Improving the AAPCC With Health-Status Measures From the MCBS

Leonard Gruenberg, Ph.D., Eugenia Kaganova, Ph.D., and Mark C. Hornbrook, Ph.D.

Using data from the 1991 Medicare Current Beneficiary Survey (MCBS), multiple regression-based models predicting 1992 Medicare costs are developed and compared. A comprehensive model incorporating demographic, diagnostic, perceived-health, and disability variables is shown to be stable and to fit the data well over the full range of Medicare-covered annual per capita expenses and for a variety of beneficiary subgroups defined by their health and functional status. This model produces stable unbiased estimates of expenditures on validation samples. A variant of this model is being considered for use in setting Medicare capitation payments for the second phase of the social/health maintenance organization (S/HMO) demonstration.

BACKGROUND AND GOALS

Managed care and risk contracting represent an important policy option to balance the conflicting demands made by U.S. citizens for high-quality medical care on the one hand and cost containment on the other. This potential has led to an unprecedented growth of enrollment in health maintenance organizations (HMOs) and other forms of managed care (Group Health Association of America, 1994), but until recently this growth has been restricted primarily to the population under 65 years of age. Legislative proposals under serious consideration would lead to a much greater reliance on managed care for the Medicare population as well.

If managed care is to be the centerpiece of efforts to enhance quality of care and to control overall Medicare outlays, better risk payment methods are needed. A principal shortcoming of the existing method is its inability to adequately adjust capitation levels for systematic differences in the health status of enrolled groups (Gruenberg, Wallack, and Tompkins, 1986; Newhouse, 1986). Indeed, the findings of the Medicare HMO evaluation strongly suggest that a new payment approach must be found to meet the cost-containment goals of Medicare, and to provide adequate payments to HMOs that seek to provide care to the most sick (Brown et al., 1993).

Current Medicare HMO capitation payments are based on an estimate of 95 percent of the amount HMO enrollees would have cost Medicare had they remained in the local fee-for-service (FFS) sector. This hypothetical amount is calculated using the average adjusted per capita cost (AAPCC) methodology as the average cost in the county of residence for groups of FFS Medicare beneficiaries, classified by age, sex, welfare status, institutional status, and employment status (Kunkel and Powell, 1981). The last of these categories was introduced in 1995 with the inclusion of a set of rate cells for the working aged.

Although a great deal of research has been carried out regarding the impact of health-status measures on health care costs, only very limited experience has been obtained

The research presented in this article was supported by the Health Care Financing Administratiuon (HCFA) under Contract Number 94-0808. Leonard Gruenberg and Eugenia Kaganova are with DataChron Health Systems, Inc. Mark C. Hornbrook is with Kaiser Permanente and Oregon Health Sciences University. The views expressed in this article are those of the authors and do not necessarily represent those of DataChron Health Systems, Inc., Kaiser Permanente, Oregon Health Sciences University, or HCFA.

regarding the payment of real-world programs using a model based on this research. A few demonstration programs have provided a fertile laboratory for testing various alternatives. One such example is the S/HMO, which is testing a modified AAPCC formula that includes special payment rates by Medicare for frail individuals who reside in the community and who are classified as "nursing home certifiable" (NHC).

The study reported here was undertaken in response to a need articulated by participants in the second phase of the S/HMO demonstration for a more comprehensive health-status payment adjustment than the one based on NHC status. In response, the study was designed to examine the variation in health care costs across a broad cross-section of the elderly population, rather than focusing on the cost differences between the small group of those who are classified as NHC and others who are less frail. Although specifically directed toward developing a payment formula for use in the S/HMO, several of the models discussed here could also be considered as potential AAPCC health adjusters for use in setting Medicare's payments to HMOs.

PRIOR RESEARCH

Risk-adjustment research for elderly populations has established the following as significant predictors of future Medicare expenditures:

- Demographic variables included in the current AAPCC: age, sex, institutional status, and work status (Kunkel and Powell, 1981; Gruenberg, Tompkins, and Porell, 1989).
- Prior use of acute-care hospital and physician services (Beebe, Lubitz, and Eggers, 1985; Gruenberg, Tompkins, and Porell, 1989; Freeborn et al., 1990; Anderson et al., 1990).
- Morbidity indicators constructed from data on prior use of acute-care hospital services associated with non-discretionary conditions (Anderson et al., 1986; Ash et al., 1989; Ellis and Ash, 1989).
- Disability-related Medicare eligibility before age 65 (Riley, 1987).
- Self-reported health status, including diagnoses and self-ratings of health (Lichtenstein and Thomas, 1987a, 1987b; Whitmore et al., 1989; Hornbrook and Goodman, 1991; Brown et al., 1993).
- Disability level as measured by limitations in activities of daily living (ADLs) and/or instrumental activities of daily living (IADLs) (Gruenberg and Stuart, 1982; Thomas and Lichtenstein, 1986) or by nursing home certifiability (Gruenberg, Tompkins, and Porell, 1990; Gruenberg, Silva, and Leutz, 1993; Gruenberg, Kaganova, and Rumshiskaya, 1993).
- Medical risk factors, including physiologic measures and results of certain laboratory tests (Howland et al., 1987; Schauffler, Howland, and Cobb, 1986).
- Cost-weighted disease-specific mortality rates (Tolley and Manton, 1984).

The selection of models tested in the current study was influenced generally by the studies cited but was limited to models sought by participants in the third phase of the S/HMO demonstration and by the experience gained during the first phase. In addition to examining models incorporating various measures of health and functional status that were under serious consideration for inclusion in a new payment formula, we included models that incorporate prior-use variables. These are known to have the greatest explanatory power among all of those considered to date (Howland et al., 1987; Kaganova and Gruenberg, 1996) but, as discussed later, are not being seriously considered as risk adjustors for the AAPCC.

THE AAPCC AND THE S/HMO

By assigning a higher capitation payment for nursing home as opposed to community residents, the AAPCC formula creates an incentive counter to the goals of the S/HMO, which seeks to maintain frail individuals in the community with the aid of community support services. To address this problem, a simple modification of the AAPCC was adopted: Frail community residents who were classified as NHC were assigned the same payment rate as that for individuals residing in nursing homes. As data became available from the National Long-Term Care Survey, an improved model was developed and implemented with payment rates associated with an NHC subgroup based on an analysis of Medicare claims (Gruenberg, Tompkins, and Porell, 1990; Gruenberg, Silva, and Leutz, 1993; Gruenberg, Kaganova, and Rumshiskaya, 1993).

The second phase of the S/HMO demonstration focuses attention on developing improvements in geriatric care that it is hoped will lead to better management of chronic conditions and prevention of or delay in the onset of disability among aged beneficiaries. For this reason, a new model of payment is being sought to adequately account for the variations in medical costs over a much broader range of persons than those who are classified as NHC. Variables to be considered for inclusion are: disability level, including ADL and IADL limitations; diagnoses; and self-reported health status. It was considered preferable that measures be based on empirical data that could be found in an individual's medical records. However, the absence of consistent information of this type led to the acceptance of the idea of using self-reported data, and for this reason, Diagnostic Cost Groups (DCGs) were not studied. It is known from the extensive research already cited that social surveys provide an efficient and valid way of collecting health and functional-status information. Therefore, until comparable information is routinely entered into patients' medical records, it may be necessary to rely on this method of data collection. However, as discussed later, further research is necessary to examine issues of reliability before there is widespread use of the models discussed here for payment purposes.

DATA

Data used for this study were obtained from the first round of the MCBS, which was conducted during the last 4 months of 1991. For each person surveyed, we linked these data with summary information obtained from Medicare claims for 1991 and 1992 total Medicare expenditures and utilization rates.

As discussed in detail by Adler (1994), the MCBS is a multipurpose survey of a representative sample of the Medicare population, conducted on a continuous basis by the HCFA Office of the Actuary. The MCBS provides a mechanism that enables HCFA to monitor the performance of the health care system with regard to the elderly population and to trace the impacts of program changes and changing health status over time on patient outcomes and program expenditures.

Interviews are conducted regardless of whether the sample person resides at home or in a long-term care facility, using a questionnaire version appropriate to the setting. Respondents for the MCBS were selected from the Medicare enrollment file to be representative of the Medicare population as a whole and by age group: under age 45, 45-64, 65-69, 70-74, 75-79, 80-84, and 85 or over. Because of interest in their special health care needs, the oldest old (85 years of age or over) and the disabled (under 65 years of age) were oversampled to permit detailed analysis of these subpopulations.

The sample used for estimation included all surveyed individuals who were:

- Alive on January 1, 1992.
- At least 65 years of age at the time of the first round interview.
- Eligible for both Part A and Part B coverage throughout 1991.
- Not a member of a group health plan at any time during 1991.
- Not diagnosed with end stage renal disease, regardless of original reason for Medicare eligibility.

Altogether, there were 9,316 individuals in the sample, with 8,640 residing in the community and 676 in institutions. The study focused on individuals residing in the community for two reasons: First, a number of key variables in the analysis were not available for persons residing in institutions, and second, the question of the reliability of self-reported data needs to be more closely examined for this population, which includes a large proportion of persons who are cognitively impaired. Data for at least one of the independent variables in the study were missing for 48 persons in the community, resulting in a study sample of 8,592.

METHODS

Conceptual Scheme

Building upon methods and concepts that evolved from other recent work in developing risk-adjustment models for the population under 65 years of age (Hornbrook and Goodman, 1995; to be published) and upon a preliminary analysis of the MCBS data (Kaganova and Gruenberg, 1996), we constructed and tested, using multiple regression analysis and a split-sample technique, a comprehensive risk-adjustment model. The model incorporated information on demographic and diagnostic characteristics, perceived health status, and functional-impairment information. Variables included in the model are shown in Table 1. We dropped variables that were found not to have stable, statistically significant coefficients and constructed a single "best" model using only significant variables.

We used the Medicare cost ratio, defined as the ratio of Medicare expenditures for the individual and mean Medicare expenditures as the dependent variable. Using the cost ratio rather than actual costs facilitates simple interpretations of regression coefficients as risk adjustors. For persons who died during the year, annualized Medicare costs were used rather than total costs to ensure that an appropriate accounting is made of the above-average costs of persons who died during the year (Ellis and Ash, 1995). For these individuals, annualized costs were determined as the ratio of total costs and the proportion of the year that the person was alive.

Defining the Risk-Adjustment Models

To gain an understanding of how the comprehensive model compares with models incorporating more limited sets of data, we subdivided the variables in the comprehensive model into the following domains: demographic, diagnostic-health, ADL, and IADL. These separate sets of variables from each domain were forced to enter each specific model in a block, in order to facilitate a clear understanding of the contributions of each domain to the explanation of the overall variance. Models examined included: (1) a demographic model

Table 1
Description of Variables and Coding Used in Study Models

Variable	Description	Coding	
Health	Self-reported health status	0	Excellent
		1	Very good
		2	Good
		3	Fair
		4	Poor
Bathing Dressing Toileting Transferring Eating Walking, ADL	Activities for which individual needs human help and/or supervision; includes bathing, dressing, toileting, transferring, eating, walking	0	No help
		1	Help
Meal Preparation Shopping Light Work Heavy Work Money Phone	Limitations in instrumental activities of daily living: preparing meals, shopping, light housework, heavy housework, money management, using phone	0	No difficulties, or does not do for reason other than health
		1	Has difficulties, but no help required
		2	Has difficulties, uses help
		3	Does not do for health reason
Lifting Walking Reaching Writing Stooping	Having difficulties in lifting or carrying objects as heavy as 10 pounds, walking one-quarter of a mile, reaching overhead, writing, stooping, or kneeling	0	No difficulty at all
		1	A little difficulty
		2	Some difficulty
		3	A lot of difficulty
		4	Not able to do it
Diagnoses	Has a doctor ever told you that you have: Hardening of the arteries Hypertension Myocardial infarction Angina pectoris or coronary heart disease Other heart conditions Stroke or brain hemorrhage Skin cancer Arthritis Rheumatoid arthritis Mental retardation Alzheimer's disease Mental disorder Emphysema, asthma, or chronic obstructive pulmonary disease Complete or partial paralysis Absence or loss of an arm or leg Diabetes Cancer or tumor (includes non-malignant) Parkinson's disease Broken hip Osteoporosis	0	No
		1	Yes
Age 65 Years	Age at the time of the survey minus 65		
Welfare Status	Welfare status	0	Non-welfare
		1	Welfare
Sex	Sex	0	Male
		1	Female

SOURCE: Health Care Financing Administration, Office of the Actuary: Data from the Medicare Current Beneficiary Survey, 1991-92.

that was the regression-based equivalent of the current AAPCC; (2) a health model that includes demographics, self-reported diagnoses, and perceived health; (3) an ADL model that includes ADL variables in addition to demographics; (4) a broader disability model that includes IADL and other impairment variables in addition to ADLs and demographics; and (5) a comprehensive model including demographic, health, ADL, and IADL variables. For the sake of comparison, we also included a sixth model incorporating prior-use variables in addition to all others included in the comprehensive model.

Estimation Methods

We used a weighted least-squares regression to estimate the 1992 Medicare cost ratio as a function of demographic characteristics, health status, functional status, and 1991 utilization rates. The weighting factor was equal to the product of the sample weight and the proportion of the year the sample person was alive. This procedure was necessary to ensure that the regression equation would provide an unbiased estimate of Medicare costs for a person drawn randomly from the Medicare population.

The linear regression method was employed for the sake of simplicity, robustness, and ease of direct interpretation. Other more complex approaches were considered but not used because of the need to provide a model that would be simple to understand by a variety of concerned individuals at each S/HMO site.

To ensure that models were sufficiently robust, we employed methods used by Hornbrook and Goodman (1995). In estimating a comprehensive health- and functional-status model, we carried out repeated regression analyses on 25 randomly selected samples of 50 percent of the overall survey population. We selected variables for the comprehensive model with the aid of an analysis of the significance level of each coefficient in the course of the 25 estimates, selecting variables having significant coefficients, as discussed in greater detail in the next section. Finally, a comprehensive model was estimated on the full sample, using a reduced set of independent variables based on the stability analysis. We applied this method only to the comprehensive model: It would not be appropriate to apply it to the other models because they were constructed by forcing in a specific set of variables.

Evaluating the Performance of Models

We evaluated the performance of the comprehensive model in two ways. First, we compared the model with the other five risk-adjustment models in their ability to fit the data. Second, we validated the model by simulating the effects of using it on another sample and observing the consequences.

Comparison With Other Models

We performed several comparisons, using models developed on the entire sample. First, we compared the adjusted R^2 of the models; ideally, a better model should have a higher R^2. Second, we compared the standard deviation of predicted values. As pointed out by Hornbrook and Goodman (1995), in addition to having a high R^2, a good model should "stretch out" predictions relative to other models; otherwise, it neither explains the wide variation in actual health care costs nor addresses important policy issues.

Third, we compared how well the various models fit across the full range of health care costs by dividing the population into subgroups in order of increasing expected costs, where expected costs were defined as those predicted from the comprehensive model. Fourth, we compared the models' performance in accurately predicting the Medicare costs for various biased subgroups defined in terms of health, functional status, and prior-use variables.

Validating Models

We sought to validate the comprehensive model in order to determine what would occur if it were used as a payment formula, i.e., how accurately it would determine actual FFS expenditures for groups of persons different from the one on which the model was developed. We performed a series of validation tests by determining a payment formula from a 50-percent test sample and using it to estimate what payments would be made to individuals in the remaining 50-percent validation sample. We did this using the 25 randomly drawn samples.

Several tests were carried out. We first examined the extent of errors in prediction by looking at the mean and standard deviation and distribution of errors. Second, we examined the overall fit of actual to predicted costs over the whole range of predicted costs. We did this in two different ways: regressing actual versus predicted costs and following the suggestion of Luft and Rosenkrantz (1993) that the critical issue is how well the model is able to predict for groups of persons. They proposed grouping individuals into 50 equally sized subgroups chosen after sorting individuals in order of their expected costs. We carried out this grouped-R^2 analysis by regressing the average observed expense of each group against the predicted expense.

Finally, we examined whether there were biases in the predictions. This is quite an important test, because if a model is good, it should not lead to any cross-subsidies between groups of persons that are regarded as important from a policy perspective. We examined age- and sex-related biases.

FINDINGS

Comprehensive Model: Stability Analysis

Table 2 shows the results of ordinary least-squares regression analyses carried out on the 25 test samples, where every variable in Table 1 was included as an independent variable and the Medicare cost ratio was used as the dependent variable. This table displays the mean value and standard deviation of the standardized beta coefficients corresponding to each independent variable. Also shown is the mean *t* value, obtained as an average from the 25 regression runs and the number of times (out of 25) that the coefficient was found to be statistically significant at the 0.05 and 0.20 levels. Finally, the table shows the beta values and regression coefficients that were determined from the pooled data (i.e., the full sample).

We found that some variables appeared to be statistically significant in a regression analysis using the entire sample but were found to have coefficients that were unstable as a result of the analysis shown in Table 2. Examples of these variables were reaching and writing. We chose to include variables that were found to be significant at the 0.20 level or higher in 12 or more regression runs, except for the ADL variables. The ADL variables were examined more carefully because of their high degree of intercorrelation. After repeating the analysis shown in Table 2 with dressing and transferring (variables with a very low stability level) removed, we found that the remaining three ADL variables—bathing, toileting, and eating—were all significant by the above definition.

Among IADLs and other impairments, only meal preparation assistance and having difficulty in lifting and in walking two to three blocks were found to be significant. The welfare variable was not found to be significant, indicating that the health-status variables by themselves are sufficient in accounting for differences in costs between Medicare-only and dually eligible populations. Demographic variables (age and sex) were significant, but at a reduced level compared with health-status and disability variables.

Table 2

Derivation of Comprehensive Model: Stability Analysis

Variable	Beta Mean	Beta Standard Deviation	Mean *t* Value	Pooled Beta	Pooled Coefficient	Significance Less Than 0.05	Significance Less Than 0.20
Age 65 Years	0.030	0.013	1.825	0.0304	0.0123	11	17
Sex	-0.024	0.008	-1.482	0.0260	-0.1450	6	15
Welfare Status	0.001	0.011	0.057	0.0025	0.0257	0	3
Bathing	0.049	0.018	2.254	0.0446	0.4441	14	23
Dressing	0.011	0.015	0.491	0.0128	0.1567	1	4
Transferring	-0.010	0.017	-0.474	-0.0100	-0.1210	2	4
Toileting	0.025	0.025	1.216	0.0264	0.4585	8	10
Eating	0.035	0.024	2.094	0.0305	0.7830	12	15
Walking, ADL	-0.017	0.012	-0.852	-0.0130	-0.1190	0	6
Health	0.034	0.010	1.798	0.0325	0.0740	9	21
Artery	0.036	0.008	2.237	0.0327	0.2632	18	24
Hypertention	-0.006	0.010	-0.371	-0.0070	-0.0370	0	3
Myocardial Infarction	0.036	0.014	2.120	0.0362	0.2842	14	20
Angina	0.014	0.014	0.810	0.0135	0.1088	2	9
Other Heart Disease	0.028	0.011	1.721	0.0298	0.1877	10	20
Stroke	-0.003	0.010	-0.202	-0.0070	-0.0670	0	1
Skin Cancer	-0.005	0.012	-0.355	-0.0030	-0.0260	0	4
Cancer	0.020	0.009	1.329	0.0188	0.1358	4	14
Diabetes	0.037	0.010	2.404	0.0342	0.2632	19	24
Rheumatoid Arthritis	-0.013	0.010	-0.800	-0.0120	-0.1110	2	4
Arthritis	-0.003	0.008	-0.205	-0.0030	-0.0170	0	1
Mental Retardation	-0.016	0.006	-1.057	-0.0170	-0.8980	0	8
Alzheimer's Disease	-0.014	0.010	-0.895	-0.0150	-0.4060	1	6
Mental Disorder	-0.008	0.013	-0.543	-0.0050	-0.0890	1	6
Osteoporosis	0.020	0.011	1.283	0.0179	0.1878	4	12
Broken Hip	-0.002	0.012	-0.115	-0.0050	-0.0670	1	3
Parkinson's Disease	0.040	0.015	2.643	0.0372	0.8833	19	24
Emphysema	0.029	0.012	1.889	0.0318	0.2613	14	19
Partial Paralysis	-0.001	0.010	-0.057	0.0020	0.0229	0	1
Amputation	0.030	0.019	1.966	0.0290	0.7717	13	18
Meal Preparation	0.046	0.019	1.940	0.0505	0.2061	11	21
Shopping	-0.003	0.020	-0.109	0.0041	0.0130	0	4
Light Work	0.012	0.018	0.533	0.0094	0.0341	2	4
Heavy Work	0.026	0.013	1.229	0.0214	0.0534	3	11
Money	-0.002	0.013	-0.110	-0.0030	-0.0120	0	2
Phone	-0.002	0.013	-0.093	0.0045	0.0248	1	2
Lifting	0.039	0.018	1.768	0.0426	0.0815	8	20
Walking	0.072	0.013	3.121	0.0671	0.1180	25	25
Reaching	-0.022	0.012	-1.183	-0.0200	-0.0480	5	8
Writing	-0.022	0.018	-1.202	-0.0190	-0.0510	6	11
Stooping	-0.018	0.014	-0.854	-0.0190	-0.0380	0	6
Intercept	—	—	—	—	0.3712	—	—
Adjusted R^2	0.0601	0.0068	—	—	0.0602	—	—

NOTE: ADL is activity of daily living.

SOURCE: Health Care Financing Administration, Office of the Actuary: Data from the Medicare Current Beneficiary Survey, 1991-92.

Comparison of Alternative Risk Models

Specification of Models

Table 3 displays the results of the linear regression analyses for the six alternative risk models. Specification of each model except the comprehensive model was forced. The adjusted R^2 values varied from 0.7 percent for the demographic model to 13.2 percent for the prior-use model. The health, ADL, and disability models have adjusted R^2 values ranging from 3.1 percent for the ADL model to 4.9 percent for the disability model. The comprehensive model has an R^2 (6.0 percent) that is greater than the R^2 of the health, ADL, and disability models but is substantially less than the R^2 of the prior-use model.

Table 3
Regression Statistics for Alternative Risk-Adjustment Models

	Demographic		Health		ADLs		Disability		Comprehensive		Prior-Use	
Variable	Coefficient	Standard Error	Coefficient	Standard Error	Coefficient	Standard Error	Coefficient	Standard Error	Coefficient	Standard Error	Coefficient	Standard Error
Age 65 Years	0.0339	0.0044	0.0293	0.0043	0.0216	0.0044	0.0099	0.0045	0.0126	0.0045	0.0140	0.0043
Sex	-0.0691	0.0592	-0.0569	0.0599	-0.1036	0.0586	-0.2114	0.0592	-0.1393	0.0604	-0.0893	0.0581
Bathing	—	—	—	—	1.1999	0.1232	0.4572	0.1394	0.4272	0.1388	0.1982	0.1340
Toileting	—	—	—	—	0.7397	0.2201	0.3707	0.2218	0.3827	0.2208	0.0896	0.2138
Eating	—	—	—	—	0.8999	0.2947	0.6810	0.2941	0.6434	0.2930	0.3076	0.2817
Lifting	—	—	—	—	—	—	0.1091	0.0266	0.0694	0.0272	0.0412	0.0261
Walking	—	—	—	—	—	—	0.1636	0.0241	0.0989	0.0254	0.0761	0.0244
Meal Preparation	—	—	—	—	—	—	0.2267	0.0567	0.2100	0.0565	0.1540	0.0545
Health	—	—	0.2040	0.0261	—	—	—	—	0.0598	0.0290	0.0142	0.0279
Artery	—	—	0.3259	0.0894	—	—	—	—	0.2518	0.0888	0.1710	0.0853
Myocardial Infarction	—	—	0.3498	0.0892	—	—	—	—	0.3203	0.0885	0.2097	0.0851
Other Heart Disease	—	—	0.2119	0.0709	—	—	—	—	0.1861	0.0703	0.1009	0.0676
Cancer	—	—	0.1274	0.0768	—	—	—	—	0.1327	0.0761	-0.1033	0.0737
Diabetes	—	—	0.3549	0.0833	—	—	—	—	0.2606	0.0830	0.1386	0.0798
Osteoporosis	—	—	0.3071	0.1143	—	—	—	—	0.1680	0.1139	0.0942	0.1094
Parkinson's Disease	—	—	1.0364	0.2520	—	—	—	—	0.7951	0.2502	0.7322	0.2403
Emphysema	—	—	0.3083	0.0887	—	—	—	—	0.2612	0.0883	0.2254	0.0848
Amputation	—	—	1.1518	0.2814	—	—	—	—	0.8362	0.2802	0.6057	0.2692
Part B	—	—	—	—	—	—	—	—	—	—	0.0004	0.0000
Home Health Visits	—	—	—	—	—	—	—	—	—	—	0.0083	0.0017
Number of Inpatient Days	—	—	—	—	—	—	—	—	—	—	0.0057	0.0051
Intercept	0.6948	0.0599	0.0783	0.0722	0.7100	0.0592	0.6158	0.0596	0.3185	0.0740	0.1494	0.0715
Adjusted R^2	0.0068	—	0.0412	—	0.0317	—	0.0493	—	0.0602	—	0.1339	—
F-Ratio	30.22	—	31.77	—	57.32	—	56.64	—	31.56	—	64.23	—
Standard Deviation	0.2245	—	0.5542	—	0.4828	—	0.6016	—	0.6698	—	0.9907	—
Risk Score	0.0726	—	0.8874	—	0.0918	—	0.552	—	0.9128	—	0.9697	—

NOTE: ADL is activity of daily living.

SOURCE: Health Care Financing Administration, Office of the Actuary: Data from the Medicare Current Beneficiary Survey, 1991-92.

Table 3 also provides two additional measures of how successful the alternative models are in accounting for the differences among individuals in predicted costs. The first of these is the standard deviation of predicted costs. A high standard deviation is indicative of a successful model because it indicates that predicted costs vary across a broad range, and that the model is able to capture substantial amounts of differences in expected costs. A low value indicates a poor model because it indicates that expected costs vary over a narrow range. Standard deviations vary considerably among models, ranging from 0.25 for the demographic model to 0.98 for the prior-use model.

Finally, Table 3 provides an additional measure of degree of success of a model, the risk score, which we define to be the percent of persons for whom predicted costs are found to be influenced by characteristics other than age and sex. The risk score as defined in this manner is a convenient tool for measuring the extent to which a model is differentiating the overall population. For example, the current AAPCC has a very low risk score because it only differentiates persons on welfare or in institutions from all others. In the demographic model, only 7.3 percent of persons (i.e., those on welfare) are differentiated from all others on the basis of their characteristics. In contrast, in the prior-use model, 97 percent of persons are differentiated. It can be seen that the comprehensive model (risk score = 91 percent) is quite close in this measure to the prior-use model, although it differs considerably from the latter model in its R^2.

Goodness of Fit of Alternative Models

Table 4 compares the degree of fit of the six models across the full range of expected costs. The comprehensive model was used as the basis for defining quartiles of costs, with individuals grouped according to their expected costs; cutoffs are defined so that equal numbers of persons are included in each group. We also include a fifth group consisting of persons in the uppermost 5 percent of expected costs.

For the comprehensive model, expected costs range from 0.38 in the lowest quartile to 2.9 in the uppermost 5 percent. The fit with actual costs across is quite good, i.e., expected costs are quite close to the observed costs for each subgroup: In no case is the difference more than 5 percent. The fit across this grid appears quite similar to what is found for the better (but unsuitable) prior-use model.

In contrast, the demographic model provides the poorest fit to the data, overestimating the expected costs by more than 125 percent in the lowest quartile and underestimating these costs by more than 57 percent for the highest quartile.

The health model appears to fit the data quite well in the lowest quartile, but it overestimates costs somewhat in the second and third quartiles and substantially underestimates costs in the fourth quartile. The error is quite serious in the uppermost 5 percent subgroup, where costs are underestimated by more than 33 percent. The health model fits the data much better than the demographic model but does not do very well in singling out and estimating the costs of high-cost elderly.

In contrast, the two disability (ADL and disability) models provide better fits than the health model at the high end, i.e., for the upper quartile and uppermost 5 percent of the population. In fact, the better of these, the disability model, underestimates costs in each of these domains by less than 12 percent. However, neither the ADL nor the disability model is as good as the health model at the low end in estimating costs. The ADL model overestimates costs

Table 4

Comparison Between Models of Predicted and Observed Cost Ratios, by Level of Predicted Expenditures and Model Type

Model Type	Quartiles of Expected Cost Ratios				Uppermost 5 Percent of Expenditures
	First	Second	Third	Fourth	
Demographic	0.86	0.97	1.01	1.08	1.15
Health	0.41	0.76	1.10	1.65	2.02
ADLs	0.77	0.85	0.90	1.40	2.40
Disability	0.55	0.70	0.94	1.72	2.71
Comprehensive	0.38	0.63	0.99	1.92	2.97
Prior-Use	0.37	0.63	1.02	1.89	2.98
Observed	0.38	0.65	0.99	1.90	3.12

NOTES: Expected cost ratios were defined using the comprehensive model. ADL is activity of daily living.

SOURCE: Health Care Financing Administration, Office of the Actuary: Data from the Medicare Current Beneficiary Survey, 1991-92.

in the lower quartile by nearly 100 percent. The disability model is somewhat better, but it also overestimates the costs of this group by nearly 50 percent.

We interpret these findings to mean that to adequately fit the data over the full range of costs, it is necessary to include disability measures as well as health and diagnostic ones. The data suggest that a combination of disability, diagnostic, and self-reported health variables are needed to adequately account for severity and instability of medical conditions that have substantial impacts on health care costs. It is not sufficient to identify diagnoses alone or even to include self-reported health variables. We suspect that the same limitations will be found in models using claims data, unless these models are built to incorporate severity measures.

Comparison of Models for Biased Subgroups

Table 5 provides a comparison of how well the models perform in estimating costs for a variety of special subgroups defined according to diagnostic, functioning, and prior-use variables. The subgroups were created by identifying individuals: (1) with one or more heart conditions of the three included in the list; (2) with diabetes; (3) with arthritis; (4) with stroke; (5) needing supervision or hands-on help in walking; (6) needing help in managing money; (7) with a hospitalization in 1991; (8) with Part B costs more than one standard deviation above the mean in 1991 and no hospitalizations in 1991; and (9) whose health is limiting their social activities.

The comprehensive model, the prior-use model, and the health model predict the costs of groups defined according to diagnoses quite accurately. In contrast, the demographic, ADL, and disability models underestimate these costs, with the errors being least for the disability model and most for the demographic model. The extent of underestimation made by the two disability models is greater for heart conditions and diabetes than for arthritis and stroke. We interpret this as meaning that the effect of the latter two are to a large extent explained by disability-level variables in the ADL and disability models.

Overall, the comprehensive model yields predicted costs that are within 5 percent of actual costs for all of the biased subgroups except the two that were chosen on the basis of high prior utilization. For these subgroups, the comprehensive model underestimates actual costs by 31 percent (for the hospitalized subgroup) to 43 percent (for the "high Part B" subgroup). The prior-use model is the only risk model that provides reasonable estimates for these subgroups, but it overestimates costs for these subgroups by 5-7 percent.

It should be noted that groups 3-9 are defined by variables that are not included in the comprehensive model. It is therefore a quite important finding that the comprehensive model fits relatively well for all of these subgroups except those defined in terms of their prior use (i.e., groups 7 and 8).

In contrast to the comprehensive and prior-use models, the ADL, disability, and

Table 5

Comparison Between Models of Predicted and Observed Cost Ratios for Biased Subgroups

Model Type	Heart Condition	Diabetes	Arthritis	Stroke	Help With Walking	Help Managing Money	Hospitalized in 1991	High Part B Costs[1]	Health Limits Social Activities
Demographic	1.00	0.97	0.99	1.04	1.09	1.17	1.03	1.01	1.09
Health	1.36	1.47	1.08	1.37	1.53	1.55	1.30	1.34	1.62
ADLs	1.05	1.07	1.01	1.35	1.76	1.84	1.20	1.23	1.69
Disability	1.14	1.21	1.07	1.53	2.05	2.12	1.33	1.39	2.06
Comprehensive	1.36	1.47	1.09	1.62	2.09	2.15	1.43	1.51	2.13
Prior-Use	1.37	1.47	1.10	1.66	2.10	2.10	2.18	2.80	2.37
Observed	1.36	1.47	1.08	1.57	2.03	2.07	2.08	2.63	2.41
Number of Observations	3,018	1,217	4,038	813	991	713	911	1,653	695

[1] Part B costs in 1991 were greater than $1,864, with no inpatient costs.

NOTE: ADL is activity of daily living.

SOURCE: Health Care Financing Administration, Office of the Actuary: Data from the Medicare Current Beneficiary Survey, 1991-92.

health models display significant errors in estimating costs for various groups. The health model is quite poor in estimating costs for "needs help walking" and "needs help managing money." It is also interesting to note that the health model underestimates costs by nearly one-third for subgroup 9. In comparison, the comprehensive models provide good estimates (error of 12 percent) for these individuals, demonstrating the usefulness of disability measures as a severity measure for risk modeling.

The two disability models provide estimates for subgroups 4 and 5 that are more accurate than those of the health model, but they provide much poorer fits to the data than the health model for the diagnostically defined subgroups 1, 2, and 3.

These analyses of the comparative fit of different models for subgroups that are biased according to either their expected costs or specific health and functional criteria demonstrate that neither health nor disability variables alone are able to adequately fit the observed cost data. Disability variables appear better able to provide accurate cost estimates for high-cost individuals, and health variables facilitate better differentiation among those whose average costs are relatively low. Taken together, the two types of variables provide a reasonably good fit over the whole range of costs and for key subgroups of individuals defined according to health status and functioning.

Validation of Model

We simulated the effects of using the comprehensive model as a payment formula by applying predicted costs derived from each of the 25 test samples to estimate the costs on the 25 validation samples. Table 6 shows the results of this test.

When averaged over the 25 runs, the predicted mean differs insignificantly (by 0.3

Table 6

Validation of Comprehensive Model: Summary of Prediction Results From 25 Validation Samples

Predicted Mean	0.981
Mean Standard Deviation	0.681
Standard Deviation of Means	0.030
Actual Mean	0.978
Standard Deviation of Actual Means	0.028
Predicted Error	-0.003
Maximum Error	0.130
Minimum Error	-0.093
Range	0.223
Intercept of Predicted Regression	0.065
Slope of Predicted Regression	0.933
Adjusted R^2	0.063
Grouped R^2	0.834
Sex-Related Prediction Error	
Males:	0.015
Standard Deviation	0.091
Maximum	0.195
Minimum	-0.139
Range	0.334
Females:	-0.015
Standard Deviation	0.060
Maximum	0.103
Minimum	-0.119
Range	0.222
Age-Related Prediction Error	
Males:	
Age	-0.002
Significant at the 0.05 Level	0
Adjusted R^2	-0.0002
Females:	
Age	0.007
Significant at the 0.05 Level	1
Adjusted R^2	0.0002

SOURCE: Health Care Financing Administration, Office of the Actuary: Data from the Medicare Current Beneficiary Survey, 1991-92.

percent) from the actual mean. Errors range from -9 to 13 percent. The standard deviation of the mean value of actual and predicted costs for the 25 runs is small (about 3 percent). However, there are some significant outliers: for 5 of the 25 samples, the error was greater than 5 percent, and in 2 of the samples the error was 10 percent or greater. It should be noted that the frequency with which errors of this magnitude occur will be reduced if samples of greater size are used (the samples used in the analyses all had 4,300 persons).

A regression analysis of actual versus predicted costs was carried out for of the 25 validation samples. As anticipated, the R^2 (6 percent) was quite similar to what we found estimating the equation for the test sample. The average slope (0.933) and intercept (0.065) indicate that there is a good overall fit between predicted and actual costs, in spite of wide variability found for individual cases. This is further substantiated by the grouped R^2 analysis (Luft and Rosencrantz, 1993), which indicated a high degree of correlation (grouped R^2 = 0.83) between expected and actual costs for subgroups of 86 persons formed in order of increasing expected costs.

To examine whether there were gender-related biases in prediction, we determined, for each of the 25 validation samples, the mean prediction error for males and females. As shown in Table 5, the average difference (.015) is much less than the standard deviation of the difference, indicating that there was not a statistically significant bias. Similarly, a regression of prediction error against age, carried out separately for males and females, showed an insignificant relationship except in 1 out 25 of the regressions and for females only.

Estimating the Effects of Biased Selection

The magnitude of effects of incorporating the comprehensive model into the AAPCC can be appreciated by using the model to predict costs under several alternative case-mix scenarios. The results of such an analysis are shown in Table 7. This table shows the financial effects of various enrollment patterns in an HMO, where enrollment pattern or case-mix is measured using the comprehensive model. Five cost strata are defined in order of increasing costs, as predicted by this model. This table shows that a small change in the number of high-risk enrollees (chosen from individuals in the uppermost 5 percent of expected costs) and a similar offsetting change in the corresponding number of low-risk enrollees (from among those in the lowest 25 per-

Table 7

Effect of Biased Selection on Health Maintenance Organization Profits or Losses

Cost	Total Cost Ratio	Percent Profit or Loss	Percent From Quartile				Uppermost 5 Percent
			First	Second	Third	Fourth	
$4,242	1.13	-13.2	20	25	25	20	10
$4,143	1.11	-10.6	21	25	25	20	9
$4,043	1.08	-7.9	22	25	25	20	8
$3,944	1.05	-5.3	23	25	25	20	7
$3,845	1.03	-2.6	24	25	25	20	6
$3,746	1.00	0.0	25	25	25	20	5
$3,647	0.97	2.6	26	25	25	20	4
$3,548	0.95	5.3	27	25	25	20	3
$3,448	0.92	7.9	28	25	25	20	2
$3,349	0.89	10.6	29	25	25	20	1
$3,250	0.87	13.2	30	25	25	20	0

SOURCE: Health Care Financing Administration, Office of the Actuary: Data from the Medicare Current Beneficiary Survey, 1991-92.

cent of expected costs) leads to a large change in total costs. For example, each 1-percent change of this type for an enrolled population of 10,000 would lead to an increase or decrease in costs of 2.6 percent that would translate into a gain or loss of $1 million. The use of the proposed model could protect both health plans and the Medicare program against these unpredictable losses or gains.

SUMMARY AND CONCLUSIONS

The first major conclusion is that our primary hypothesis is supported, namely, that direct health-status measures (diagnoses, perceived health status, and functional level) and indirect health-status measures (demographic characteristics) all make, to a significant degree, independent contributions to forecasting health care costs. We recommend that these factors be incorporated in the payment formula for the S/HMO and be considered for inclusion in the AAPCC to improve the currently inadequate risk-adjustment model. We need to stop thinking of the individual items mentioned as substitute measures of risk; rather they are complementary measures because they are all needed to capture the underlying structure of risk in the heterogenous population of Medicare aged beneficiaries. It is especially important to note the critical contribution of disability in accounting for Medicare costs, especially for the oldest old, and our analysis suggests that revisions to the AAPCC based on diagnostic measures alone are not likely to be satisfactory.

Second, although the comprehensive model is not nearly as successful as the prior-use model (which incorporates prior year's utilization and cost measures) in explaining the variance in health care costs, it does not suffer from the well-known drawbacks of relying on prior-use measures in a payment formula. These include, most importantly, the potentially inappropriately high payment amount assigned by the prior-use model to a patient cared for by an "elaborate provider" (McClure, 1984) and the underpayment for persons or populations who are underserved. In addition, the use of utilization in one year in the HMO as a modifier of payments in the next year would create perverse incentives for the HMO (Ash et al., 1989).

Several authors (Newhouse, 1986; Gruenberg, Wallack, and Tompkins, 1986) have argued that it may be preferable for Medicare to base payments to HMOs on a blended payment method rather than on a full capitation method. In a blended payment method, the total payments to an

HMO are based in part on a prospectively determined amount and in part on actual utilization experience. Our analysis suggests that it is possible to develop a prediction model (and a payment formula) that is robust, that fits reasonably well across the full range of health costs, and that provides reliable cost estimates for a variety of clinically defined subgroups, without the need to incorporate utilization measures. We believe that the comprehensive model would address the major concerns regarding full capitation and would thus obviate the necessity of moving toward a blended payment approach.

A key issue that needs to be addressed is whether social survey techniques can be used to collect all of the required data. There is considerable reluctance on the part of payers and providers to rely on self-report for payment purposes. To some extent, objections to the use of self-reports could be overcome by further analyses of data from the MCBS, and the strength of having these data available on an ongoing basis should not be overlooked. However, it is also necessary to test whether the procedures can be implemented in a manner that ensures that real-world use of self-reported data does not lead to gaming on the part of HMOs, and that individuals' responses to the survey will not be influenced by the knowledge that payment rates to HMOs will be affected by these responses.

In weighing the potential for biases that may be introduced by incorporating measures proposed here in the AAPCC, one must weigh theses biases against those that would occur in an alternative approach. It is already well known that there are considerable biases in the current AAPCC that lead to a considerably inequitable pattern of expenditures. We believe that the current biases in the AAPCC are more serious than those that could occur if a self-report-based model were implemented. In addition, alternative morbidity-based AAPCC models that are under consideration would rely on medical records data. These alternatives are also subject to the potential for gaming, because the data required to implement them would be under the direct control of the HMO. Moreover, a morbidity-based model using ambulatory care records would fail to account for unmet need.

Further refinements of the model need to be explored. Interaction effects between health and functioning can, to some extent, be examined using MCBS data, but a full examination of this issue would require a much larger data set.

Questions regarding the effects of biases in the model resulting from the use of survey data and methods for compensating for these biases can be addressed by further research, including: examining whether there are biases in the model when proxies are used in collecting survey data; studying effects on the model of non-response; making possible improvements of modeling techniques by considering two-part models; developing methods for taking account of geographic variations; examining the effect of sociocultural variations on the risk-adjustment models; and studying the stability of the model with respect to the time of measurement of health care costs in relation to the time of the survey. A validation of the model by testing it out using a later round of MCBS data can also be undertaken to establish the extent of stability over time in the model's predictions.

REFERENCES

Adler, G.: A Profile of the Medicare Current Beneficiary Survey. *Health Care Financing Review* 15(4):153-164, Summer 1994.

Anderson, G., Cantor, J., Steinberg, E., and Holloway, J.: Capitation Pricing: Adjusting for Prior Utilization and Physician Discretion. *Health Care Financing Review* 8(2):27-34, 1986.

Anderson, G., Steinberg, E., Powe, N., et al.: Setting Payment Rates for Capitated Systems: a Comparison of Various Alternatives. *Inquiry* 27:225-233, 1990.

Ash, A., Porell, F., Gruenberg, L., et al.: Adjusting Medicare Capitation Payments Using Prior Hospitalization Data. *Health Care Financing Review* 10(4):17-29, 1989.

Beebe, J., Lubitz, J., and Eggers, P.: Using Prior Utilization to Determine Payments for Medicare Enrollees in Health Maintenance Organizations. *Health Care Financing Review* 6(3):27-38, 1985.

Brown, R., Clement, D., Hill, J., et al.: Do Health Maintenance Organizations Work for Medicare? *Health Care Financing Review* 13(1):7-24, 1993.

Ellis, R., and Ash, A.: *The Continuous-Update Diagnostic Cost Group Model*. Final Report. Grant No. 18-C-98526/1-03. Boston. Health Policy Research Consortium, Boston University, 1989.

Ellis, R., and Ash, A.: Refinements to the Diagnostic Cost Group (DCG) Model. *Inquiry* 32(4):418-429, 1995.

Freeborn, D., Pope, C., Mullooly, J., and McFarland, B.: Consistently High Users of Medical Care Among the Elderly. *Medical Care* 28(6):527-540, 1990.

Group Health Association of America: *Annual Industry Survey*. Washington, DC. 1994.

Gruenberg, L., Kaganova, E., and Rumshiskaya, A.: *Updating the Social HMO AAPCC*. Final Report to the Health Care Financing Administration under Purchase Order Number HCFA-94-0105. Cambridge, MA. DataChron Health Systems, Inc., 1993.

Gruenberg, L., Silva, A., and Leutz, W.: *An Improved Medicare Payment System for the Social/HMO*. Final Report to the Health Care Financing Administration under Cooperative Agreement Number 99-C-98526/1-07. Cambridge, MA. DataChron Health Systems, Inc., 1993.

Gruenberg, L., and Stuart, N.: *A Health Status-Based AAPCC: The Disability-Level Approach*. Unpublished Working Paper. Waltham, MA. Brandeis University, Heller Graduate School, 1982.

Gruenberg, L., Tompkins, C., and Porell, F.: The Health Status Patterns of the Elderly: Implications for Setting Medicare Payments to HMOs. *Advances in Health Economics and Health Services Research* 10:41-73, 1989.

Gruenberg, L., Tompkins, C., and Porell, F.: *Capitation Rates for the Frail Elderly*. Final Report to the Health Care Financing Administration. Waltham MA. Bigel Institute Health Policy, Brandeis University, 1990.

Gruenberg, L., Wallack, S., and Tompkins, C.: Modifying Capitation Payments: Rewarding Efficiency, Not Selection. *Health Care Financing Review* Annual Supplement:35-44, 1986.

Hornbrook, M., and Goodman, M.: Adjusting Employer Health Benefits Contributions to Reflect Risk. *Advances in Health Economics and Health Services Research* 12:41-76, 1991.

Hornbrook, M., and Goodman, M.: Assessing Relative Health Plan Risk With the RAND-36 Health Survey. *Inquiry* 32(1):56-74, 1995.

Hornbrook, M., and Goodman, M.: Chronic Disease, Functional Health Status, and Demographics: A Multi-Dimensional Approach To Risk Adjustment. *Health Services Research*. To be published.

Howland, J., Stokes, J., Crane, S., and Belanger, A.: Adjusting Capitation Using Chronic Disease Risk Factors: A Preliminary Study. *Health Care Financing Review* 9(2):15-23, 1987.

Kaganova, E., and Gruenberg, L.: *An Examination of Health Status Adjustments to the AAPCC Using the MCBS*. Final Report to the Health Care Financing Administration under Purchase Order Number HCFA-94-0808. Cambridge, MA. DataChron Health Systems, Inc., 1996.

Kunkel, S., and Powell, C.: The Adjusted Average per Capita Cost Under Risk Contracts With Providers of Health Care. *Society of Actuaries Transactions* 33:221-230, 1981.

Lichtenstein, R., and Thomas, J.: A Comparison of Self-Reported Measures of Perceived Health and Functional Health in an Elderly Population. *Journal of Community Health* 12:213-230, 1987a.

Lichtenstein, R., and Thomas, J.: Including a Measure of Health Status in Medicare's Health Maintenance Organization Capitation Formula: Reliability and Issues. *Medical Care* 25(2):100-110, 1987b.

Luft, H., and Rosenkrantz, S.: *Changing the Focus: Applying Subscriber Level Expenditure Models to Groups of Enrollees*. Paper presented at the Annual Meeting of the Association for Health Services Research. Washington, DC. 1993.

McClure, W.: On the Research Status of Risk-Adjusted Capitation Rates. *Inquiry* 21:205-213, 1984.

Newhouse, J.: Rate Adjustors for Medicare Under Capitation. *Health Care Financing Review* Annual Supplement:45-55, 1986.

Riley, G.: *Revised Underwriting Factors for the AAPCC Incorporating Former Entitlement Under the Social Security Disability Program in Place of Institutionalization Status*. Health Care Financing Administration, Office of Research and Demonstrations, 1987.

Schauffler, H., Howland, J., and Cobb, J.: Using Chronic Care Disease Factors to Adjust Medicare Capitation Payments. *Health Care Financing Review* 14(1):85-95, 1986.

Thomas, J.W., and Lichtenstein R.L.: Functional Health Measures for Adjusting Health Maintenance Organization Capitation Rates. *Health Care Financing Review* 7:85-95, 1986.

Tolley, H., and Manton, K.: Assessing Health Care Costs in the Elderly. *Transactions of the Society of Actuaries* 36:579-598, 1984.

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

Reprint Requests: Leonard Gruenberg, Ph.D., President, DataChron Health Systems, Inc., 763 Massachusetts Avenue, Suite 7, Cambridge, Massachusetts 02139.