part A and Part B costs in 1979 is:				
\$0	8,052	40	551	0.50
\$1-69	3,057	15	821	0.74
\$69-218	3,056	15	1,057	0.95
\$218-1,354	3,046	15	1,460	1.31
More than \$1,354	3,052	15	2,597	2.33
65YRF—Women age 65-69 not receiving welfare and without prior disability	3,593	18	790	0.71
NOCOST—Those in 65YRF who also have no 1979 costs	1,594	8	386	0.35
CVD/CA—1 1979 admission or more for cerebrovascular disease or cancer	526	3	2,805	2.52
2HOSPS—At least 2 1979 admissions	1,242	6	3,195	2.87

¹Ratio of the mean 1980 cost for the group to the mean 1980 cost for the whole population.

SOURCE: Health Care Research Unit, Boston University: Analysis of data from the Health Care Financing Administration's Continuous Medicare History Sample.

Testing models

Our methodology for comparing models is similar to that of Beebe, Lubitz, and Eggers (1985). Each of several candidate models (six of which are described in detail here) was fitted to the estimation file of 1975-77 data described in the "Data" section. Least-squares regressions with predictor variables relating to usage in 1975 and 1976 (the base years) were constructed to predict annualized total costs for eligible persons in 1977 (the target year). Observations were weighted by the fraction of the target year during which the person remained alive and eligible for FFS reimbursement. Models were then compared by studying their behavior on the test file of data relating to the period 1978-80. This procedure simulates the way in which the methodology might actually be applied, because some gap will certainly exist between the availability of data for model fitting and the time at which prospective payment levels must be set.

The fitted models, with coefficients for the base year variables derived from the estimation file, were applied to the new base years, 1978 and 1979, in order to predict costs in the new target year, 1980. To adjust for inflation, each predicted value was multiplied by the ratio of total actual 1980 costs to total predicted costs. This had the effect of making the sum of all the new predictions for 1980 add to the total 1980 actual costs without changing the relative size of cost predictions for individuals.

Each model was thus determined on the basis of information obtained from an earlier period and used to predict 1980 costs for anyone for whom 1978 and 1979 usage data were available. Models were then evaluated on the basis of how well their predictions fit the actual 1980 costs of the people in the test set, as described next.

Measures of goodness of fit

We have found it useful to compare the performance of models for predicting future costs using several different measures of a model's goodness of fit. Brief descriptions of these follow. Explicit definitions can be found in the "Technical note."

The first measure is the conventional R^2 calculated by regressing 1977 costs on 1975-76 information. We call it the 1977 R^2 . A cross-validating, or 1980, R^2 is also reported. This is a measure of how well the 1980 predictions made using the estimation file (as just described) match 1980 expected costs.

The R^2 is a measure of how well a formula predicts actual costs for individuals. For our purposes, a formula need only be able to make good predictions for the expected costs of groups of people who, if not priced correctly, might be discriminated against in HMO's. (Newhouse, 1986, discusses this point.) The current AAPCC can make accurate predictions for a good-sized group of randomly selected individuals. However, models that are able to successfully

distinguish future costs for groups with atypical health histories are required so that the right payments can be made when an HMO experiences biased enrollment.

Four measures of prediction bias, called predictive ratios (PR's), were produced to evaluate how closely each model predicts the average 1980 cost for four different subgroups of the test file, which were selected on the basis of information available during the base years, 1978 and 1979. The predictive ratio for a particular group and a given model is formed by dividing the total costs that the model predicts for the group by the total costs actually incurred by the group in 1980. PR's greater than 1 indicate groups for which the model will lead to overpayment; PR's less than 1 reflect groups whose costs are higher than the model predicts. The best models will have all PR's for a wide selection of subgroups quite close to 1.

For this article, predictive ratios have been computed for four groups, constructed to address a range of questions about our ability to correctly predict costs for people with specific characteristics. Information about the number of people in these groups and their average 1980 cost experience is given in Table 3. The four groups are named on the left and defined as follows:

65YRF	Women 65-69 years of age (in the first base year) who were not receiving welfare and whose original reason for Medicare entitlement was becoming 65 years old (rather than a disability identified at an earlier age).
NOCOST	A subset of those in group 65YRF defined by having no Medicare reimbursement in the prior year.
CVD/CA	Anyone hospitalized in the prior year for cerebrovascular disease and/or cancer (<i>Eighth Revision International Classification of Diseases</i> codes 140-201 or 430-438).
2HOSPS	Anyone with at least two hospitalizations in the prior year.

The first group was selected by demographic characteristics alone and should otherwise exhibit a representative distribution of prior and future health utilization needs. The other groups all had atypical prior health utilization. The expected Medicare costs of these four groups differ considerably. The relative costliness index (ratio of the mean cost of a subgroup to the total population mean cost) for these groups varied from a low of .35 (NOCOST group) to 2.87 (2HOSPS group). Models were compared on their ability to yield PR's near 1 for each of the four groups.

In summary, we have six different measures of performance for each of the models under consideration: its R^2 for both the estimation and the test data sets and its predictive ratio for the four subgroups.

Comparing regression models

In this article, six regression models that could be used in setting HMO payments are compared for predictive power. The first four regressions, which have appeared in the literature, establish benchmarks against which the performance of the two new diagnostic models may be judged. All models except Model 4 use the age, sex, and welfare status information that makes up the current AAPCC. (Institutional status, the final factor in the current AAPCC, could not be included because the data were unavailable.) Model 1 uses only this demographic information. Models 2 and 3, originally proposed by Beebe, Lubitz, and Eggers (1985), also use administrative information relating to utilization, which is maintained for all beneficiaries in HCFA's Health Insurance Master Accretions (HIMA) file. Neither model distinguishes among hospital episodes on the basis of diagnostic information. Model 2 uses the presence or absence of any hospital admission in the 1 year immediately preceding the target (or prediction) year; thus, we can evaluate the additional benefit of diagnostic information by comparing its performance with that of the DCG regressions (Models 5 and 6). Model 3, implemented in the Senior Health Plan demonstration, uses 2 years of data and counts total days of hospital stay as well as whether Part B (physician expense) deductibles were met. The last of the benchmark models, Model 4, bases its predictions on a single variable that cannot be obtained from the HIMA file, that is, the dollar amount of Part B expenses.

Models 5 and 6 employ the DCG methodology. The first of these differentiates among five different groups of people classified using the original DCG designations given in Table 1. Model 6 is coarser grained, using fewer groups (DCG's 0-3, 4 and 5, and 6-8), and is based on the revised DCG categories. (Categories footnoted in Table 1 were downgraded into the DCG 0 group.) In developing Model 6, we assumed that individuals in DCG 9, who were hospitalized for treatment related to renal failure, would be reimbursed by a separate, noncapitated mechanism; these individuals were dropped from the analysis.

Results

In this section, we examine the relative empirical performance of the models just described. Table 4 contains the estimated regression coefficients for each of the models. All variables were statistically significant at the 0.05 level (two-sided). Table 5 displays the six goodness-of-fit measures for each model.

The best models cannot predict individual annual costs well. Even with the most complete and complex models that we have ever specified with these data, the R^2 values obtained have not exceeded 10 percent

Table 4
Regression coefficients for models to predict cost for Medicare enrollees: United States

Variable	Model					
	1	2	3	4	5	6 ¹
CONSTANT	556	395	91	702	443	597
AGE	23.9	20.3	17.2	—	21.0	35.4
MALE	234	211	264	—	192	126
WEL	245	171	55	—	162	179
ANYHOSP	—	1,075	—	—	—	—
ANYB1	—	—	157	—	—	—
ANYB2	—	—	494	—	—	—
DYS	—	—	35	—	—	—
BCOST	—	—	—	1.94	—	—
DCG4	—	—	—	—	861	—
DCG5	—	—	—	—	1,013	—
DCG678	—	—	—	—	2,068	—
DCG9	—	—	—	—	20,382	—
HIGH	—	—	—	—	—	2,537
VERYHIGH	—	—	—	—	—	3,994

¹The variables CONSTANT, AGE, MALE, and WEL of Model 6 are used only for predictions in the BASE group. In the revised DCG's of this model, diagnoses were removed and placed in the BASE group, as indicated by footnotes 2 and 3 in Table 1, because of concerns about the discretionary nature of hospitalization for these reasons. The DCG 9 individuals were removed entirely under the assumption that they would be paid through the end stage renal disease program.

NOTES: DCG is diagnostic cost group. Cost is total Part A plus Part B dollars this year. AGE is age in years minus 67. MALE is 1 if male; otherwise 0. WEL is 1 if welfare buy-in last year; otherwise 0. ANYHOSP is 1 if hospitalized at all during previous 2 years; otherwise 0. ANYB1 is 1 if Part B deductibles were met 2 years ago; otherwise 0. ANYB2 is 1 if Part B deductibles were met last year; otherwise 0. DYS is total days in hospital in the 2 base years. BCOST is Part B dollars used last year. DCGxyz is 1 if most costly diagnosis last year is in DCGx, DCGy, or DCGz; otherwise 0. VERYHIGH is 1 if most costly diagnosis last year is in revised DCG 6, 7, or 8; otherwise 0. HIGH is 1 if most costly diagnosis last year is in revised DCG 4 or 5; otherwise 0. BASE is 1 if not HIGH or VERYHIGH; otherwise 0.

SOURCE: Health Care Research Unit, Boston University: Analysis of data from the Health Care Financing Administration's Continuous Medicare History Sample.

Table 5
Goodness of fit of measures of models for predicting annual rates of Medicare costs: United States

Model and description	Data needs ¹	<i>R</i> ² × 100		Predictive ratio			
		1977	1980	65YRF ²	NOCOST ²	CVD/CA ²	2HOSPS ²
1. Adjusted average per capita cost (demographics only)	H	0.5	-1.4	1.00	2.05	0.38	0.34
2. Beebe ³ (presence of 2nd base year admissions)	H	2.6	1.2	0.98	1.47	0.74	0.66
3. Beebe ² ³ (presence of Part B cost and number of hospital days during both years)	H	4.9	4.4	0.98	0.62	0.71	0.76
4. Part B cost in 2nd base year alone	B	8.5	4.6	1.02	1.00	1.02	1.09
5. Original DCG's (5 groups)	D	4.7	3.8	1.00	1.72	0.87	0.77
6. Revised DCG's (3 groups, demographics)	D	4.5	3.1	1.00	1.73	0.92	0.70

¹H—Uses Health Insurance Master Accretions (HIMA) file information only. B—Requires total Part B costs, not available on the HIMA file. D—Requires diagnostic information on hospitalizations, not available on the HIMA file.

²Definition can be found in Table 3.

³(Beebe, Lubitz, and Eggers, 1985).

NOTES: Explicit model descriptions can be found in Table 4. DCG is diagnostic cost group.

SOURCE: Health Care Research Unit, Boston University: Analysis of data from the Health Care Financing Administration's Continuous Medicare History Sample.

for the conventional 1977 *R*² and 7 percent for the cross-validated 1980 *R*². Most of the variability in annual costs remains unexplained. Even if we knew a person's expected cost exactly, actual expenditures in any given year would vary widely around this average. For example, Welch (1985) shows that certain plausible assumptions about the functional form of the autocorrelation of costs between 2 years lead to the conclusion that the maximum achievable *R*² is .20. In light of this, the *R*² values achieved by the prior-use models of this article may be seen as explaining a substantial part of the explainable variance.

Although each model has a smaller cross-validating 1980 *R*² than its conventional 1977 *R*², the relative rankings of the models by these two measures is the same. The current AAPCC formula is the least predictive of the models. It has a 1977 *R*² of only 0.5 percent, and its 1980 *R*² is even slightly negative. Model 4 is the best predictive model tested. It has a 1977 *R*² of 8.5 percent, and its 1980 *R*² is 4.6 percent. Models 5 and 6 (the two DCG models) and Model 3, which uses information from the HIMA file on hospital days and Part B deductibles, all have similar explanatory power. Somewhat lower *R*² values are

obtained for Model 2, which does not distinguish among hospitalizations through diagnostic information. Only a small loss in explanatory power results from the revisions concerning discretionary admissions and the coarser classification of DCG's in Model 6 relative to Model 5.

All models yielded predictive ratios close to unity for the group 65YRF (defined only on the basis of AAPCC demographic attributes). Thus, the AAPCC, despite its low R^2 , appears to be fully adequate for paying HMO's that enroll a representative mix of enrollees. On the other hand, its predictive ratios were considerably more variable for the other three subgroups than the predictive ratios for the other models were. The current AAPCC formula, for example, dramatically overpays the NOCOST subgroup and underpays those in the two hospitalized groups (CVD/CA and 2HOSPS). Like the R^2 results, the subgroup predictive ratios were best (that is, closest to unity) for Model 4 (the Part B reimbursement model) and by far the worst for Model 1. Among the remaining models, the HIMA file models (2 and 3) have better predictive ratios than the DCG models (5 and 6) for the NOCOST subgroup, but the DCG models have better PR's for the CVD/CA subgroup. There is little to recommend one DCG model over the other on the evidence of these four predictive ratios.

That the ability to capture future cost differences increases with each model as one moves from Model 1 to Model 4 could have been predicted on the basis of previous research and logic. Demographics alone (Model 1) are known to have low predictive power. Classification by whether a person was hospitalized (Model 2) only splits off the 20 percent who are hospitalized from a heterogeneous residual group. Of the residual group, about one-half have no Medicare expenses, and the rest incur ambulatory care costs ranging from minimal to extremely high. Model 3 has more power because it distinguishes the healthy subset of individuals with no costs to Medicare. Model 4 has the most power of all because even better distinctions among those who use different amounts of ambulatory care can be made by using Part B dollars.

The promising feature of the DCG models is that, despite their inability to make any distinctions among the 80 percent who are not hospitalized, they perform only a little less well than Model 3, which does make such distinctions. In the case of Model 6, in particular, this is true even after the deletions for discretion, which means that less than 10 percent of the people are being distinguished for special payments based on past utilization.

As noted by Thomas et al. (1983), predictive performance is not the sole criterion for evaluating alternative payment models for HMO's. Administrative feasibility and the invulnerability of risk classification to potential manipulation by providers are crucially important for a successful payment system. Despite its strong predictive performance, the Part B reimbursement model has significant drawbacks on grounds of administrative

infeasibility and the sensitivity of payment to potential provider manipulation. Part B reimbursement data are not contained in readily accessible HCFA files, and payment rates in a subsequent year would increase for each dollar expended by providers on Part B services. The strong predictive results of the Part B reimbursement model are indicators of the value of research aimed at improving the AAPCC further by using health status information obtained during ambulatory encounters.

Although the HIMA file models are quite attractive on grounds of administrative feasibility, they are also more likely to be subject to provider manipulation than the DCG models are. For example, increased rates associated with meeting Part B deductibles may encourage providers to incur minimal costs so that enrollees meet such a threshold. The DCG models did not surpass all others in predictive performance, but the relatively modest decrease in predictive performance associated with their use may be reasonably traded off with other factors in choosing among alternative HMO payment models.

Discussion and conclusions

The diagnostic models developed here could be used to adjust HMO payments to reflect HMO and local FFS differences in population health status composition in the following way. The first two components of the current AAPCC formula, namely, the projected U.S. per capita Medicare cost and the county geographic adjustment, would be unaffected by the use of the new model. However, the third component of the AAPCC formula, which is a ratio of the average risks for the HMO's enrollees and for members of the FFS population in the counties from which HMO enrollees are drawn, would be estimated using a DCG model.

HMO payments would change under a DCG payment model to the extent that the average risk factor of HMO enrollees relative to the average risk factor of their FFS counterparts differs between the current AAPCC and DCG risk classifications. For example, an HMO may enroll beneficiaries who are quite similar to local FFS beneficiaries in terms of their distribution across current AAPCC risk categories; this would yield a ratio near unity for the third component of the current AAPCC formula. Suppose, however, that significantly greater fractions of the HMO's enrollees than local FFS beneficiaries have experienced hospitalizations with DCG primary diagnoses. Then this ratio would be greater than unity, and HMO payments would be higher than payments under the DCG model.

The data needs for implementing such DCG models are far more demanding than those of the current AAPCC; they are also more demanding than those of prior-use models that require only HIMA file data. However, DCG information is much easier to collect than prior-use expenditure information or any data relating to outpatient utilization would be. This is an important factor affecting administrative feasibility,

because the current regulations of the Tax Equity and Fiscal Responsibility Act require HMO payments to be based on expected FFS experience. Thus, any risk classification employed for HMO enrollees must also be used to classify FFS beneficiaries. Because Medicare already requires hospitals to collect diagnostic information for reimbursement purposes under its prospective payment system, the additional data burden would not be as great as it could be with other types of utilization information that might be used as proxies for health status.

An attractive feature of risk classification based on hospital diagnostic information is that it opens up the possibility of setting capitation rates for newly eligible beneficiaries, for enrollees switching from other HMO's, and for continuing HMO enrollees whose health status should be updated. The required diagnostic information for such people should be retrievable through hospital records that can be verified by Medicare audits. Implementation of a model of the kind discussed here would require further work aimed at assuring interphysician recording consistency and, in addition, testing to assure that manipulation resulting in a form of diagnostic coding upgrading ("creep") does not occur.

The value of other data or a longer base period should continue to be explored. Other data could include more detailed information regarding the hospital stay, such as surgical or other procedures, comorbidities, and severity measures. Ambulatory care diagnoses or risk factor information not associated with a particular hospital stay, such as laboratory test results or documentation of functional impairments, might also be used. If predictions can be made substantially more accurate with such data, then the benefit of ensuring that people are properly identified and priced will have to be weighed against the cost of additional data collection and auditing.

When a payment model is used to update risk classifications for HMO enrollees on the basis of HMO utilization, the issue of whether the diagnostic classifications reward inefficient provider behavior becomes key. In recently completed research, Ellis and Ash (1988) developed DCG classifications using the better data reported under Medicare's prospective payment system for hospitals and addressed the issue of physician discretion with more rigor than in the research reported here. In that work, level of physician discretion is defined on the basis of clinical judgments of a physician panel. Approaches for empirically testing the reliability of less discretionary codes (for example, via studies of recording consistency or comparisons of HMO and FFS usage) need to be developed. For the present, however, the incorporation of health status adjustments into an HMO payment model through diagnostic information

holds significant promise for improving the current AAPCC method of paying Medicare HMO's.

Acknowledgments

We are grateful to Michael Shwartz, Randall Ellis, Susan Payne, John Lamperti, Joseph Restuccia, and three careful anonymous reviewers for helpful comments.

Technical note: Goodness-of-fit measures

Conventional and cross-validated R^2

When a model is fit to data, a convenient measure of how well its predicted values match the actual ones is the R^2 . For each model fitted to 1977 outcomes based on 1975-76 data, an R^2 for the 1977 predictions is computed in the usual way as:

$$R^2 = 1 - SS(\text{predicted} - \text{actual})/SS(\text{total}),$$

where $SS(\text{predicted} - \text{actual})$ is the sum, over all observations in the estimation file, of the squares of differences between predicted and actual costs in the target year 1977. With weighted regressions, all sums are weighted sums. For each model, the predicted costs are computed by applying the regression formula to the base-year data, the same data that were used to fit the formula. Then $SS(\text{total})$ is the sum of the squared deviations of each actual cost in the estimation sample from the average such cost.

Because the models are fit using one set of data and will be used to make predictions at a later time, these 1977 R^2 values overstate our ability to predict costs. Thus, we tend to put more faith in a second and more realistic measure of goodness of fit, the 1980 R^2 . It has the same form as the equation previously given. However, in this case, the predicted values are made using a formula that employs the estimated parameters obtained from the 1975-77 sample applied to the test-set base-year information (1978-79) and multiplied by an inflation factor. Actual costs are from 1980.

With this methodology, there is no guarantee that a model's 1980 R^2 will be positive. From the equation, we can see that the 1980 R^2 would achieve its maximum value of 1.00 only if predicted and actual costs were identical for every person.

In computing these R^2 values, each person's actual costs are compared with predicted annual rates of expenditure multiplied by the fraction of the year for which the individual remained alive. This reflects the reality that when a person dies in midyear, the HMO receives only one-half of the yearly premium.

Predictive ratios

The predictive ratio for a given subgroup and a given model is defined as

$$PR = \frac{\text{Average of the model's predictions}}{\text{Average of actual costs}},$$

where each of the two averages is taken over the individuals in the subgroup.

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*. Vol. 8, No. 2. HCFA Pub. No. 03226. Office of Research and Demonstrations, Health Care Financing Administration. Washington. U.S. Government Printing Office, Winter 1986.
- Anderson, G. F., and Knickman, J.: Patterns of expenditures among high utilizers of medical care services. *Medical Care* 22(2):143-149, Feb. 1984.
- Anderson, J., Resnick, A., English, P., and Gertman, P. G.: Prediction of Subsequent Year Reimbursement Using the Medicare History File: I. Results from the Preliminary Model. Discussion Paper No. 45a. University Health Policy Consortium, Brandeis University. Waltham, Mass. 1982.
- Anderson, J., Resnick, A., and Gertman, P. G.: Prediction of Subsequent Year Reimbursement Using the Medicare File: II. A Comparison of Several Models. Discussion Paper No. 45b. University Health Policy Consortium, Brandeis University. Waltham, Mass. 1983.
- Ash, A., Porell, F., Gruenberg, L., et al.: *An Analysis of Alternative AAPCC Models Using Data from the Continuous Medicare History Sample*. Grant No. 99-C-98526/1. Prepared for Health Care Financing Administration. University Health Policy Research Consortium, Brandeis/Boston University. Waltham, Mass. Aug. 1986.
- Beebe, J., Lubitz, J., and Eggers, P.: Using prior utilization information to determine payments for Medicare enrollees in health maintenance organizations. *Health Care Financing Review*. Vol. 6, No. 3. HCFA Pub. No. 03198. Office of Research and Demonstrations, Health Care Financing Administration. Washington. U.S. Government Printing Office, Spring 1985.
- Brown, R.: Biased Selection in the Medicare Competition Demonstration. Unpublished report. Princeton, N.J. Mathematica Policy Research, Inc. Mar. 1988.
- Eggers, P. W., and Priboda, R.: Pre-enrollment reimbursement patterns of Medicare beneficiaries enrolled in "at-risk" HMO's. *Health Care Financing Review*. Vol. 4, No. 2. HCFA Pub. No. 03146. Office of Research, Demonstrations, and Statistics, Health Care Financing Administration. Washington. U.S. Government Printing Office, Sept. 1982.
- Ellis, R., and Ash, A.: *Refining the Diagnostic Cost Group Model: A Proposed Modification to the AAPCC for HMO Reimbursement*. Grant No. 18-C-98526/1-03-HCFA. Prepared for Health Care Financing Administration. University Health Policy Consortium, Brandeis/Boston University. Waltham, Mass. Feb. 1988.
- Gruenberg, L., Wallack, S. S., and Tompkins, C. P.: Pricing strategies for capitated delivery systems. *Health Care Financing Review*. 1986 Annual Supplement. HCFA Pub. No. 03225. Office of Research and Demonstrations, Health Care Financing Administration. Washington. U.S. Government Printing Office, Dec. 1986.
- Halvorson, G. C., and Stix, P.: Hospital use by new and continuing HMO enrollees. *HMO Practice* 2(3):97-100, May-June 1988.
- Lubitz, J., Beebe, J., and Riley, G.: Improving the Medicare HMO payment formula to deal with biased selection. In Scheffler, R. M., and Rossiter, L. F., eds. *Advances in Health Economics and Health Services Research*. Vol. 6. Greenwich, Conn. JAI Press, 1985.
- McCall, N., and Wai, H.: An analysis of the use of Medicare services by the continuously enrolled aged. *Medical Care* 21(6):567-585, June 1983.
- Newhouse, J. P.: Rate adjusters for Medicare capitation. *Health Care Financing Review*. 1986 Annual Supplement. HCFA Pub. No. 03225. Office of Research and Demonstrations, Health Care Financing Administration. Washington. U.S. Government Printing Office, Dec. 1986.
- Public Health Service and Health Care Financing Administration: *International Classification of Diseases, 9th Revision, Clinical Modification*. DHHS Pub. No. (PHS) 80-1260. Public Health Service. Washington. U.S. Government Printing Office, Sept. 1980.
- Thomas J., and Lichtenstein, R.: Functional health measure for adjusting health maintenance organization capitation rates. *Health Care Financing Review*. Vol. 7, No. 3. HCFA Pub. No. 03222. Office of Research and Demonstrations, Health Care Financing Administration. Washington. U.S. Government Printing Office, Spring 1986.
- Thomas, J., Lichtenstein, R., Wyszewianski, L., and Berki, S.: Increasing Medicare enrollment in HMO's: The need for capitation rates adjusted for health status. *Inquiry* 20(3):227-239, 1983.
- Thomas, J., Lichtenstein, R., and Wyszewianski, L.: Including health status in Medicare's adjusted average per capita cost capitation formula. *Medical Care* 24(3):259-275, 1986.
- Trapnell, G. R., McKusick, D. R., and Genuardi, J. S.: An Evaluation of the Adjusted Average Per Capita Cost (AAPCC) Use in Reimbursing Risk-Basis HMO's Under Medicare. Unpublished Report. Actuarial Research Corporation, 1982.
- U.S. General Accounting Office: *Medicare: Issues Raised by Florida Health Maintenance Organizations*. Report to the Congress. Pub. No. GAO/HRD-86-97. Washington. July 1986.
- Welch, W. P.: Medicare capitation payments to HMOs in light of regression toward the mean in health care costs. In Scheffler, R. M., and Rossiter, L. F., eds. *Advances in Health Economics and Health Services Research*. Vol. 6. Greenwich, Conn. JAI Press, 1985.