Evaluating Alternative Risk Adjusters for Medicare

Gregory C. Pope, M.S., Killard W. Adamache, Ph.D., Edith G. Walsh, Ph.D., and Rezaul K. Khandker, Ph.D.

In this study the authors use 3 years of the Medicare Current Beneficiary Survey (MCBS) to evaluate alternative demographic, survey, and claims-based risk adjusters for Medicare capitation payment. The survey health-status models have three to four times the predictive power of the demographic models. The risk-adjustment model derived from claims diagnoses has 75-percent greater predictive power than a comprehensive survey model. No single model predicts average expenditures well for all beneficiary subgroups of interest, suggesting a combined model may be appropriate. More data are needed to obtain stable estimates of model parameters. Advantages and disadvantages of alternative risk adjusters are discussed.

INTRODUCTION

Although it is a goal of the Medicare program to enroll more of its beneficiaries in managed care programs, as of January 1997, only about 13 percent were enrolled in health maintenance organizations (HMOs) or competitive medical plans (Health Care Financing Administration, 1997). Greater managed care enrollment has the potential to reduce the growth in Medicare expenditures and improve the quality of care Medicare beneficiaries receive and is consistent with private sector trends. Medicare's managed care enrollment shortfall has been attributed in part to the inadequacy of its current payment formula for HMOs—the adjusted average per capita cost (AAPCC)—in accounting for expenditure differences among beneficiaries. The AAPCC considers only sociodemographic factors (age, sex, private insurance coverage, welfare status, and institutional status), location (county of residence), and reason for Medicare eligibility (aged, disabled, or having end stage renal disease [ESRD]). Many studies have shown that the AAPCC factors inadequately predict medical expenditures, creating inequities among HMOs that enroll healthier or sicker beneficiaries, and also large financial incentives for HMOs to try to attract healthier beneficiaries. Numerous proposals have been advanced to incorporate additional factors into the AAPCC, such as health status, prior medical care use, diagnoses, and medical risk factors. The additional factors that could be added to the AAPCC generally are available from two sources: surveys of beneficiaries or medical claims data. If Medicare adopts a revised capitated payment methodology, it is likely to incorporate factors collected from either surveys or claims or both. It is thus of interest to examine the merits of alternative survey and claims-based risk adjusters for the Medicare population.

The largest survey of the Medicare population currently available is the MCBS. This study employs 3 years of the MCBS and associated claims data to evaluate alternative survey and claims-based risk adjusters on a common sample. With the requirement that Medicare claims for ambulatory patients contain diagnostic

codes, there have recently been substantial innovations in claims-based risk adjusters (Ellis et al., 1996; Weiner et al., 1996; Kronick et al., 1996). One of the latest generation of claims-based adjusters is included in our evaluation.

Both survey adjusters and claims-based adjusters have been extensively studied in the past and continue to be a subject of intensive current research (Ellis et al., 1996). But very few studies have compared both survey and claims-based measures. For example, Gruenberg, Kaganova, and Hornbrook (1996) used the first 2 years of the MCBS to analyze survey risk adjusters. That study did not include any claims-based measures, was limited to the elderly, non-institutionalized Medicare population, and evaluated models using a different methodology than this study. Hornbrook and Goodman (1995) assessed the RAND-36 Health Survey in a population predominantly under the age of 65 enrolled in a large prepaid group-practice HMO in the Pacific Northwest but did not consider claims-based measures. Ellis et al. (1996), Weiner et al. (1996), and Kronick et al. (1996) all analyzed only claims-based measures. Fowles et al. (1996) and Fowles, Weiner, and Knutson (1994) are the studies most closely related to this one in that they evaluated both survey and claims-based risk adjusters. But the specific survey and claims models studied were different, and the population studied was very different—a predominantly younger, employed sample of enrollees in a Minnesota HMO versus the nationally representative Medicare elderly and disabled sample here.

DATA

The MCBS is an ongoing, multipurpose survey of a nationally representative sample of the Medicare population, including both aged and disabled enrollees who live in the community or are institutionalized. A key advantage of the MCBS for use in this study is that it links survey responses to Medicare administrative claims, enabling us to compare the performance of survey- and claims-based risk adjusters in predicting actual Medicare payments. Also, survey responses allow performance of alternative risk adjusters to be compared for groups—such as supplemental insurance status—not identifiable from Medicare administrative records.

The MCBS is a population-based survey that employs a panel design. Each round of the MCBS includes survey data and Medicare claims data collected for the same individuals. The claims data include diagnostic codes and the Medicare expenditures associated with each claim. For this study, we used data from rounds 1, 4, and 7. Survey data were collected from 12,674 persons in round 1 (September-December, 1991). During round 4, approximately 1 year later, 10,388 of these persons completed their followup interviews. An additional 1,995 persons were added to the sample and interviewed in round 4 to account for attrition (because of death, relocation, or non-response). Round 7 interviews were completed in fall 1993, with 10,936 individuals who had participated in the earlier rounds, as well as 1,927 sample replacements. Response rates for all three rounds were between 87 and 94 percent.

We used elements of the survey data from each round in several ways: to describe the sample for each year (rounds 1, 4, and 7), as independent variables in survey-based estimation (round 1) and validation models (round 4), and to define validation groups (round 7). We used the claims data to develop claims-based diagnostic groups that were used as independent variables in claims-based estimation and validation models. Total annual

Medicare expenditures from the subsequent round served as the dependent variable in all models. That is, we used round 1 (1991) survey data and claims-based diagnostic groups as predictors of round 4 (1992) expenditures, and round 4 (1992) survey and claims-based diagnostic groups as predictors of round 7 (1993) expenditures. We also used total annual Medicare expenditures as an independent variable in prior-use models, where expenditures in a given year were used as predictors of expenditures in the subsequent year.

STUDY DESIGN

The goal of this study is to evaluate the performance of alternative risk-adjustment models. It is important to do this on a validation sample that differs from the estimation sample used to establish parameters for the models. If a single sample is used for both estimation and validation, the explanatory power of the models will be overstated in general, and certain models may be unduly favored relative to others. Avoiding "overfitting" a single sample is especially important when the available sample is small, as is true with the MCBS.

Typically, a "split-sample" design is employed, where models are estimated on a portion of a cross-sectional sample, then validated on the remainder of the sample. The relatively small sample sizes available from the MCBS in any 1 year make this cross-sectional split-sample design unattractive. Using it would result in highly unstable parameter estimates and validation results. Instead, we exploit the longitudinal nature of the MCBS by estimating our models using 1991 survey and claims data to predict 1992 expenditures. We then validate the models using 1992 survey and claims data to predict 1993 expenditures. Two years of data are necessary for both estimation and validation because we are evaluating prospective risk-adjustment models that use beneficiary characteristics to predict expenditures in the subsequent year. That is, using the regression parameters from the 1991-92 sample, the validation model uses 1992 Medicare beneficiary characteristics to predict 1993 Medicare expenditures. Then the predicted expenditures for 1993 are compared with the actual expenditures for 1993.

Because the MCBS is a panel survey, many of the same individuals are present in both our estimation and validation samples. Because of the lack of full independence of the validation sample, our validation results probably slightly overstate the predictive power of the risk-adjustment models. However, the year-to-year correlation of medical expenditures is small (Ellis et al., 1996), so this bias should not be large, and even a partially independent validation is better than relying solely on the estimation results to compare the models. We emphasize performance of models on our validation rather than the estimation sample.

SAMPLE SELECTION

To create the 1991-92 estimation file, we eliminated from the sample individuals who died before January 1, 1992, lived outside the United States, were entitled for the ESRD program, were not eligible for both Part A and Part B of Medicare for all of 1992, or had missing values on any analysis variable. Respondents who were enrolled at any time during 1991 or 1992 in a managed care organization were excluded because they have no Medicare claims for their period of managed-care enrollment. Total Medicare payments were constructed by summing total Part A and Part B Medicare payments available from the MCBS. Data on 1992 Medicare payments for individuals who did not respond to round 4 of the survey (i.e., those lost to

attrition from round 1) were obtained from HCFA. Parallel methods were used to construct the validation file that contains beneficiary characteristics reported in 1992 and expenditures for 1993. As before, expenditure data for those lost to attrition between 1992 and 1993 were obtained from special files provided by HCFA.

The final estimation sample consisted of 10,893 individuals for 1991-92, and the final validation sample consisted of 10,532 individuals for 1992-93. Table 1 shows estimation sample characteristics overall and for three important subsamples: the non-institutionalized elderly, the non-institutionalized disabled, and the institutionalized elderly or disabled. (Validation sample characteristics are similar.) Although the three subsamples differ substantially on demographic and health characteristics, we included all of them in our analysis to obtain the greatest generality and information on model properties for different populations and to maximize limited sample sizes. Future work could examine risk-adjustment models specialized for segments of the Medicare population, for example, aged versus disabled.

METHODOLOGICAL CONSIDERATIONS

The MCBS employs a complex sample design, with both stratification and clustering. Although means and regression coefficients are not affected by stratification or cluster sampling, their standard errors are. Comparisons were performed between ordinary least squares (OLS) regressions weighted by the MCBS sampling weights and regressions corrected for the MCBS' complex sampling design using a specialized software package, SUDAAN. The standard errors in the SUDAAN regressions were generally smaller than in the weighted OLS regressions. This affected

the statistical significance of some variables. We adopted a conservative and more computationally convenient approach using weighted OLS regressions.

To get the correct predicted average payments for all beneficiaries in a payment class, including those who died, we adjusted the MCBS sampling weights using a process described in Ellis and Ash (1995). First, total payments were annualized by dividing by the fraction of the year (measured in months) each beneficiary was alive. Then we adjusted the MCBS sample weights by multiplying them by the fraction of the year that the person was eligible for coverage. This process of annualizing and reweighting observations results in unbiased estimates of the average and total payments for a group in which individuals are eligible for different fractions of the year.

Specialized statistical models have been developed to account for the unusual distributional properties of medical expenditures, namely, extreme skewness with a small proportion of people accounting for a large proportion of expenditures, and a substantial proportion of people with no expenditures in a year. The "two-part" model is the best known of these specialized models (Duan et al., 1983). However, the two-part model suffers from the disadvantages of being computationally burdensome and more difficult to interpret. We estimated weighted OLS and two-part variants of a few of our risk-adjustment models, then compared their predictive power on the validation sample. The predictive power of the OLS and two-part models was about the same (more details are available in Pope et al., 1997). Thus, we based our analysis on the computationally more convenient weighted OLS models.

We also considered the effect of payment outliers on our analyses (Pope et al., 1997). With sample sizes available from the MCBS, outliers have some effect on

Table 1

Estimation Sample Characteristics, Overall and by Subsample[1]

Variable	Full Sample	Non-Institutionalized Elderly	Non-Institutionalized Disabled	Institutionalized (Elderly and Disabled)
Observations	10,893	8,526	1,622	745
Mean 1992 Expenditures	$3,795	$3,752	$3,583	$4,951
Hierarchical Coexisting Condition Scores[2]	1.00	0.96	0.94	1.70
Age		Percent		
0-64 Years	8.7	NA	100.0	11.9
65-74 Years	49.7	56.4	NA	13.1
75-84 Years	31.5	34.6	NA	30.0
85 Years and Over	10.1	9.1	NA	45.0
Male	41.1	39.9	61.1	29.5
Medicaid	11.6	7.3	33.1	53.7
Self-Rated Health Status				
Poor	10.3	7.7	35.8	13.8
Fair	21.0	19.1	30.1	41.8
Good	29.7	30.5	20.1	32.2
Very Good	23.0	25.1	8.8	9.9
Excellent	15.9	17.6	5.2	2.4
Functional Status				
5-6 ADLs	7.0	3.6	10.3	63.2
3-4 ADLs	9.3	8.0	17.6	17.8
1-2 ADLs	24.9	24.4	33.8	19.1
IADLs Only	14.2	13.8	26.7	0.0
None	44.7	50.3	11.5	0.0
Chronic Conditions				
Arteriosclerosis	13.4	12.7	11.0	29.8
Heart Attack	13.6	13.4	16.9	9.8
Angina	13.3	12.8	15.4	19.6
Other Heart Conditions	24.5	23.8	25.2	35.0
Hypertension	47.4	48.7	40.4	35.8
Stroke	9.7	8.6	12.7	24.8
High-Cost Cancer[3]	2.8	2.9	3.2	0.4
Low-Cost Cancer[3]	12.8	13.3	10.3	7.7
Skin Cancer	13.3	14.5	5.5	4.5
Diabetes	14.4	14.2	15.7	16.0
Rheumatoid Arthritis	10.0	9.7	15.2	6.1
Osteoarthritis	44.8	46.0	38.0	34.1
Osteoporosis	7.2	6.9	6.7	14.9
Mental Retardation	2.1	0.2	16.7	10.4
Alzheimer's Disease	3.0	1.1	1.0	42.1
Mental Disorders	5.5	2.3	29.0	24.1
Hip Fracture	4.4	3.6	3.5	20.6
Parkinson's Disease	1.6	1.3	1.3	6.9
COPD	12.7	12.1	19.9	12.4
Partial Paralysis	7.6	5.7	21.7	17.8
Amputation of Arm/Leg	1.3	1.1	2.0	4.1
Lost Urine More Than Once per Week	10.4	7.9	11.2	55.3

[1] Weighted by MCBS sampling weights.

[2] Predicted expenditure scores based on 1991 claims diagnoses, with 1.00 representing average predicted expenditures.

[3] The following cancers were classified as high-cost: lung, ovarian, stomach, kidney, brain, throat, and head. All other cancers included in the MCBS, except for skin (which has its own category), were classified as low-cost. The assignments to high- and low-cost cancer were derived by Ellis et al. (1996).

NOTES: ADL is activity of daily living. IADL is instrumental activity of daily living. COPD is chronic obstructive pulmonary disease. MCBS is Medicare Current Beneficiary Survey. NA is not applicable.

SOURCE: Data from the Medicare Current Beneficiary Survey; analysis by Pope et al., 1997.

both estimated coefficients and R^2 values. However, the relative ranking of alternative models, our main interest, is largely unaffected by top-coding[1] expenditures at an upper threshold such as $50,000 or $25,000. Some have suggested that top-coded expenditures be analyzed to simulate reinsurance purchased by HMOs. However, Medicare does not currently have an outlier policy for HMO risk contracts, so participating HMOs are at risk for these costs. We analyzed untransformed expenditures.

RISK-ADJUSTMENT MODELS

We developed nine risk-adjustment models using the information available from the MCBS. They are:

- Demographic.
- Self-rated health status.
- Self-reported chronic conditions.
- Functional status.
- Short form (SF)-36 simulation.
- Comprehensive survey.
- Claims diagnoses.
- Claims diagnoses plus survey.
- Prior use.

Each model was estimated using 1991 (round 1) survey characteristics or claims data to predict 1992 Medicare program expenditures. Estimates are shown in Table 2.

All models include categorical variables for age and sex. The categorical variable for the 0-64 age category allows the intercept to shift for the disabled-entitled population, all of whom are under age 65. In addition to a basic age/sex model (not shown in Table 2), a second demographic model incorporating additional factors used in Medicare's current AAPCC methodology (Medicaid enrollment status and institutionalization) was estimated.

A model is defined for each of three major domains of survey health-status measures so that the properties of each measure can be isolated. These are self-rated health (also called "general" or "perceived" health status), self-reported chronic conditions, and functional status.[2] These measures, along with the measures of having limitations in walking two to three blocks or lifting 10 pounds, are also combined into a comprehensive survey model to analyze their joint properties. The MCBS measures self-rated health status using the standard question "In general, compared to other people your age, would you say that your health is excellent, very good, good, fair, or poor?"

Our functional-status variable is a count of the number of activities of daily living (ADLs) that a respondent reports difficulty with or inability performing, with an additional category for difficulty or inability with at least one instrumental activity of daily living (IADL) but not with any ADL. We considered individual ADLs and IADLs as well as the count scale but found the latter to be more stable across samples. In the MCBS, ADLs are bathing, dressing, walking, toileting, transferring in and out of chairs, and eating. IADLs are light housework, heavy housework, meal preparation, using the telephone, managing money, and shopping for personal items.[3] The premise of the scale is that impairments in more domains indicate greater disability.

---

[1] Top-coding establishes an upper threshold for a variable and sets all greater values at that level. For example, expenditures of $50,001 or greater are set equal to $50,000, if that is the chosen threshold. Unlike truncating, topcoding keeps the observations in the model while decreasing the outlier effect.

[2] The MCBS collects additional health risk factors that are not reported here. Refer to Pope et al. (1997) for analyses of these variables and additional analysis of disability and social functioning variables.

[3] The nursing home sample members were only asked about shopping for personal items, use of the telephone, and money management. Beneficiaries in the nursing home portion of the MCBS were coded as "having difficulty" with light and heavy housework and meal preparation.

We chose to define functional impairment based on a report of difficulty or inability to perform for a health reason, rather than on reported receipt of help, for several reasons. First, our purpose is to evaluate risk adjusters for the general Medicare population, unlike payment methodologies for demonstrations targeted to the smaller segment of the population that is at risk of institutionalization. We believe the salient cut for Medicare risk adjustment is between those who are healthy and those who are impaired to any degree. Risk adjusters based on this distinction will be predictive for larger segments of the Medicare population than those distinguishing only beneficiaries with a high degree of functional impairment. Close to 50 percent of the sample report difficulty with at least one ADL, although slightly less than 25 percent report receipt of help with at least one ADL. Second, the wording of the MCBS, which asks about receipt of help but not need for help, is dependent on supply of and access to help, not just health status. We believe it is inappropriate for a payment model to use a measure of impairment that is confounded by availability of help and the provision of care. Finally, use of report of difficulty will focus health plans on identifying persons with difficulty and addressing the underlying health problems.

We also developed a model simulating four of the eight scales from the SF-36 to provide a comparison to our survey models and to other work done using the SF-36 for risk adjustment (Hornbrook and Goodman, 1995). The SF-36 is a widely used 36-item health-status questionnaire developed to measure outcomes of medical care (Ware, 1993; Ware and Sherbourne, 1992). Although the MCBS and SF-36 questions differ in details of wording, we were able to construct simulated scores for the physical-functioning, general-health,

social-functioning, and role-physical scales.[4] These are four of the five scales that Hornbrook and Goodman (1995) found to be predictors of medical costs. Our SF-36-like scales have not been tested for equivalence to the actual SF-36 scales.

Our physical-functioning scale is composed of responses to questions about difficulty in lifting or carrying 10 pounds, walking two to three blocks, bending, stooping or kneeling, heavy housework, and bathing or dressing. Our general health scale is derived from the MCBS self-rated health-status question, scored according to SF-36 guidelines. Our social-functioning scale is based on an MCBS question about restrictions in social activities due to health. Finally, our role-physical scale uses difficulties in performing IADLs as a proxy for SF-36 questions about limitation in "usual" activities.

The MCBS asks respondents if a doctor has ever told them that they have any of a list of specific medical conditions (heart attack, diabetes, cancer, etc.). We measured each of these with a dichotomous yes/no variable.[5] We also included the response to the question "Have you lost urine beyond control in the past 12 months (more than once per week)?" in the list of conditions. Conditions that were not positive and statistically significant in a preliminary model were eliminated from the final model, which is shown in Table 2. This same list of selected conditions was used for the comprehensive survey model combining self-rated health, functional status, and self-reported chronic conditions.

For comparison to the survey models, we included a claims-based diagnostic model using the diagnoses recorded on

---

[4] Refer to Pope et al. (1997) for a description of the crosswalk between the MCBS and the SF-36 and for the scoring methodology.

[5] The MCBS collects history of cancer by anatomical site (throat, lung, etc.). Based on Ellis et al., 1996, we divided cancer sites into high-cost (e.g., lung cancer) and low-cost (e.g., breast cancer).

Table 2
Regression Estimates of Alternative Risk-Adjustment Models

Variable	AAPCC-Like[1]		Self-Rated Heath Status		Self-Reported Chronic Conditions		Functional Status	
	Coefficient	Standard Error	Coefficient	Standard Error	Coefficient	Standard Error	Coefficient	Standard Error
Intercept	3,033	***196	1,708	***323	1,851	***213	1,978	***223
Age (65-74 Years Omitted)								
0-64 Years	-40	439	-1,132	***439	-158	***427	-1,125	**440
75-84 Years	924	***263	839	***261	693	***261	447	*264
85 Years and Over	2,260	***416	2,138	***398	1,860	***404	721	*416
Male	226	238	224	236	127	242	477	**237
Medicaid	1,399	***388	—	—	—	—	—	—
Institutionalized	-190	584	—	—	—	—	—	—
SF-36 Like Scales[2]								
Physical Functioning (0-100)	—	—	—	—	—	—	—	—
General Health (1-5)	—	—	—	—	—	—	—	—
Social Functioning (1-4)	—	—	—	—	—	—	—	—
Role-Physical (2-8)	—	—	—	—	—	—	—	—
Self-Rated Health Status (Excellent Omitted)								
Poor	—	—	5,197	***470	—	—	—	—
Fair	—	—	2,702	***383	—	—	—	—
Good	—	—	1,335	***356	—	—	—	—
Very Good	—	—	486	373	—	—	—	—
Functional Status (No Limitations Omitted)								
5-6 ADLs	—	—	—	—	—	—	5,589	***488
3-4 ADLs	—	—	—	—	—	—	3,517	***428
1-2 ADLs	—	—	—	—	—	—	2,537	***294
IADLs Only	—	—	—	—	—	—	1,129	***357
Difficulty Walking 2-3 Blocks	—	—	—	—	—	—	—	—
Difficulty Lifting	—	—	—	—	—	—	—	—
Chronic Conditions								
Arteriosclerosis	—	—	—	—	913	**354	—	—
Heart Attack	—	—	—	—	1,626	**358	—	—
Other Heart Conditions	—	—	—	—	904	**281	—	—
High-Cost Cancer	—	—	—	—	1,402	**689	—	—
Diabetes	—	—	—	—	1,774	**329	—	—
Osteoporosis	—	—	—	—	1,312	**453	—	—
Parkinson's Disease	—	—	—	—	2,446	**923	—	—
COPD	—	—	—	—	1,452	**347	—	—
Partial Paralysis	—	—	—	—	1,174	**444	—	—
Amputation of Arm/Leg	—	—	—	—	3,989	**1,021	—	—
Lost Urine More Than Once per Week	—	—	—	—	1,744	**391	—	—
Hierarchical Coexisting Condition Score[3]	—	—	—	—	—	—	—	—
Previous Medicare Payments	—	—	—	—	—	—	—	—
Adjusted R^2	0.0045	—	0.0178	—	0.0240	—	0.0200	—
F-Ratio	***9.08	—	***25.62	—	****18.85	—	***28.71	—
Observations	10,893	—	10,893	—	10,893	—	10,893	—

*Statistically significant at 10-percent level.

**Statistically significant at 5-percent level.

***Statistically significant at 1-percent level.

[1] Except for employment status and geographic location, this model includes the factors used in Medicare's AAPCC payment system.

[2] Number in parentheses are the range of each scale, with a larger value indicating better health.

[3] Predicted expenditure score (mean = 1.000) based on claims diagnoses, age, and sex.

Table 2—Continued
Regression Estimates of Alternative Risk-Adjustment Models

SF-36-Like		Comprehensive Survey		DCG-HCC (Claims Diagnoses)		Claims Diagnoses Plus Survey Measures		Prior Use	
Coefficient	Standard Error	Coefficient	Standard Error	Coefficient	Standard Error	Coefficient	Standard Error	Coefficient	Standard Error
10,876	***937	1,093	***331	-207	177	-546	*301	2,190	***191
-1,818	***435	-1,304	***449	—	—	—	—	284	413
339	264	433	264	—	—	—	—	691	***255
484	428	974	**422	—	—	—	—	1,847	***389
692	***240	392	246	—	—	—	—	38	230
—	—	—	—	—	—	—	—	—	—
—	—	—	—	—	—	—	—	—	—
-34	***6	—	—	—	—	—	—	—	—
-350	***116	—	—	—	—	—	—	—	—
-658	***162	—	—	—	—	—	—	—	—
-190	*113	—	—	—	—	—	—	—	—
—	—	1,890	***542	—	—	1139	**496	—	—
—	—	579	423	—	—	331	399	—	—
—	—	235	368	—	—	37	355	—	—
—	—	121	373	—	—	42	363	—	—
—	—	2,326	***624	—	—	1,773	***501	—	—
—	—	732	526	—	—	897	**437	—	—
—	—	764	**368	—	—	961	***300	—	—
—	—	59	376	—	—	73	348	—	—
—	—	581	*320	—	—	—	—	—	—
—	—	786	***298	—	—	—	—	—	—
—	—	551	354	—	—	—	—	—	—
—	—	1,381	***359	—	—	—	—	—	—
—	—	590	**283	—	—	—	—	—	—
—	—	1,226	*687	—	—	—	—	—	—
—	—	1,257	***333	—	—	—	—	—	—
—	—	787	*456	—	—	—	—	—	—
—	—	1,527	*925	—	—	—	—	—	—
—	—	908	***352	—	—	—	—	—	—
—	—	347	453	—	—	—	—	—	—
—	—	3,155	***1,021	—	—	—	—	—	—
—	—	680	415	—	—	—	—	—	—
—	—	—	—	4,014	***138	3,688	***148	—	—
—	—	—	—	—	—	—	—	0.446	***0.017
0.0281	—	0.0322	—	0.0716	—	0.0744	—	0.0601	—-
***39.98	—	***15.47	—	***841.01	—	***98.37	—	***140.29	—-
10,845	—	10,892	—	10,893	—	10,893	—	10,893	—-

NOTES: Dependent variable is annualized 1992 Medicare payments. AAPCC is adjusted average per capita cost. ADLs is activities of daily living. IADLs is instrumental activities of daily living. COPD is chronic obstructive pulmonary disease. DCG is diagnostic cost group. HCC is hierarchical coexisting condition.

SOURCE: Data from the Medicare Current Beneficiary Survey; analysis by Pope et al., 1997.

MCBS-linked hospital and physician claims to predict future expenditures. This is the hierarchical coexisting conditions (HCC) variant of the diagnostic cost group (DCG) model, described in Ellis et al. (1996). The MCBS sample is too small to properly establish parameters for the DCG-HCC model. Instead, parameter estimates from Ellis et al. (1996) were combined with MCBS claims diagnoses and age and sex to produce a predicted expenditure "score," normalized to have a mean of 1.00. The results of regressing 1992 expenditures on the DCG-HCC predicted expenditure score are shown in Table 2.[6] In addition to the DCG-HCC model itself, we added survey measures to the claims-based score to form an additional model. Only self-rated health and functional status are added to the DCG-HCC score, because the self-reported chronic conditions largely duplicate the claims diagnoses already incorporated into the DCG-HCC score. This "combined" model allows us to evaluate the incremental contribution of survey variables to the claims-based model. Finally, we developed a model based on prior use of medical services for comparison to the survey and claims-diagnosis models. In the prior-use model, total Medicare payments for an individual in the previous year are used to predict current year payments. For example, medical service use in 1991 is used to predict expenditures in 1992.

Table 2 presents estimates of the alternative risk-adjustment models on the 1991-92 MCBS sample. The magnitudes and patterns of the coefficient estimates are plausible. For example, poorer self-rated health or more ADL limitations are associated with greater future Medicare payments. The intercept is large and the coef-

ficients in the SF-36 simulation are negative, because higher scores are associated with better health in the SF-36 scales. Most coefficients are statistically significant but have large confidence intervals. Our focus, however, is on model performance on the 1992-93 validation sample, to which we now turn.

VALIDATION RESULTS

We estimated the models using 1991 beneficiary characteristics to predict 1992 Medicare payments. We then applied the estimated parameters to beneficiary characteristics reported in 1992 to predict 1993 Medicare payments. The 1993 Medicare payments were deflated to have the same mean as 1992 payments.[7] Predicted 1993 payments were compared with actual 1993 payments to judge the models' predictive power. Two measures of predictive accuracy were computed for each estimation model and validation group, one for individuals and one for groups. The individual measure is the R^2 statistic, defined as the proportion of variation in actual 1993 payments accounted for by predicted 1993 payments. The group measure is the predictive ratio, defined as the ratio of the aggregate predicted payments for a group of beneficiaries divided by the aggregate actual payments for this group. Each of these measures—R^2 and predictive ratio—is examined for the overall sample and for various subgroups that are of interest.

PREDICTIVE ACCURACY FOR INDIVIDUALS

Table 3 shows R^2 values for the overall validation sample and selected subgroups. Models are arrayed in increasing order of

---

[6] Because the DCG-HCC score incorporates age and sex, these variables were not entered separately in the regression reported in Table 2 that includes the HCC score.

[7] For the DCG-HCC model, 1993 expenditures were predicted from DCG-HCC scores computed using the parameters estimated by Ellis et al. (1996), not the results shown in Table 2.

Table 3
Explanatory Power (R^2) of Alternative Risk-Adjustment Models, by Validation Subgroup

Validation Group	Age-Sex	AAPCC-Like	Functional Status	Self-Reported Chronic	Self-Rated Health	SF-36-Like	Prior Use	Comprehensive Survey	DCG-HCC (Claims Diagnosis)	Combined Survey and Claims
					Percent					
Overall Sample	0.70	0.93	2.52	2.74	3.11	4.05	4.13	4.18	7.27	7.85
Age										
0-64 Years	0.0	0.6	0.1	2.1	0.1	1.6	18.5	2.6	14.4	13
65-74 Years	-0.0	0.5	2.6	3.0	3.4	4.4	1.6	5.0	8.6	9.3
75-84 Years	-0.0	-0.1	1.8	1.7	2.1	3.0	2.9	2.9	4.1	4.8
85 Years and Over	-0.0	0.3	0.5	0.4	1.9	2.6	1.8	1.6	5.1	5.6
Sex										
Female	0.8	1.1	3.1	3.1	3.6	4.8	5.4	4.8	8.2	8.8
Male	0.5	0.7	1.7	2.3	2.5	3.2	2.6	3.5	6.1	6.7
Medicare Status										
Elderly	0.7	0.9	2.7	2.8	3.3	4.2	3.0	4.3	6.7	7.4
Disabled	0.0	0.6	0.1	2.1	0.1	1.6	18.5	2.6	14.4	13.8
Institutional Status										
Non-Institutionalized	0.7	0.9	2.5	2.7	3.1	3.9	4.3	4.2	7.1	7.8
Institutionalized	-2.9	-2.6	-1.2	-1.6	-0.9	2.3	-2.7	-0.1	5.6	5.2
Self-Rated Health Status										
Poor	-5.6	-5.1	-0.8	-1.4	1.1	2.7	5.6	2.0	6.7	8.1
Fair	-0.7	-0.6	-0.3	1.3	0.5	1.1	1.8	1.7	6.8	6.8
Good	0.6	0.6	0.9	1.9	0.4	1.7	0.8	2.3	5.7	5.9
Very Good	-0.6	-0.4	0.5	-0.7	0.8	0.9	0.9	0.7	0.6	1.3
Excellent	-10.3	-9.4	-1.4	-4.0	1.0	2.4	-5.6	1.9	-0.1	2.5
Functional Status										
5-6 ADLs	-7.5	-6.3	0.0	0.2	-0.3	3.5	1.6	3.1	7.3	9.0
3-4 ADLs	-2.0	-2.0	0.6	1.2	1.3	2.9	4.2	2.7	7.2	7.8
1-2 ADLs	-0.8	-0.6	0.3	0.3	0.5	1.8	3.3	1.7	4.9	5.4
IADLs Only	-0.3	-0.2	0.1	0.4	1.0	1.8	0.1	1.9	6.3	6.5
None	-1.0	-0.8	0.3	0.6	1.3	1.2	1.0	1.7	3.5	4.1
Elderly Helped With 3 or More ADLs	-7.1	-6.5	-0.8	-2.2	-0.9	2.2	0.9	1.0	5.8	7.0

See notes at end of table.

Table 3—Continued

Explanatory Power (R^2) of Alternative Risk-Adjustment Models, by Validation Subgroup

Validation Group	Age-Sex	AAPCC-Like	Functional Status	Self-Reported Chronic	Self-Rated Health	SF-36-Like	Prior Use	Compre-hensive Survey	DCG-HCC (Claims Diagnosis)	Combined Survey and Claims
					Percent					
Expenditures, 1992										
First Quintile (Lowest)	-4.5	-4.6	-1.5	-3.4	-1.9	-1.2	-0.4	-1.2	-0.1	0.6
Second Quintile	-2.4	-2.5	-1.9	-2.0	-1.7	-1.8	-0.1	-1.4	0.0	-0.2
Third Quintile	-0.7	-0.4	-1.2	-5.9	-0.5	-0.2	1.2	-3.8	-1.2	-0.7
Fourth Quintile	0.2	0.7	1.0	0.6	1.6	1.8	-0.0	1.8	3.1	3.6
Fifth Quintile (Highest)	-9.3	-8.9	-5.9	-3.1	-4.7	-2.6	-3.0	-1.4	4.5	5.5
Top 5 Percent	-21.3	-20.7	-14.9	-13.0	-14.0	-9.9	-15.7	-9.3	2.9	4.4
Hospital Admissions, 1992										
No Admissions	0.2	-0.1	1.5	1.2	0.2	2.0	2.8	1.6	3.3	3.8
1 Admission	-3.4	-3.5	-1.0	-1.7	-0.4	-0.0	-2.9	0.9	3.5	4.1
2 or More Admissions	-17.8	-18.1	-12.0	-13.9	-9.9	-9.1	-9.1	-7.7	1.9	2.8

NOTES: R^2 is negative if mean payment is a better predictor than risk-adjusted payments for a subgroup. AAPCC is adjusted average per capita cost. ADLs is activities of daily living. IADLs is instrumental activities of daily living. DCG is diagnostic cost group. HCC is hierarchical coexisting condition.

SOURCE: Data from the Medicare Current Beneficiary Survey; analysis by Pope et al., 1997.

predictive power from left to right. Note that negative R^2 values occur. This happens when mean payment is a better predictor than risk-adjusted payments for a subgroup.

As expected, the demographic models—age/sex and AAPCC—are the least predictive models, explaining less than 1 percent of the variance in actual payments in the overall sample. Even the least powerful model incorporating health status, the functional-status model, triples the predictive power of the demographic models. The comprehensive survey and SF-36-like models, which measure multiple dimensions of health status, are the most predictive survey models. But the relatively modest gain in R^2 over the single-dimension survey models indicates considerable redundancy among the survey measures. The prior-use model is more powerful than the individual survey measures but less powerful than the comprehensive survey model and the claims-based model.

The claims-diagnosis-based DCG-HCC model is more predictive than any of the survey models, with an R^2 that exceeds that of the comprehensive survey model by about 75 percent. Adding survey functional and self-rated health status to the DCG-HCC model results in a gain in predictive power of 0.58 percentage points, or about 8 percent. Thus, these survey variables appear to contain only a limited amount of information relevant to predicting expenditure differences among individuals not already incorporated into the DCG-HCC model. But the incremental explanatory power of the survey variables may be important in "getting payment right" for certain policy-relevant subgroups.

The predictive advantage of the claims-based model differs greatly by aged versus disabled subsamples. For the disabled the DCG-HCC model is clearly more predictive, with an R^2 of 14.4 percent versus only 2.6 percent for the comprehensive survey model. Among the elderly the DCG-HCC model is still better by more than 50 percent, but the gap in R^2 is narrowed to 6.7 percent versus 4.3 percent. In addition, survey variables add more predictive power at the margin to claims diagnoses among the elderly. Prior use also does dramatically better among the disabled than the elderly. Expenditures among the disabled are more predictable and are relatively strongly related to past expenditures and to diagnoses recorded on medical claims, making the disabled particularly suitable for claims-based risk adjustment.

Consistent with its greater overall predictive power, the claims-based DCG-HCC model predicts better among individuals in most subgroups than the survey or prior-use models. The DCG-HCC model tends to do better at predicting expenditure differences among individuals in poorer health than among those in better health. This is consistent with its emphasis on multiple, serious, high-cost conditions (Ellis et al., 1996). Adding survey measures (in the combined survey/claims model) improves the ability of the DCG-HCC model to predict expenditure differences among individuals in relatively good health, as well as differences among individuals in the worst health. Quite often, the models are less successful at predicting expenditure differences among individuals in a subgroup than among the overall sample. This is because much of the models' overall explanatory power results from predicting differences among groups.

PREDICTIVE ACCURACY FOR GROUPS

Table 4 reports predictive ratios for the overall validation sample and validation subgroups. A predictive ratio closer to 1.00 indicates better prediction. A predictive

Table 4

Predictive Ratios for Alternative Risk-Adjustment Models, by Validation Subgroup

Validation Group	Age-Sex	AAPCC-Like	Functional Status	Self-Reported Chronic Conditions	Self-Rated Health Status	SF-36 Like	Prior Use	Comprehensive Survey	DCG-HCC (Claims Diagnosis)	Combined Survey and Claims
Overall Sample (normalized)[1]	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Age										
0-64 Years	1.04	1.04	0.97	1.01	1.01	0.96	1.08	0.97	1.05	**1.19
65-74 Years	*1.07	*1.07	1.05	1.06	1.05	0.95	1.05	1.05	1.01	0.96
75-84 Years	0.92	*0.92	0.94	0.93	0.94	1.02	0.93	0.94	0.98	0.98
85 Years and Over	1.00	0.99	1.03	1.01	1.01	1.05	1.01	1.03	0.99	1.05
Sex										
Female	1.04	1.04	1.05	1.05	1.04	1.04	1.05	1.05	1.01	1.02
Male	0.94	0.94	0.94	*0.93	0.95	0.95	*0.93	0.93	0.99	0.97
Medicare Status										
Elderly	1.00	1.00	1.00	1.00	1.00	1.00	0.99	1.00	1.00	0.98
Disabled	1.04	1.04	0.97	1.01	1.01	0.96	1.08	0.97	1.05	**1.19
Institutional Status										
Non-institutionalized	1.02	1.02	0.98	1.00	1.01	0.99	1.00	0.99	0.99	0.98
Institutionalized	***0.77	**0.82	***1.21	1.03	*0.86	*1.15	0.96	**1.17	1.12	***1.27
Self-Rated Health Status										
Poor	***0.49	***0.50	***0.69	***0.68	0.94	*0.88	***0.72	0.93	***0.76	0.94
Fair	***0.77	***0.78	0.92	0.93	1.00	1.06	**0.87	1.02	0.96	1.03
Good	0.97	0.97	0.96	0.96	0.93	0.94	0.95	0.93	0.98	0.94
Very Good	***1.42	***1.40	***1.20	***1.22	1.05	1.02	***1.23	1.06	***1.13	1.02
Excellent	***2.15	***2.11	***1.69	***1.64	***1.31	***1.25	***1.77	***1.28	***1.47	***1.25
Functional Status										
5-6 ADLs	***0.58	***0.61	1.10	***0.82	***0.76	1.02	**0.85	1.08	*0.88	1.08
3-4 ADLs	***0.66	***0.67	0.97	**0.84	**0.83	1.06	**0.83	0.94	*0.85	0.95
1-2 ADLs	***0.79	***0.80	1.03	**0.88	**0.89	0.99	***0.85	1.03	**0.90	1.03
IADLs Only	1.04	1.05	0.95	1.04	1.04	1.04	1.05	0.95	1.04	0.96
None	***1.44	***1.41	0.97	***1.21	***1.23	0.97	***1.22	0.99	***1.16	0.98
Elderly Helped With 3 or More ADLs	***0.55	***0.58	***0.71	0.91	***0.82	0.94	***0.84	0.97	*0.88	1.00
Chronic Conditions										
Any Chronic Condition	*0.95	*0.95	0.97	0.99	0.97	0.98	0.98	1.00	0.99	1.00
Arteriosclerosis	***0.74	***0.74	**0.87	1.08	**0.86	0.93	***0.85	1.09	0.95	1.00
Heart Attack	***0.61	***0.61	***0.68	0.96	***0.71	***0.74	***0.77	0.97	**0.86	*0.89
Angina	***0.62	***0.63	***0.70	*0.88	***0.73	***0.76	***0.75	*0.90	**0.86	*0.89

See notes at end of table.

Table 4—Continued

Predictive Ratios for Alternative Risk-Adjustment Models, by Validation Subgroup

Validation Group	Age-Sex	AAPCC-Like	Functional Status	Self-Reported Chronic Conditions	Self-Rated Health Status	SF-36 Like	Prior Use	Comprehensive Survey	DCG-HCC (Claims Diagnosis)	Combined Survey and Claims
Other Heart Conditions	***0.70	***0.71	***0.78	0.95	***0.80	***0.83	***0.83	0.96	0.94	0.96
Hypertension	***0.89	***0.90	***0.93	0.97	0.94	0.96	**0.93	0.99	0.96	0.97
Stroke	***0.73	***0.75	0.95	1.01	**0.87	1.01	0.93	1.06	1.01	*1.09
High-Cost Cancer	*0.80	*0.81	0.87	1.18	0.91	0.95	0.96	*1.20	1.15	1.17
Low-Cost Cancer	*0.89	*0.88	0.93	0.96	0.93	0.94	1.00	0.96	1.04	1.05
Skin Cancer	0.97	0.95	0.97	1.01	0.98	0.97	0.99	1.00	1.02	1.00
Diabetes	***0.62	***0.64	***0.71	0.95	***0.73	***0.77	***0.73	0.96	0.95	0.98
Rheumatoid Arthritis	***0.77	***0.78	0.91	0.92	0.90	0.98	***0.84	1.00	**0.87	0.94
Osteoarthritis	*0.93	*0.94	0.99	1.01	0.99	1.03	0.97	1.04	0.98	1.00
Osteoporosis	0.95	0.96	**1.17	***1.39	1.08	***1.22	*1.14	***1.40	1.10	***1.20
Mental Retardation	**1.33	***1.47	***1.39	1.19	1.10	1.21	**1.31	1.09	***1.35	***1.47
Dementia	0.94	0.98	***1.39	**1.21	1.12	***1.42	1.08	***1.38	**1.21	***1.38
Mental Disorders	*0.81	*0.84	0.89	0.88	0.88	0.93	0.91	0.91	0.93	1.01
Hip Fracture	0.92	0.94	**1.20	1.07	1.00	**1.20	1.12	**1.19	1.13	***1.24
Parkinson's Disease	0.76	0.77	1.04	1.32	0.93	1.08	0.87	1.34	1.07	1.18
Chronic Obstructive Pulmonary Disease	***0.71	***0.71	***0.80	1.03	***0.84	**0.87	***0.81	1.03	0.95	0.99
Partial Paralysis	***0.73	***0.74	0.98	1.10	*0.86	1.02	0.93	1.10	0.97	1.08
Amputation of Arm/Leg	***0.57	***0.59	0.77	*1.26	**0.71	0.85	0.86	*1.27	0.88	0.96
Lost Urine More Than Once per Week	***0.70	***0.72	0.96	1.06	***0.83	0.99	**0.88	1.07	*0.91	1.02
Expenditures, 1991										
First Quintile (Lowest)	***1.99	***1.97	***1.77	***1.64	***1.79	***1.66	***1.34	***1.55	0.97	0.92
Second Quintile	***1.68	***1.66	***1.54	***1.51	***1.55	***1.47	*1.16	***1.44	1.13	1.10
Third Quintile	***1.39	***1.39	***1.37	***1.39	***1.36	***1.34	1.01	***1.37	***1.22	***1.22
Fourth Quintile	0.92	0.92	0.95	0.99	0.96	0.97	***0.79	0.99	1.02	1.03
Fifth Quintile (Highest)	***0.46	***0.47	***0.55	***0.56	***0.54	***0.60	0.97	***0.61	***0.88	***0.90
Top 5 Percent	***0.31	***0.31	***0.42	***0.41	***0.39	***0.47	***1.25	***0.47	***0.86	*0.88
Hospital Admissions, 1991										
No Admissions	***1.27	***1.27	***1.23	***1.22	***1.23	***1.20	0.97	***1.19	1.02	1.02
1 Admission	***0.63	***0.63	***0.72	***0.73	***0.70	***0.77	1.04	***0.78	1.04	1.06
2 or More Admissions	***0.33	***0.34	***0.41	***0.44	***0.41	***0.47	1.07	***0.49	**0.86	*0.87
Supplemental Insurance[2]										
Medicaid	***0.74	0.91	*0.90	***0.86	***0.83	0.93	**0.88	0.93	0.93	1.01
Medicare Only	1.15	1.10	1.17	1.15	*1.22	*1.22	1.06	*1.20	1.01	1.07
Other Supplemental Coverage	*1.05	1.01	1.00	1.02	1.02	0.99	1.03	0.99	1.02	0.99

See notes at end of table.

Table 4—Continued

Predictive Ratios for Alternative Risk-Adjustment Models, by Validation Subgroup

Validation Group	Age-Sex	AAPCC-Like	Functional Status	Self-Reported Chronic Conditions	Self-Rated Health Status	SF-36 Like	Prior Use	Compre-hensive Survey	DCG-HCC (Claims Diagnosis)	Combined Survey and Claims
Income										
$15,000 or Less	**0.93	0.95	0.96	0.95	0.96	0.98	0.95	0.98	0.96	0.98
$15,001-$25,000	1.09	1.05	1.05	1.06	1.07	1.04	1.04	1.04	1.04	1.01
More Than $25,000	***1.19	***1.15	1.09	*1.11	1.07	1.03	**1.14	1.03	*1.11	1.04
Not Reported	*0.67	*0.66	1.01	0.88	0.79	0.98	0.82	1.00	0.92	1.04
Education										
Less Than 12 Years	**0.92	0.94	0.95	0.95	0.98	1.00	*0.93	0.99	0.96	0.99
12 Years	*1.08	1.06	1.04	1.05	1.05	1.02	1.06	1.03	1.04	1.02
More Than 12 Years	1.09	1.06	1.02	1.02	0.98	0.96	1.08	0.96	1.00	0.96
Not Reported	0.86	0.92	*1.23	1.06	0.96	*1.20	0.98	1.18	1.15	**1.30
Race										
White	1.01	1.00	1.01	1.01	1.00	1.00	1.01	1.00	1.01	1.00
Black	0.93	1.00	0.97	0.93	1.01	1.03	0.94	1.00	0.97	1.01
Other	0.82	0.88	0.87	0.86	0.89	0.92	0.89	0.92	0.89	0.92
Living Status										
Living Alone	0.97	0.97	*0.91	0.94	0.94	**0.90	0.95	*0.91	0.95	0.93
Living With Spouse	***0.85	**0.88	0.99	0.93	**0.89	1.00	0.95	0.99	0.98	1.05
Living With Others	***1.11	***1.09	1.06	**1.08	**1.10	1.07	1.06	1.07	1.05	1.01

*Predictive ratio is significantly different from 1 at the 0.10 level.

**Predictive ratio is significantly different from 1 at the 0.05 level.

***Predictive ratio is significantly different from 1 at the 0.01 level.

[1] Predictive ratios were normalized by dividing them by the predictive ratio of the overall sample.

[2] Other supplemental coverage includes individually purchased (IP), employer sponsored (ES), both IP and ES, and public coverage other than Medicaid, as well as private plans held by a small number of working elderly.

NOTES: AAPCC is adjusted average per capita cost. ADLs is activities of daily living. IADLs is instrumental activities of daily living. DCG is diagnostic cost groups. HCC is hierarchical coexisting conditions.

SOURCE: Data from the Medicare Current Beneficiary Survey; analysis by Pope et al., 1997.

ratio greater than 1.00 indicates overprediction, whereas a predictive ratio less than 1.00 indicates underprediction. The predictive ratios are subject to random variation because of the limited MCBS sample size. Accordingly, statistical significance of the predictive ratios (difference from 1.00) is indicated in Table 4. Among the large number of predictive ratios in Table 4, some will be statistically significant by chance. To avoid predictive ratios different than 1.00 merely due to random error in predicting overall mean 1993 expenditures, we normalized the predictive ratios by dividing by the predictive ratio for the overall sample.

Although one would expect predictive ratios closer to 1.00 for validation groups that are defined by elements of the predictive model, these predictive ratios are still of interest to determine reliability, because estimation and validation are on different years. Moreover, there is no guarantee that models comprising multiple variables will predict well for validation groups defined by a single variable.

Dually Eligible Beneficiaries

Persons dually eligible for Medicare and Medicaid (identified by "Medicaid" under "Supplemental Insurance" in Table 4) are a group of particular interest to State and Federal policymakers. Only the combined claims and survey model predicts this group's expenditures accurately. All other models underpredict for this group, although the underpredictions of the AAPCC-like, SF-36-like, comprehensive survey and DCG-HCC models are not statistically significant. Larger sample sizes are needed to confirm these findings.

Institutionalized Persons

The demographic models underpredict spending for institutionalized persons, but many of the health-status models overpredict spending. This indicates that nursing home residents are absolutely more expensive but are less expensive to Medicare than community residents with the same diagnoses or functional status. Institutionalized beneficiaries may be less expensive, controlling for diagnoses or functional status, because of the substitution of nursing home care for the acute care services covered by Medicare.

Self-Rated Health and Functional Status

Not surprisingly, survey models including these variables predict well across validation groups. Other demographic, survey, prior-use, and claims-based models are less successful. Comparison of the claims-diagnosis DCG-HCC model with the combined survey/claims model indicates that survey variables can improve predictions of the claims model across health- and functional-status groups.

Elderly Receiving Help with ADLs

We included an "elderly receiving help with three or more ADLs" validation group because policymakers and providers are interested in the ability of risk-adjustment methodologies to pay accurately for the more functionally impaired elderly at risk of institutionalization. Only the combined model, including both the claims-based DCG-HCC and the survey measures, predicts accurately for this group, with the

comprehensive survey model a close second. The self-reported chronic conditions and the SF-36-like model predict reasonably well, even though there are no functional-status measures in the first, and very little functional-status information in the second. All other models substantially underpredict expenditures for this group.

Prior Utilization

The models using claims information predict payments better for persons with varying levels of prior-year payments than the survey or demographic models. The DCG-HCC model underpredicts by only 14 percent among the highest 5-percent prior-year spenders. For the lowest quintile, the DCG-HCC model underpredicts by only 3 percent, versus 64-79 percent overprediction by the survey variables. The survey variables do only somewhat better than demographics across prior-year expenditure quintiles. The combined survey and claims model does not do much better than the DCG-HCC model alone, that is, survey measures do not add much for predicting across prior-expenditure quintiles. For prior-year hospital admission categories, the prior-use model does best, with the two models including the DCG-HCC score a close second.

Chronically Ill Persons

Across groups of people reporting chronic conditions,[8] the models using diagnostic information (self-reported chronic conditions, comprehensive survey, DCG-HCC, and combined survey/claims) show the fewest statistically significant under- or overpredictions. The SF-36-like model also does well, despite utilizing no diagnostic

information. The most important chronic-condition indicators to include in survey models appear to be heart disease, diabetes, and chronic lung disease. All models overpredict expenditures for the mentally retarded, and all except the demographic models overpredict for dementia. This could be attributable to underprovision of care to these groups or substitution of Medicaid for Medicare expenditures.

Demographic Groups

All the models predict mean expenditures reasonably well across income, education, and race groups, with the exception of the age/sex model. Predicted spending is in general higher than actual spending for beneficiaries who live with individuals other than their spouse, which could reflect substitution of nursing home care for acute medical care or underservice to these beneficiaries. Living alone, on the other hand, has a tendency to raise actual compared with predicted expenditures.

CONCLUSIONS

No one risk-adjustment model is best on all empirical criteria considered in this article. The claims-diagnosis-based DCG-HCC model has greater overall predictive power than the survey models and predicts average expenditures as well as or more accurately for most of the validation subgroups we considered. It appears to be the best single model empirically. However, for certain subgroups (for example, the elderly receiving help with ADLs) it does not appear to predict expenditures as accurately as certain of the survey models. No model predicts uniformly well for all groups. Thus, which model is preferred depends in part on what relative weight policymakers put on "getting payment right" for different subgroups of Medicare beneficiaries.

---

[8] Chronic-condition groups could be defined using either survey self-reports or diagnoses recorded on claims. We use self-reports.

Practical and administrative considerations are also important in evaluating claims versus survey adjusters. Claims adjusters require encounter data systems, which are expensive and time-consuming to develop, although useful for a variety of purposes once implemented. Moreover, claims adjusters are sensitive to intentional and unintentional variations in diagnostic coding (e.g., "upcoding"). Surveys have lower startup costs and are available more immediately but are expensive and burdensome to conduct on an ongoing basis.[9] They suffer from non-response and biased and inaccurate responses (e.g., what does self-rated health mean from someone with dementia?). Providers may be able to influence survey responses (e.g., by "prescribing" disability), and beneficiaries may respond strategically once they realize that provider reimbursement depends on their survey answers. Survey responses may deviate from "objective" criteria along sociodemographic or regional lines and are difficult to audit or verify.

Adding survey variables to a claims-based model such as the DCG-HCC increases overall explanatory power and improves predictions for key subgroups such as the elderly receiving help and dually eligible beneficiaries. But a combined model requires obtaining both survey and encounter data, which might be prohibitively expensive. Also, overpredictions for certain diagnoses (osteoporosis, hip fracture, dementia) are increased. Combined models warrant more research as a means of combining diagnoses from claims with severity/disability information (subjective health, functional status) from survey responses into a single powerful model.

Substantial redundancy exists among the various survey adjusters. Their combined explanatory power is much less than the sum of their individual explanatory power. Nevertheless, independent dimensions of health status are measured by the different survey variables. A multidimensional survey model such as our comprehensive survey model or the SF-36 simulation is necessary to predict expenditures well across the range of subgroups. The disadvantage of multidimensional survey models is that survey instruments must be longer, increasing survey expense and respondent burden and lowering response rates.

Although multiple domains of health status need to be surveyed, redundancy implies that some pruning of questions based on other criteria is possible and desirable. Other desirable characteristics for risk adjusters include resistance to manipulations by providers or beneficiaries, objectivity, reliability, parsimony, and face validity. In our opinion certain survey variables rank higher on these criteria than others. We would place chronic conditions (diabetes, heart disease) and physical functioning ("Can you walk two blocks?") higher on this scale, and social functioning ("Has your health interfered with your social activities?") and self-rated health ("Is your health excellent, good, fair, or poor?") lower. Others might disagree with our assessment. More research and practical experience are needed on pertinent aspects of survey adjusters other than predictive power.

Even the best survey models do not predict accurately for groups defined by prior medical expenditures. Providers will be able to practice substantial risk selection against survey models by employing their knowledge of the medical care use of actual or potential enrollees. Although prior-use and claims-diagnosis models worked well,

---

[9] The marginal cost of using surveys for risk adjustment may be low if the same survey instrument (e.g., the SF-36) that is used for outcomes assessment can also be used for risk adjustment (perhaps with a few added questions).

our survey models did not perform well for the disabled. At a minimum, parameters for the elderly and the disabled appear to be different and require separate estimates. Perhaps totally different survey models need to be developed for the disabled.

A final and very important point is that more data are needed to obtain stable and reliable estimates of risk-adjustment models before they can be implemented. Although we believe our estimates (Table 2) using a single year of MCBS data are plausible, they are clearly not very precise (as indicated by the large standard errors of estimates). Not surprisingly, comparison of parameter estimates for 1991-92 versus 1992-93 data shows substantial differences (Pope et al., 1997). For example, osteoporosis, which has a highly statistically significant coefficient of $1,312 in the 1991-92 chronic conditions model, has a negative and statistically insignificant coefficient of -$288 in the 1992-93 model. "Very good" in the self-rated health-status model has a statistically insignificant coefficient of $486 in 1991-92 versus a highly significant coefficient of $1,003 in 1992-93. More data will also increase the sensitivity of risk-adjustment models by allowing the estimation of health-status scales with more response levels (e.g., "a lot/some/a little/no difficulty" versus "some/no difficulty").

LIMITATIONS

This study has several significant limitations. We analyzed a particular set of survey models, albeit a wide range of this class of models. We analyzed only one claims-diagnosis-based model, the DCG-HCC model, not, for example, the ambulatory care groups model (Weiner et al., 1996, 1991). Our results may not generalize beyond the particular survey and claims-based models we analyzed. Nor will our results necessarily generalize to other populations. For example, the differences between our results and those of Fowles et al. (1996) may be attributable to that study's evaluation of the actual SF-36 survey scales and the ambulatory care groups claims model on a mixed population of persons under 65 years of age as well as the elderly.

Because many high-cost medical conditions that account for a large portion of expenditures are rare, large sample sizes, such as are available from claims files, are desirable for estimating and validating risk-adjustment models. Sample sizes comparable to claims samples are not currently available for survey variables. Our MCBS results—both estimation and validation—may not fully generalize to other samples because of the MCBS' limited sample size. That is, our results are influenced to some extent by random error. Nevertheless, we believe that most of our qualitative findings will generalize to other and larger samples.

Another technical limitation is that we did not have a fully independent validation sample, which may tend to overstate the predictive power of all risk-adjustment models. We did not include individuals in our sample for whom we did not have survey responses, either because a person did not respond to the MCBS at all, or because he or she did not answer a specific question. To the extent that survey non-respondents are sicker on average than respondents, our results may somewhat overstate the predictive power of models (especially survey models) compared with a full set of responses.

ACKNOWLEDGMENTS

this research. Anonymous reviewers and Randall Ellis also provided very useful comments on the research reported in this article.

REFERENCES

Duan, N., Manning, W.G., Jr., Morris, C.M., and Newhouse, J.P.: A Comparison of Alternative Models for the Demand of Medical Care. *Journal of Business and Economic Statistics* 1:115-126, 1983.

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

Ellis, R.P., and Ash, A.: Refinements to the Diagnostic Cost Group Model. *Inquiry* 32(1):1- 12, Spring 1995.

Fowles, J.B., Weiner, J.P., Knutson, D., et al.: Taking Health Status Into Account When Setting Capitation Rates: A Comparison of Risk-Adjustment Methods. *Journal of the American Medical Association* 276(16):1316- 1321, Ocyober 23/30, 1996.

Fowles, J.B., Weiner, J.P., and Knutson, D.: *A Comparison of Alternative Approaches to Risk Measurement: Final Report to Physician Payment Review Commission.* Nicollet Medical Foundation Grant 93-G07. Minneapolis, MN: Park Nicollet Medical Foundation, 1994.

Gruenberg, L., Kaganova, E., and Hornbrook, M.: Improving the AAPCC With Health Status Measures From the MCBS. *Health Care Financing Review* 17(3):59-76, Spring 1996.

Health Care Financing Administration: HCFA *Statistics.* HCFA Publication Number 03403. Baltimore, MD. October 1997.

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

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

Pope, G.C., Adamache, K.A., Khandker, R., and Walsh, E.: *Evaluating Alternative Risk Adjusters for Medicare.* Draft Report to the U.S. Health Care Financing Administration. Waltham, MA: Center for Health Economics Research, 1997.

Ware, J.E.: SF-36 *Health Survey Manual and Interpretation Guide.* Boston, MA: The Health Institute, New England Medical Center, 1993.

Ware, J.E., and Sherbourne, C.D.: The MOS 36-Item Short-form Health Survey (SF-36). *Medical Care* 30(6):473-483A, June 1992.

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

Weiner, J.P., Starfield, B., Steinwachs, D., Mumford, L.: Development and Application of a Population Oriented Measure of Ambulatory Care Case-Mix. *Medical Care* 29(5):452-472, May 1991.