

Adjusting capitation rates using objective health measures and prior utilization

by Joseph P. Newhouse, Willard G. Manning, Emmett B. Keeler, and Elizabeth M. Sloss

Several analysts have proposed adding adjusters based on health status and prior utilization to the adjusted average per capita cost formula. The authors estimate how well such adjusters predict annual medical expenditures among non-elderly adults. Both measures substantially improve on the variables currently used. If only health measures are added,

20-30 percent of the predictable variance is explained; if only prior use is added, more than 40 percent is explained; if both are added, about 60 percent is explained. The results support including some measure of use in the formula until better health measures are developed.

Introduction

A persistent theme in recent literature is the need for improvement in the adjusted average per capita cost (AAPCC), the method Medicare uses to pay health maintenance organizations (HMO's) and other capitated delivery systems (McClure, 1984; Thomas et al., 1983; Newhouse, 1986; Anderson and Knickman, 1984a, 1984b; Lubitz, Beebe, and Riley, 1985; Thomas and Lichtenstein, 1986; Anderson et al., 1986; Howland et al., 1987). Medicare pays the HMO an amount per enrollee that is based on Medicare payments per fee-for-service beneficiary in the county. Using the AAPCC formula, the amount is adjusted for differences between the HMO's enrollees and fee-for-service users with respect to age, sex, welfare status, institutional status, and basis for Medicare eligibility (over 65 years of age, disabled, or end stage renal disease). This method of computing payment rates poses several problems, two of which are relevant to this article.

- The county average may not apply to HMO enrollees if those using the fee-for-service system comprise an atypical group along dimensions that are not included in the formula and that affect utilization (for example, if those in the fee-for-service system constitute an abnormally sick group).
- An HMO has incentives to exclude those whose expenses will be predictably above the amount that the HMO is reimbursed.

A common conclusion in the literature is that the adjustments embodied in the AAPCC—age, sex, welfare status, and institutional status—are too crude. For example, the AAPCC pays an HMO the same amount for all 70-year-old, noninstitutionalized females who are not on welfare, live in the same county, and do not have end stage renal disease, but

there is an obvious disparity in the likelihoods that individual women within this group will use services during a year. A woman with cancer of the lung at the beginning of the year will almost surely use more services than a woman with no chronic disease. Moreover, the HMO will know or quickly learn the number of services that an individual person is likely to utilize, with corresponding incentives to encourage the person to remain enrolled or to disenroll. The same problem exists at a group level. If an HMO can identify relatively healthy groups of elderly, it will profit from enrolling them; conversely, there will be an incentive to not enroll unhealthy groups. Furthermore, some HMO's, because of their location or policies, may attract sicker patients than average. The ability of these HMO's to continue in business may depend on their being paid more than average. (The HMO could also economize by providing lower quality care. However, we assume that patients could detect such efforts and would disenroll as a result.)

Although these problems with the AAPCC are well known, their quantitative importance is still in some doubt. The most common metric for judging the AAPCC is explained variance or, more accurately, the lack of it. For example, Lubitz, Beebe, and Riley (1985) show that age, sex, welfare status, and institutional status explain only 0.6 percent of the variation in annual Medicare-covered expenditures for the elderly. Note that location is omitted from this list of variables. An estimate is not given of how much additional variance location would explain, but other results suggest that its inclusion would cause the figure of 0.6 percent to rise to about 1 percent (Anderson et al., 1986).

Although it seems clear that adjusters that can explain only 0.6 percent of the variance are scarcely better than no adjusters at all, it is less clear what percentage of variance explained would be satisfactory. Newhouse (1982) previously estimated that the maximum percentage that one should expect to explain is about 20 percent; McCall and Wai (1983) estimated the percentage to be 14 percent. The maximum is, in any event, much less than 100 percent because many health expenditures cannot be foreseen by either the individual or the HMO; that is, they are truly random. Such unforeseen expenditures should not cause an access-to-care problem as long as

This work was supported by the Health Care Financing Administration as part of its research program at the Center for Health Care Financing Policy of the RAND Corporation, the University of California at Los Angeles, and Harvard University. None of the three organizations necessarily endorses the views expressed here.

Reprint requests: Joseph P. Newhouse, Harvard University, Division of Health Policy Research and Education, 25 Shattuck Street, Parcel B, First Floor, Boston, Massachusetts 02115.

Medicare pays the average amount of these random expenditures. Moreover, adjustment could be too good: Any set of adjusters that explained 100 percent of the variance would not be capitation, as it is usually understood, but simply cost-based reimbursement.

Although there is considerable support for the notion that the AAPCC needs modification, there is less agreement on how it should be modified. Some propose adding measures of health status to the adjuster list. For example, Medicare might pay an HMO more for covering someone with lung cancer than it would for someone with no chronic disease (McClure, 1984; Thomas et al., 1983; Howland et al., 1987). Others are skeptical that usable measures of health status can be developed, at least in the short run, and propose taking account of actual utilization in the payment formula, either through adjustments based on prior use or by using a blend of capitation and actual use in the present year (Lubitz, Beebe, and Riley, 1985; Newhouse, 1986; Anderson and Knickman, 1984a, 1984b; Anderson et al., 1986). However, to many, taking account of use in any fashion represents an undesirable dilution in the incentives of capitation. The use of diagnostic cost groups represents an intermediate position because these groups are based on hospitalization for a specified chronic condition or a nondiscretionary acute condition (Physician Payment Review Commission, 1988).

All sides agree that, if it were practical, it would be desirable to incorporate health status measures into the formula. Those advocating a blend of capitation and fee for service suggest that more weight could be placed on the capitated amount as the state of the art with respect to health status adjusters improves. The health status measures most commonly proposed for inclusion are those that pertain to chronic diseases and functional status. The focus on chronic problems is proper if, as seems likely, acute problems and their associated expenditures are not foreseeable.

Our principal objective in this article is to assess the probable gain from developing measures of health status and prior use for inclusion in a modified formula. We did not attempt to develop a particular formula for the AAPCC, only to provide a general indication of the amount of improvement that might be possible from alternative types of variables. For that reason we did not conduct tests of specification. Moreover, the imperfections of explained variance as a criterion (for example, that it does not distinguish appropriate from inappropriate care) may be less serious for this objective than they would be for an absolute appraisal of any given formula.

In studying these problems, we used a unique body of data from the RAND Health Insurance Experiment (the Experiment), including measures of individuals' utilization and expenditures for 3 or 5 years, as well as both objective and subjective measures of health status. The data thus permit an assessment of how much variance in utilization could be explained by commonly proposed adjusters.

The major drawback of these data for the purposes of the Medicare program is that the individuals in the sample were all under 65 years of age and not eligible for Medicare. Elderly Medicare eligibles are hospitalized for different reasons (e.g., more frequently for hip fracture and not at all for pregnancy) and make much greater use of skilled nursing facility and home health services than persons under 65 years of age do. Nonetheless, the results for persons under 65 years may indicate the relative performance of these generic types of adjusters among the elderly, especially among the younger elderly (65-74 years of age). Furthermore, the results of Howland et al. (1987) for the elderly (although they are for only the number of hospitalizations, not expenditures) appear to be consistent with ours, suggesting that our results do apply to the elderly. Irrespective of their applicability to the elderly, our findings are relevant to rate setting for persons under 65 years of age, whose enrollment in capitated delivery systems is increasing.

Methods

Maximum R²

In the body of this article, we use an example to explain our methods in a relatively nontechnical manner. More detail can be found in the technical note. We begin by introducing some notation.

If medical expenditures follow the following model, determining how much variance one could possibly explain in a cross-sectional regression of annual expenditures is straightforward.

$$\text{Expenditure}_{it} = \beta X + u_i + e_{it}, \quad (1)$$

where i indexes person; t indexes year; X is a vector of adjusters; u_i is a person-specific, time-invariant component of variance; and e_{it} is a person-specific, time-varying component of variance.

If the last term, e_{it} , were in fact random (meaning that it could not be predicted by the HMO or by the patient), the maximum variance that could be explained by an AAPCC-like formula would be that accounted for by the first term plus any portion of u_i that might be explained by including additional adjusters whose effect did not vary over time.

We now work through an example of our method. Suppose the adjusters in the first term are those now in the AAPCC formula. Consider adding the person's cholesterol level to the formula. Because cholesterol level is a reasonably stable characteristic, its effect is now included approximately in the u_i term. Consider the group of persons who are in every way average except that they have high cholesterol levels, thus being at greater risk of a heart attack and, consequently, above average expenditures. For the high-cholesterol group, u_i is positive and equals the difference between the group's average expenditure and the average expenditure among all people.

Whether a heart attack actually occurs for any individual in the high-cholesterol group is, of course, uncertain. If a heart attack does not occur in any given year (and all other factors in that year are average) or if a massive heart attack occurs and the person dies immediately without being treated, e_{it} will be negative. In contrast, if a heart attack occurs that causes the person to spend many days in the intensive care unit recovering, e_{it} will be positive.

The variance in total expenditures is the sum of the variances of the three terms in equation 1. The amount of variance explained by the formula is the amount explained by the first term (βX). If cholesterol were added as an adjuster, the variance explained by the first term would increase, and the variance explained by the second term would decrease by the same amount. The variance explained by the third term, e_{it} , would not change, because adding cholesterol level would not explain the variance attributable to whether persons within the high-risk group actually had a heart attack.

It is important that the AAPCC include variables such as cholesterol level if HMO's can observe cholesterol levels and act on that information to encourage or discourage subsequent enrollment; that is, it is important that variance shift from u_i to βX . It is not important to explain the variation in the e_{it} (e.g., whether a person in the high-cholesterol group actually has a heart attack) because the HMO cannot know whether that will happen.

Although we want the formula to minimize the variance explained by the second term, no actual formula will make the variance explained by that term equal to zero; i.e., make u_i equal zero for each person. For example, another reasonably stable characteristic about an individual is whether the person is a hypochondriac. Because hypochondriacs make more physician visits, they too will have a positive u_i (all else being average), although each hypochondriac may have greater or lesser expenses (positive or negative e_{it}) depending on whether, for example, he or she happens to be injured in an automobile accident in any one year. It is not likely that hypochondriasis would ever be an adjuster; rather, its effect would be a permanent part of u_i , and the variance attributable to it would not be explained.

In order to judge the goodness of any formula, we need an estimate of how much variance one could expect to explain, or what we term the maximum R^2 . One way to estimate a maximum R^2 is to ask how much variance is explained by stable or nearly stable characteristics, such as age, sex, cholesterol level, and hypochondriasis. That can be done by regressing expenditures for a representative group of people for a series of years on dummy variables for each person, which is the same as determining how much of the total variance is between persons and how much is within person.

Such a calculation, in fact, yields a lower bound on the maximum R^2 that one can explain because it does not take account of measurable, time-varying characteristics. An example of such a characteristic is

the presence of a terminal illness. The spending rate rises in the penultimate month of life, and it rises still further in the last month of life. Thus, adding an indicator variable, "has terminal illness," will explain variation over and above the maximum R^2 computed in the fashion discussed. Nevertheless, most of the predictable variation is probably from stable characteristics, such as chronic diseases, habits, or risk factors (for instance, elevated cholesterol); the terminal disease example is probably exceptional. Put another way, much of the variation in the e_{it} probably reflects acute illness or injury or other unforeseeable demand for care, and it is not important to explain that variation. Hence, our lower bound on the maximum R^2 is probably a reasonably close lower bound.

The method we actually used to estimate the maximum R^2 was analogous but not identical to the method of including a dummy variable for each person. We discuss the details of the method we used in the technical note.

A competing model to that of equation 1 has been proposed by Welch (1985). The difference between Welch's model and that of equation 1 is explained in the technical note. A test that can be used to distinguish the two is the pattern of correlations over time in expenditures. Equation 1 predicts that those correlations will be constant; that is, if we consider a group of people, the correlation between their spending in year 1 and year 2 will be the same as the correlation between their spending in year 1 and year 3. Welch's model predicts that these correlations will decline geometrically; that is, the correlation between year 1 and year 3 will be a certain percent less than between year 1 and year 2, the correlation between year 1 and year 4 will be a certain percent less than that, and so on. Therefore, we also present the pattern of correlations in our data and conclude that equation 1 is a sufficiently good approximation for our purposes. Details can be found in the technical note.

Source of data

The data we use come from the RAND Health Insurance Experiment, the design of which has been described in many places (Newhouse et al., 1981; Brook et al., 1983; Manning et al., 1987). In this experiment, operated from 1974 to 1982, families in six areas of the Nation (Seattle, Washington; Dayton, Ohio; Charleston, South Carolina; Fitchburg-Leominster, Massachusetts; and two rural areas, Franklin County, Massachusetts, and Georgetown County, South Carolina) were randomized to insurance plans with varied cost sharing. Because variation in spending that resulted from cost sharing was induced by the experiment, we removed the effect of cost sharing from all observations (that is, we have removed the between-plan variance). Thus, we ask how well various explanatory variables or adjusters account for within-plan variance.

We removed the between-plan variance, so our results can be generalized to a population with one plan. Because of medigap insurance policies and some variation in the real value of deductibles and coinsurance across counties, not all of the Medicare population has a single plan. However, faced with the choice of explaining total variance or within-plan variance, we decided that within-plan variance is a better approximation of the Medicare situation than total variance is. Because plan explains relatively little variance, this choice is not critical.

The families that participated in the Experiment were randomly assigned to a 3-year or 5-year participation period, during which time the Experiment acted as their insurance company. (Participants formally assigned the benefits of any insurance for which they were eligible to the Experiment.) According to independent verification of physician office claims, the families filed claims with the Experiment for about 90 percent of their physician utilization (Rogers and Newhouse, 1985, p. 128).¹ Hence, we believe we have a nearly complete record of utilization for the period of participation.

The families invited to participate in the Experiment were randomly selected, subject to a number of qualifications. By far the most important for the purpose of this article is that no one eligible for Medicare was included in the Experiment. Other qualifications on the population sampled that are less important for the purposes of this article are that active-duty and retired military were excluded; veterans with service-connected disabilities were excluded; persons institutionalized indefinitely (principally prisoners and those in long-term psychiatric hospitals) were excluded; and, in five of the six sites (all but Seattle), low-income individuals were mildly oversampled.

A total of 3,958 individuals 14 years of age or over enrolled in the Experiment. All those living at a given dwelling unit who met eligibility requirements were offered enrollment. Hence, the 3,958 observations are not all independent; for example, a husband's and wife's utilization could be expected to be correlated. Our estimation methods do not account for this correlation, but accounting for it would not greatly affect our estimates of the proportion of variance that various sets of adjusters can explain. In addition, as we will show, there is dependence over time within person. The essence of our problem is to estimate the dependence in the residuals over time.

The sample used for the regression equation included only those participants who completed the study and completed the final exit examination because we did not want to impute missing physiological health data.²

¹The figure of 90 percent is estimated as $(53.88 / (53.88 + 6.50))$, where 53.88 is the amount matched between the billing record of the physician and claims filed with the Experiment and 6.50 is the amount on the physician's billing record not matched with claims.

²We did not use entry physiological variables because 40 percent of the participants in the Experiment were randomly assigned to not take the physical examination.

We included in this analysis only persons 14 years of age or over because our measures of health status are different for persons of younger ages. In the regression analysis we did not use those in their first year of participation because we did not have comparable prior-use data for them. We did use first-year data in examining the stability of year-to-year correlations.

We excluded persons with any missing data for physiologic variables. Such persons included those who did not complete the Experiment, those who moved out of area during the Experiment and so did not complete an exit screening examination, and those who for any other reason had missing physiological data resulting from nonresponse. Not completing the Experiment was relatively uncommon; more than 90 percent of the participants completed the Experiment and exited normally, and another 1 percent died. Persons who did not complete the Experiment (except those who died) had a rate of expenditures while they were participating that was statistically indistinguishable from the rate of those who did complete the Experiment. Hence, bias from attrition should be minimal.

Data on one-quarter of the enrolled persons are missing for one reason or more. In all, our sample consisted of 7,690 person-years. There are 818 person-years with any inpatient use.

Dependent variables

Our major interest was to estimate or predict annual expenditures for medical care services in constant dollars. The unit of analysis was the person-year rather than the family-year because the primary determinants of utilization are individual characteristics. The services included in the analysis were virtually all medical services other than dental and outpatient mental health services. Prescription drugs were included (but accounted for only about 10 percent of the expenditures), as were eyeglasses, hearing aids, and other supplies and appliances. Over-the-counter (OTC) drugs were included if an individual had a chronic condition for which an OTC drug might be the treatment of choice (such as aspirin for those with arthritis). Further description of services included in the analysis is available in Newhouse et al. (1981). In addition to examining total medical expenditures, we examined separately how well one can predict variation in expenditures for inpatient and outpatient services.

Calculations of R^2 can be distorted by outliers. For that reason, we calculated not only the conventional R^2 but also the R^2 with the dependent variables trimmed; that is, if a dependent variable was above the 98th percentile, it was set equal to the mean of the upper 2 percent of the observations. For example, for total medical expenditures, the 98th percentile was 2.28 standard deviations above the mean; all expenditures greater than this were set equal to the mean of the upper 2 percent of the distribution. This preserved the overall mean. The proportion of

variance explained by various adjusters was similar for both trimmed and untrimmed data; hence, we present only the untrimmed results here.

Potential adjusters

Because we wished to ignore between-plan variation, we began by regressing expenditures on plan, which, by design, is approximately orthogonal to all other covariates (Morris, 1979). Hence, we ask, what is the increment in explained variance from adding a series of adjuster variables over adding plan alone? We show the R^2 from the plan-only regression. We then present:

$$(R^2(b) - R^2(a))/(1 - R^2(a)), \quad (2)$$

where a indexes the specification with only the plan variables, and b indexes any of the more complete specifications.

For purposes of removing variation because of plan, we used the logarithm of the coinsurance rate plus a dummy variable for the individual deductible plan, ignoring the small amount of variation induced by the varying percentage of income ceilings on out-of-pocket expenditures.

We used the sets of explanatory variables shown in Table 1 as possible adjusters. First, we approximated the variables used in the current AAPCC formula: age; sex; Aid to Families with Dependent Children status (Supplemental Security Income recipients, who are eligible for Medicare, are not in the sample population); and site, which approximately corresponds to county. Then we added the following four sets of variables to the AAPCC set of variables.

Dichotomous physiologic health—This set of dummy variables indicates the presence or absence of the physiologic conditions shown in Table 1. Variables defined in Table 1 as (0,1) were included in the regression unchanged. Variables defined in Table 1 as the maximum of zero and the test value minus some cutoff point were dichotomized before being included in this set of variables; that is, if X is greater than the cutoff point, then Z equals 1. For example, a dummy variable for hypertension assumes the value 1 if the individual has a diastolic pressure of 90 mmHg (millimeters of mercury) or higher, has a systolic blood pressure of 140 mmHg or higher, or is under treatment for hypertension. These physiological measures were derived from data collected at exit from the study.

Continuous physiologic health—This set of variables indicates the presence or absence of the physiologic conditions shown in Table 1 and, for some conditions, serves as a measure of severity. Variables were included in the regression as defined in Table 1. For example, two variables related to hypertension were included in the regression: diastolic blood pressure (DBP), coded as the maximum of zero and $(DBP - 89)$, and the dummy variable for hypertension described in the previous paragraph.

In principle, the dummy variable measures the fixed

costs of treating the condition, and the continuous variable measures the variable cost of increased severity. All variation below a cutoff point, for example, 90 mmHg diastolic blood pressure, is suppressed. The cutoff points reflect a judgment about values below which most physicians would not treat; for example, most physicians would probably not prescribe treatment for diastolic blood pressure values below 90 mmHg. At or above the cutoff point, we simply entered the physiologic measure linearly. It is quite possible, indeed probable, that the true functional form above the cutoff point is nonlinear, but with a limited sample with each specific condition, we felt that experimenting with nonlinear functional forms would mean overfitting the data and thus overstating the probable performance of these measures. Put another way, our principal purpose was to gauge the amount of variance one might be able to explain with a set of health status measures and a set of use measures that were reasonably complete. We were not attempting to determine the appropriate functional form. Our linear form can, of course, be regarded as a first-order Taylor Series approximation to the true form (above the cutoff point). For the same reason, we did not explore interactions; for example, we treated the effects of having high blood pressure and diabetes mellitus as additive.

Although it may seem that expenditures should increase with less healthy values—for example, higher blood pressure—such a relationship will not necessarily hold in the data. Specifically, it will not necessarily hold if treatment alters the physiologic measure and less healthy patients utilize more resources (or if not all individuals are under treatment). For example, a hypertensive individual whose blood pressure is controlled at 90 mmHg but whose uncontrolled value is 105 mmHg could be expected to have higher medical expenditures during our period of observation than an otherwise identical hypertensive individual who is not under treatment; in such a case, the relationship between observed blood pressure and medical expenditures would be negative.

An extension that partially allows for this difficulty is to enter a dichotomous variable for being in treatment, a specification we also estimate. (The variable took the value 1 if a physician indicated a diagnosis of a condition in Table 1 on a claims form.) Incorporating such an adjuster has the additional advantage that the relevant information can, in principle, be collected solely from claims forms. Nonetheless, such an approach is only a partial solution because it does not allow for bias within the treated group. For example, one person may have an uncontrolled diastolic blood pressure of 110 mmHg and another of 100 mmHg. Both individuals may have their blood pressure controlled to 90 mmHg, but the costs of treating the first person may be greater because the case is more severe. Yet, this cost difference would appear to the analyst as unexplained.

A set of measures on functional status (physical health), self-rated general health perceptions, mental health, and the presence of a variety of self-reported

Table 1
Definitions of health status measures

Measure	Definition
Physiologic	
Elevated cholesterol	0 if 0-259 mg/dL; X - 259 if 260 mg/dL or more
Hypertension	Elevated blood pressure (systolic pressure of 140 mmHg or more or diastolic pressure of 90 mmHg or more or taking blood pressure medication: 1 = present, 0 = absent
Elevated diastolic blood pressure	0 if 0-89 mmHg; X - 89 if 90 mmHg or more
Diabetes	Elevated glucose level (160 mg/dL or more) or taking insulin or oral agents: 1 = present, 0 = absent
Elevated glucose	Measured as random glucose: 0 if 0-159 mg/dL; X - 159 if 160 mg/dL or more
Gout	Reported diagnosis of gout by physician: 1 = present, 0 = absent
Chronic joint symptoms	Self-reported symptoms characteristic of moderate joint disorders: 1 = present, 0 = absent
Hay fever	Self-reported hay fever during lifetime: 1 = present, 0 = absent
Hay fever scale	Self-reported amount of time per year bothered by hay fever on a natural logarithm scale ranging from 0 (none) to 6.4 (6 months or more)
Impaired natural far or near vision	Measured without corrective lenses, better eye: 0 if 10/20-20/20; X - 20 if 25/20 or more
Impaired hearing	Measured as simple average of thresholds at 500, 1,000, and 2,000 hertz for worse ear: 0 if 0-25 decibels; X - 25 if 26 decibels or higher
Shortness of breath scale	Self-reported measure based on dyspnea questionnaire, ranging from 0 (no shortness of breath) to 4 (severe shortness of breath)
Impaired forced expiratory volume in 1 second (percent of predicted)	Measured through spirometric testing and expressed as percentage of predicted volume based on published equation (Knudson et al., 1976) ¹ —best of three tries: 0 if 80 percent or more; 80 - X if less than 80 percent
Electrocardiogram abnormalities	Presence of one or more of the following findings: intraventricular conduction abnormalities, ventricular enlargement (including left ventricular hypertrophy), atrial fibrillation, ST-segment and T-wave changes, Q-wave abnormalities, ventricular dysrhythmias, artificial pacemaker rhythm; 1 = present, 0 = absent
Anemia	Low hemoglobin, current treatment for anemia, or previous diagnosis of anemia: 1 = present, 0 = absent
Low hemoglobin	Measured automatically by the Coulter Model S machine (Coulter Electronics, Inc., Hialeah, Florida). For males under 18 years of age: 0 if 12.5 grams/100 milliliter (g/100 mL) or more; 12.5 - X if less than 12.5 g/100 mL. For males 18 years of age or over: 0 if 13.0 g/100 mL or more; 13.0 - X if less than 13.0 g/100 mL. For females of all ages: 0 if 11.5 g/100 mL or more; 11.5 - X if less than 11.5 g/100 mL
Acne	"In the past 12 months, have you had trouble with pimples on your face?": 1 = yes, 0 = no
Severity of acne	Scale based on reading of facial photograph by a dermatologist: 0 = no acne; 1 = 1 comedo or papule, 2 = extensive comedos or papules, 3 = pustules, 4 = inflammatory cysts, 5 = acne conglobata
Varicose vein scale	Severity of varicose veins in the worse leg based on physical examination: 1 = absent, 2 = spider angiomas, 3 = minimal, 4 = moderate, 5 = severe
Active ulcer	Stomach pain or ache in past 3 months with previous history of physician diagnosis and ulcer confirmed by X-ray or symptom pattern characteristic of ulcer: 1 = present, 0 = absent
Dyspepsia	Self-reported episodes or attacks of stomach pain or ache in past 3 months (patients with active ulcer classified as having no dyspepsia): 1 = present, 0 = absent
Urinary tract or kidney infection	"Have you ever had a kidney, bladder, or urinary tract infection?": 1 = yes, 0 = no
Urinary tract infection	Growth greater than or equal to 100,000/milliliter of one or more pathogens or patient taking prescription medication for urinary tract infection: 1 = present, 0 = absent
Hemorrhoids	Hemorrhoids in the past 12 months: 1 = present, 0 = absent
Hernia	Hernia in the past 12 months or operation for hernia, rupture, or herniated navel during lifetime: 1 = present, 0 = absent
Angina	Symptoms of and/or physician diagnosis of angina pectoris: 1 = present, 0 = absent
Abnormal thyroid function	Classified as abnormal if using thyroid medication or if T7 measurement is low (hypothyroid) or high (hyperthyroid): 1 = present, 0 = absent

See footnotes at end of table.

Table 1—Continued
Definitions of health status measures

Measure	Definition
Subjectively rated health status	
Physical health	1 if either of 2 scales of physical health are not at maximum value; 0 otherwise (i.e., no limitation). The 2 scales are described in Sloss et al. (1986) ² and are: AROLE, a measure of role limitations (88 percent are not limited); APERS, a measure of personal limitations (85 percent are not limited)
Mental health	A 0-100 scale based on 38 items for adults and 12 items for children that was administered in sites other than Dayton. Described as the variable MHI in Sloss et al. (1986). ² Further description can be found in Brook et al. (1984) ³
General health	A 0-100 scale based on 22 items for adults administered in sites other than Dayton. Described as the variable GHINDX (adults) in Sloss et al. (1986). ² Further information can be found in Davies and Ware (1981) ⁴ and Brook et al. (1984). ³ The variable GHIOOP (Sloss et al., 1986) ² is used for Dayton participants
Disease count	Two variables are used. Dummy variable = 1 if any of 32 chronic diseases listed in Manning, Newhouse, and Ware (1982, Appendix B) ⁵ is present and = 0 otherwise. Second variable = logarithm of number of diseases
Demographic	
Age	in years
Sex	1 = female, 0 = male
Site	5 dummy variables
Aid to Families with Dependent Children (AFDC) status	1 if AFDC eligible at baseline; 0 otherwise
Utilization	
Outpatient expense in prior year	0 = no expense, 1 = positive expense
Inpatient expense in prior year	0 = no expense, 1 = positive expense
Logarithm of outpatient amount if positive outpatient expense	0 = no positive outpatient expense
Logarithm of inpatient amount if positive inpatient expense	0 = no positive inpatient expense

¹Knudson, R. J., Slatin, R. C., Leibowitz, M. D., and Burrows, B.: The maximal expiratory flow-volume curve: Normal standards, variability, and effects of age. *American Review of Respiratory Disease* 113:587-600, 1976.

²Sloss, E. M., Colbert, L. L., Wesley, D. L., et al.: *Health Status and Attitude Series, Volume 1, Codebooks for Adults and Children at Enrollment and Exit*. Pub. No. N-2447/1-HHS. Santa Monica, Calif. RAND Corporation, 1986.

³Brook, R. H., Ware, J. E., Jr., Rogers, W. H., et al.: *The Effect of Coinsurance on the Health of Adults: Results from the RAND Health Insurance Experiment*. Pub. No. R-3055-HHS. Santa Monica, Calif. RAND Corporation, 1984.

⁴Davies, A. R., and Ware, J. E., Jr.: *Measuring Health Perceptions in the Health Insurance Experiment*. Pub. No. R-2711-HHS. Santa Monica, Calif. RAND Corporation, 1981.

⁵Manning, W. G., Newhouse, J. P., and Ware, J. E., Jr.: The status of health in demand estimation: Beyond excellent, good, fair and poor. In Fuchs, V. R., ed. *Economic Aspects of Health*. Chicago. University of Chicago Press, 1982.

NOTES: mg/dL is milligrams per deciliter. mmHg is millimeters of mercury.

chronic diseases—Although the use of such variables as adjusters in the AAPCC seems problematic because of the possibilities for fraud, we thought one should ascertain the possible gains from using them. To the degree that medical care for a chronic problem affects these measures and that medical care is greater with more severe problems, the same bias described for the physiologic variables is present. These variables were collected at entry into the study.

Four variables measuring use of medical services in the previous year—These are: whether there was any outpatient expenditure, whether there was any inpatient expenditure, and the logarithms of outpatient and inpatient expenditures for those with positive expenditures.

Estimation methods

To determine the promise of various types of adjusters, we used a variant of the four-equation model we have used in other work (Duan et al., 1983; Manning et al., 1987), with the variables in Table 1 used as explanatory variables. This variant separates outpatient and inpatient expenditures rather than persons with no inpatient expenditure and persons with inpatient expenditure. We then computed the amount of explained variation as follows.

We first predicted the total expenditure of each person using the four-equation model. The predicted value equals $p_i E(1, i) + P_i E(2, i)$, where p_i is the predicted probability of positive outpatient

expenditure for person i , P_i is the probability of positive inpatient expenditure, $E(1,i)$ is the expected outpatient expenditure, and $E(2,i)$ is the expected inpatient expenditure. $E(1,i)$ and $E(2,i)$ are retransformed from logarithms using Duan's smearing estimator (Duan, 1983; Duan et al., 1983).

We then calculated a measure of R^2 suggested by Efron (1978), using the following formula:

$$\text{Efron's } R^2 = 1 - (\Sigma(y_i - \hat{y}_i)^2 / \Sigma(y_i - \bar{y})^2), \quad (3)$$

where \hat{y} is the predicted y using the four-equation model with alternative sets of explanatory variables, and \bar{y} is the sample mean of y . Thus, the numerator of the fraction in parentheses is the unexplained sum of squares, and the denominator is the total sum of squares. Although this measure of R^2 reduces to the usual measure in the case of a linear model, it can be negative when one predicts from a nonlinear model such as ours. In this application, however, it never was negative. We then computed the ratio of this R^2 to the maximum R^2 , defined earlier.

We used the four-part model to predict y rather than the more ordinary analysis of covariance because the four-part model has less tendency to overfit the sample data (Duan et al., 1983). Hence, use of analysis of covariance, which is common in the literature, overstates how well one can do. We used Efron's R^2 because the four-part model is nonlinear. We did not adjust the R^2 value for the number of parameters in the model, but the number of observations is large relative to the number of parameters, so any such adjustment would be trivial.

Results

The variance explained by the alternative specifications is shown in Table 2. Several results in Table 2 are noteworthy.

We estimate that the maximum R^2 one could achieve in explaining total expenditures is 14.5 percent. The percentage for outpatient expenditures only is much higher, almost 50 percent, but total variance is dominated by the variance of inpatient expenditures. Thus, the ability to explain total expenditures is relatively low. Recall that our estimates of the maximum R^2 , the denominator, are probably too low; hence, our percentages of the maximum explained variation are probably too high.

The AAPCC variables by themselves explain only 11 percent of the variance in total expenditures that could be explained. To be sure, 11 percent is not negligible, but substantial room for improvement remains.

The simple measures of health we use clearly are improvements on the current AAPCC variables, but all variants of the health measure are rather modest improvements on the AAPCC variables. The percentage of variance in total expenditures that is explained rises from 11 percent with the AAPCC variables alone into the range of about 20-30 percent. The continuous physiologic health measure does not

Table 2
Percentage of maximum explained variation in health care expenditures yielded by alternative specifications, by type of expenditure

Dependent variable	Total expenditures	Inpatient expenditures	Outpatient expenditures
	Percent		
Plan only	0.25	0.05	1.64
Between-person variance as a percent of total variance (maximum R^2)	14.5	8.0	48.2
AAPCC: Age, sex, site, Aid to Families with Dependent Children status	11	9	15
AAPCC plus dichotomous physiologic health	26	25	28
AAPCC plus dichotomous physiologic health based on claims	31	27	26
AAPCC plus continuous physiologic health	29	32	27
AAPCC plus subjective health	19	15	23
AAPCC plus subjective health and continuous physiologic health	32	35	30
AAPCC plus prior-year use	44	35	44
AAPCC plus dichotomous physiologic health and prior-year use	55	51	51
AAPCC plus continuous physiologic health and prior-year use	60	59	51
All variables	62	62	52

NOTES: Between-plan variation removed except from plan-only row. Denominator of percentage is value in second (maximum R^2) row. AAPCC is adjusted average per capita cost.

SOURCE: Data from the RAND Health Insurance Experiment.

do notably better than either dichotomous version. This finding is important because one set of results for the dichotomous variables is defined from claims forms (albeit diagnosis codes are not now available in the Medicare Part B data). The continuous physiologic variable obviously costs more both to collect and to audit.

The subjective health measures, including physical health (functional status), do not do as well as even the dichotomous physiologic measures and add little to the continuous physiologic measures. Functional status measures may, however, be more important in an elderly population.

The measures of use in prior year are a substantial improvement on any of the health status measures in isolation. The percentage of the maximum variance that might be explained solely using prior-year use plus the current AAPCC variables rises to 44 percent.

Table 3
Correlation coefficients for health care expenditures, by type of expenditure and year

Type of expenditure and year	Year 2	Year 3	Year 4	Year 5
	Correlation coefficient			
Total				
Year 1	.090	.054	.044	.045
Year 2		.221	.195	.184
Year 3			.265	.065
Year 4				.192
Inpatient				
Year 1	.039	.012	.009	.011
Year 2		.147	.151	.042
Year 3			.226	.025
Year 4				.114
Outpatient				
Year 1	.540	.468	.470	.416
Year 2		.530	.363	.331
Year 3			.524	.471
Year 4				.503
Probability of any inpatient admission				
Year 1	.107	.100	.110	.106
Year 2		.128	.054	.107
Year 3			.172	.107
Year 4				.154
Probability of any outpatient expense				
Year 1	.411	.351	.340	.324
Year 2		.390	.298	.311
Year 3			.403	.378
Year 4				.389

NOTE: Sample of cases is approximately 2,960 for years 1-3 and approximately 890 for years 4 and 5.

SOURCE: Data from the RAND Health Insurance Experiment.

Adding both the physiologic measures and measures of prior-year use gains approximately another 10 to 15 percentage points over the measures of prior year use and AAPCC variables in isolation.

With all variables included, 62 percent of the maximum possible variance is explained. Put another way, more than one-third of the stable variation in expenditures is not being picked up by these measures of health status and prior-year use.

The stability of year-to-year correlations in our data is shown in Table 3. Considerable sampling variance exists in the correlation matrixes for total and inpatient expenditures. These correlations tend to be dominated by those with large inpatient bills in any one year. The correlations tend to decline with time, but the tendency is not large. Given these data and data from James Beebe cited in the technical note (Welch, 1985), we conclude that our decomposition of variance based on equation 1 is approximately correct. Further discussion can be found in the technical note.

Discussion

Capitation payments reduce the incentives for overuse created by higher demand resulting from

third-party insurance and fees in excess of marginal cost to fee-for-service providers. Because the consumer agrees to receive all services from the capitated group, the group can ration services whose marginal private benefit falls short of marginal social costs. In contrast, if fee-for-service providers receive fees in excess of marginal cost, they have a reason in addition to the insurance subsidy to provide to the patient more than the economically efficient amount of services (Pauly, 1980).

In the case of both fee for service and capitation, competition among providers, if effective, can offset the distorted fee at the time of use. For example, many people may not want to join an HMO with a reputation for stinting on services in the case of illness. There will also be some willingness to pay for an effective guaranteed renewability (that is, to join HMO's that do not encourage members who become chronically ill to disenroll). However, many fear that competition will not suffice to prevent some HMO's from selectively enrolling low-risk elderly.

In addition to the possible problem of active selection because of financial incentives, a pure capitation system also poses a potential problem of passive selection because some HMO's may be attractive to enrollees who are sicker than the average person. HMO's whose members are sicker than average will have above-average costs. Unless they receive larger capitation payments, HMO's with a sicker caseload will be at a competitive disadvantage.

The AAPCC adjustments are aimed at the issue of selective enrollment. If the adjusted payment for an individual reflects HMO expectations of what that individual will cost, then the HMO has no incentive to select healthier patients, and the playing field for HMO's with varying types of patients will be level. To see how much a modified AAPCC might reduce incentives to select healthy patients, we have estimated how much the HMO gains or loses from accepting applicants it deems profitable. We have ignored the costs incurred to determine if particular patients are profitable.

Assuming that there would be no repercussions from rejecting applicants, an HMO interested only in short-term profits might reject all those whose predicted costs were higher than the AAPCC-adjusted capitation payment. The greater the HMO's ability to discriminate between people who are high or low cost relative to their AAPCC-adjusted payments, the more profits lie in a rejection program. The expected gain per case rejected depends on the standard deviation (SD) of HMO predictions of differences between the actual cost and the AAPCC-adjusted capitation amount. Because the SD is the square root of the variance and the actual expenditures have such a large variance, even a small additional percentage of variance explainable by the HMO can lead to fairly substantial gains from discriminating.

To keep numbers round, we assume that Medicare expenditures have a raw mean annual expenditure of \$3,000 and an SD of \$9,000. (Extrapolating 1986 expenditures to 1988 and correcting for beneficiary

Table 4

Results of better prediction of medical expenditures by health maintenance organizations (HMO's) than from adjusted average per capita cost (AAPCC) formula, by additional variance explained by HMO

Additional variance explained by HMO	Standard deviation of logarithm of prediction differences	Percent of predictions below mean	Profit per enrollee ¹
0 percentage point	.00	² 0	\$0
1 percentage point	.29	56	630
5.5 percentage points	.63	62	1,170
7.5 percentage points	.72	64	1,320
13 percentage points	.88	67	1,530
18.5 percentage points	.99	69	1,650

¹The formula is the mean (\$3,000 in the calculation) times 0, .21, .39, .44, .51, and .55, respectively, from top to bottom, assuming the HMO enrolls only patients it expects to be profitable.

²In this case, the HMO's predictions coincide with the AAPCC.

growth would yield a figure of slightly less than \$3,000 (Levit et al., 1985; Division of National Cost Estimates, 1987). In the Experiment, the standard deviation of annual expenditures was three to five times the mean, depending on plan (Manning et al., 1987), and for Medicare enrollees in 1976, the standard deviation of costs was about three times the average outlay per enrollee (Beebe, Lubitz, and Eggers, 1985.) Also suppose that the HMO can predict the maximum 14.5 percent of the variance (first column of Table 2). We can compute the expected standard deviation of the expected gains or losses per patient, assuming an AAPCC that explains 11 percent of the maximum variance (first column of Table 2). Under these assumptions, the HMO can predict an additional 13 percentage points of variance (13 = 14.5 (100 - 11)). More details are contained in the technical note. Ignoring any costs associated with active selection, the HMO maximizes profits by rejecting the 33 percent of applicants whose predicted costs are above the payment (Table 4, column 2, percent of predictions below mean: 33 = 100 - 67). The remaining people cost an average of 49 percent of the capitation payment, leading to an expected profit of 51 percent of the capitation payment on each enrollee, or \$1,530.

Such behavior would be extreme. We are assuming that the HMO is solely interested in short-run pecuniary gain and is risk neutral. Although these assumptions are unlikely to hold, the potentially large rewards from pursuing a policy of selective enrollment and disenrollment is indicated by the numbers in Table 4.

Suppose that the AAPCC were improved; then the additional variance explained by the HMO would fall. As this happened, the HMO still would profit from discrimination, but at a decreasing rate (Table 4). (As the predictive ability of the AAPCC rises relative to that of the HMO, one moves up in the last two columns of Table 4.) For example, if the AAPCC is based on our most complete specification, it explains 9 percent of the variance. Assuming that the HMO

still can predict 14.5 percent, it can predict 5.5 percentage points of additional variance. If the HMO can select the 62 percent of patients with predicted expenses below the AAPCC-adjusted fee, it will make \$1,170 per enrollee.

Thus, as shown in Table 4, a better AAPCC does reduce the profitability from pursuing active selection, but substantial incentives remain unless the AAPCC can explain expenses almost as well as the HMO can. Even if the HMO can explain only 1 percent more of the variance, it will still gain \$630 per accepted patient and has an incentive to reject 44 percent of the applicants.

Because the explainable variance is small in absolute terms, luck plays a much larger role than predictive ability in the gains and losses from any particular case. However, reasonable-sized HMO's should be able to rely on the law of large numbers to smooth random fluctuations in profits. (A case could be made for outlier payments to small HMO's as an alternative to basing a portion of reimbursement on total utilization.)

Our results are somewhat discouraging with respect to physiologic health measures. A considerable portion of expenditures is stable from year to year and cannot be predicted with the physiologic measures that we used. Indeed, using those measures plus the current AAPCC measures leaves unexplained about 75 percent of the variance that one might hope to explain. Results are even more discouraging if one uses subjective measures of health rather than physiologic health. Moreover, subjective measures would be more susceptible to potential fraud.

Of course, more complete measures of health status would be more useful than our simple measures. However, Howland et al. (1987) found that the physiologic variables studied in the Framingham Study, together with demographic variables, predicted only about 5 percent of the variance in the number of hospitalizations for males and about 2 percent for females. These results suggest that the gain is not of quantitative importance. Even if it were, such measures would pose several problems.

- Obtaining such data might require invasive tests, but ethical considerations preclude invasive tests on an asymptomatic population. McClure (1984) has suggested obtaining results from the health plan for those with a condition, with others assumed not to have the condition. However, such a procedure may cause auditing problems for labile conditions or conditions that change with treatment. For example, consider an individual whose diastolic blood pressure the plan correctly reports as 110 mmHg. On another day, the blood pressure might be 105 mmHg because of lability. Moreover, one might expect the plan to begin treatment to reduce the blood pressure, rendering it impossible for any later, independent verification that the blood pressure at one time was, in fact, 110 mmHg.
- More complete measures, even of a noninvasive variety, would be costly and could require an expensive patient chart audit.

- Calibrating the payment schedule for relatively rare conditions would require a large sample, particularly if interactions among conditions are important.

Thus, we are skeptical that salvation lies in a much more complete battery of physiologic measures, although there would undoubtedly be some gains from a more complete battery.

Suppose one interprets these findings as follows. Neither the adjusters currently included in the AAPCC nor those variables augmented by measures of health status are likely to produce a wholly satisfactory set of adjusters. Specifically, the AAPCC will remain vulnerable to a nonrepresentative group of risks in the fee-for-service system and there will remain an incentive for capitated plans to discriminate against bad risks. If one were to interpret our results in that light, how should Medicare reimburse capitated plans?

The usually proposed alternative to adding only health status adjusters is to account for prior or current use. According to our results, however, even prior use leaves about one-half of the explainable variance unexplained. Taken in conjunction with the gains from selection shown in Table 4, these results suggest that reimbursement should be made on the basis of a weighted average between current use and a capitated rate, which is adjusted as well as possible for differences in expected expenditures at the individual level.

We have left open the question of how much weight the capitated rate should receive in such an average, but in our view the weight should reflect a compromise between the economic incentives for overutilization in fee for service and the incentives for underutilization in pure capitation. An empirical approach that determines the sensitivity of market behavior and health outcomes to alternative weights seems to be a practical way to proceed.

Technical note

Estimating maximum R^2

We used two different methods to estimate the maximum R^2 , assuming that equation 1 was the relevant equation. The first was a two-step process. We used a four-equation model analogous to that presented in Duan et al. (1983) with our most complete specification of explanatory variables. The four equations are a probit equation to estimate the likelihood of outpatient expenditures, a probit equation to estimate the likelihood of inpatient expenditures, an equation to estimate the logarithm of outpatient expenditures for those with positive outpatient expenditures, and an equation to estimate the logarithm of expenditures for those with positive inpatient expenditures. We then calculated the predicted residuals from this model for each person. We used the predicted residuals to calculate an estimate of within-person and total residual variance.

The ratio of those two variances is an estimate of the proportion of the residual variance attributable to u_i .

A second method of estimating the proportion of variance in the u_i term is analogous to estimating the R^2 by using a dummy variable for each person. In this method, we subtracted an estimate of within-person variance from total variance, correcting for the bias that results from estimating within-person variance from a finite time series (Searle, 1971, chapter 11). We followed this approximation because of the computational problems in computing a random-effects model for the residual in total medical expenses.

In principle, the first method should yield a higher estimate than the second because it accounts for measurable time-varying covariates. In practice, however, the time-varying covariates that were included, such as self-assessed health status, were all measured at an initial point and were not updated. Perhaps for that reason, the estimate from the second method exceeded that from the first method, and we therefore have used the estimate of maximum R^2 from the second method.

The specific formulas used in the second method follow.

Let y_{it} = expenditure by person i in time t (in dollars). Let n_i = number of years person i is observed, and let a = number of persons. Calculate

$$\begin{aligned} T_o &= \sum_i \sum_t y_{it}^2; \\ T_A &= \sum_i y_i^2 / T, \text{ where } y_i = \sum_t y_{it}; \\ T_\mu &= y^2 \dots / N, \text{ where } N = \sum_i n_i; \\ S_e^2 &= (T_o - T_A) / (N - a); \text{ and} \\ S_a^2 &= (T_A - T_\mu - (a - 1)S_e^2) / (N - S/N), \\ &\text{where } S = \sum_i n_i^2. \end{aligned}$$

Then our estimate of maximum R^2 is $S_a^2 / (S_e^2 + S_a^2)$.

Autoregressive versus error components models

In a model proposed by Welch (1985) that is different from the error components model of equation 1, it is assumed that the errors follow a first-order autoregressive process.

$$u_{it} = \rho u_{i,t-1} + v_{it}, \quad (4)$$

where v_{it} is an independently and identically distributed random factor and ρ ranges from -1 to 1 . For values of ρ equal to 1 or -1 , it may appear that equation 4 reduces to equation 1, but this is not the case. If ρ equals 1 or -1 , the variance of the error term in equation 4 increases without bound as t increases (that is, it is non-ergodic), which is not the case in equation 1.

In the case of equation 4, the potential explainable variance is the variance explained by the adjusters plus the variance explained by the first term of equation 4 (because when one is predicting year t 's

expenditures, one has an estimate of $u_{i,t-1}$). Thus, in this model, the maximum R^2 is asymptotically the variance explained by the adjusters plus ρ^2 times the asymptotic variance of u , all divided by total variance, where the asymptotic variance u is $\rho^2/(1 - \rho^2)\text{var}(v)$. (The result is asymptotic because $\text{var } u$ increases with t .)

A straightforward test for distinguishing between equations 1 and 4 is to determine the pattern of correlation of the residuals over time. In equation 1, the correlation between the residuals for time periods t and $t + s$ for varying s should be constant (up to sampling error) and equal to the ratio of the variance of u to the variance of u plus the variance of e . In the second model, the correlation should decline geometrically (specifically, it should approximate ρ^s). In other investigations, health and physiological measures have been shown to follow a flat autoregressive pattern that would not differ much from a variance-components pattern over short periods. For example, the n -year correlation of cholesterol measurements in the Framingham Study is approximately $.88(.98)^n$ (Berwick, Cretin, and Keeler, 1980).

We present in the body of this article results on the time pattern of the correlations in our data, but because of a smaller sample size and shorter period of observation, they are less reliable than data from an unpublished study by Beebe that are cited by Welch (1985). Beebe estimated the correlation of expenditures by Medicare beneficiaries over a 6-year period. The correlations between expenditures in year 1 and expenditures in each of the five subsequent years were, respectively, 0.22, 0.14, 0.12, 0.13, and 0.11. (That is, 0.22 is the correlation from year 1 to year 2, 0.14 the correlation from year 1 to year 3, etc.) Although there is a decline from year 1 to year 2, there is approximate constancy after that. These data thus suggest the following model, which is a hybrid of equations 1 and 4.

$$\text{Correlation}_{i,t+T} = \rho(VC) + K(\rho(AR))^T, T = 1 - 5, \quad (5)$$

where $\rho(VC)$ is the proportion of variance attributable to the u_i term of equation 1, K scales the variance resulting from the first term on the right-hand side of equation 4, and $\rho(AR)$ is the ρ of equation 4. Fitting this equation to Beebe's data yields values of 0.12 for $\rho(VC)$, 0.69 for K , and 0.15 for $\rho(AR)$. We also fitted equation 5 to the data in Table 3 using nonlinear least squares. However, the results had such large confidence intervals as to be uninformative. In effect, we do not have enough data in our study to estimate equation 5.

One can interpret this pattern of correlations in the following way. The relatively constant correlation between year 1 and years 3 through 6 ($\rho(VC)$) could well represent a relatively constant rate of spending from chronic illness, and the higher correlation between years 1 and 2 ($\rho(AR)$) may represent acute events for which effects become negligible after a year.

If there is first-order autoregression (if $\rho(AR)$ is not equal to zero), one could do somewhat better in predicting period $t + 1$ than our estimate of the maximum R^2 suggests by an amount equal to $\rho(AR)^2$ times the variance of the estimated residual in period t . The estimate of $\rho(AR)^2$ from Beebe's data (0.0225) and our results in Table 3, however, suggest that this additional variance is small, on the order of 2 percent of the variance of v . (The value of 2 percent comes from the $\rho^2/(1 - \rho^2)$ formula.) Thus, although our estimates of the maximum R^2 are too small, they appear to be a good approximation.

Gains from better predictions

Let Y^* be the expenses for the i th enrollee; this is a function of individual characteristics x_i not taken into account in the AAPCC, characteristics a_i included in the AAPCC adjustment, and chance. Let $Z^*_i = bx_i + ca_i + e$. The AAPCC-adjusted payment is $K = E[Y^* \exp(-ca_i)]$. In what follows, we consider profits and losses after AAPCC adjustment by dealing with $Z_i = Z^*_i - ca_i = \log Y_i = bx_i + e$. To simplify calculations, we assume that:

$$\text{Var}(Y) = \text{Var}(Y^*), \quad (a)$$

when in fact $\text{Var}(Y)$ would be 0-5 percent smaller, depending on the power of the AAPCC adjustment.

Assume that bx_i is normally distributed with mean 0 and variance $\sigma(1)^2$, and assume that e is independent of bx and normally distributed with mean μ and variance $\sigma(2)^2$. Let $\sigma^2 = \text{variance}(Z) = \text{variance}(bx) + \text{variance}(e)$. Let $Yhat_i$ be the HMO's prediction of costs for person i . In our calculations of the gains from selective enrollment, we also assume that:

$$\text{Var}(Yhat) = R^2 \text{Var}(Y), \quad (b)$$

$$Yhat_i = E(Y_i), \quad (c)$$

$$Yhat \text{ is lognormally distributed,} \quad (d)$$

where R^2 is the R^2 of the HMO's prediction.

Let $R2$ be the additional variance explainable by the HMO (the R^2 of the HMO's prediction - the R^2 of the AAPCC prediction). Any lognormal $Y = \log X$ can be parametrized by its own mean M and variance S^2 or by the mean μ and variance σ^2 of the related normal variable X . The two parametrizations are related by $S^2/M^2 = \exp(\sigma^2) - 1$. If annual Medicare expenses Y^* are lognormally distributed with $M = 3,000$ and $S = 9,000$, then their log SD, σ , must satisfy $\exp(\sigma^2) - 1 = 9$. From assumptions a and b, $\text{Var}(Yhat \exp(-ca)) = R^2 \text{Var}(Y) = R^2 \text{Var}(Y^*)$, so $\exp(\sigma^2(1)) - 1 = 9R2$. This implies that $\sigma(1) = \text{the square root of } (\log(1 + 9R2))$. In any lognormal distribution, $M = \exp(\mu + \sigma^2/2)$, and so the mean occurs at $\sigma/2$ standard deviations above μ in the related normal distribution. Let $C(x)$ represent the cumulative normal distribution. Then the HMO optimally accepts the bottom $C(\sigma/2)$ of the distribution of predicted gains. Using the formula on

the moments of truncated lognormals (Aitchison and Brown, 1957, theorem 2.6), in all, the expenditures of the rejected top $1 - C(\sigma/2) = C(-\sigma/2)$ represent $1 - C(\sigma/2 - \sigma) = 1 - C(-\sigma/2) = C(\sigma/2)$ of the total spending. Thus, the profit per enrollee is (mean payment) $[1 - (C(-\sigma/2)/C(\sigma/2))]$. The percent enrolled and profits per enrollee are given for various values of $R2$ in Table 4.

References

- Aitchison, J., and Brown, J.A.C.: *The Lognormal Distribution: With Special Reference to Uses in Economics*. London. Cambridge University Press, 1957.
- Anderson, G., and Knickman, J.: Patterns of expenditure among high utilizers of medical care services: The experience of Medicare beneficiaries from 1974 to 1977. *Medical Care* 22(2):143-149, Feb. 1984a.
- Anderson, G., and Knickman, J.: Adverse selection under a voucher system: Grouping of Medicare recipients by level of expenditure. *Inquiry* 21(2):135-143, Summer 1984b.
- Anderson, G., Cantor, J., Steinberg, E., and Holloway, J.: Capitation pricing: Adjusting for prior utilization and physician discretion. *Health Care Financing Review*. Vol. 8, No. 2. HCFA Pub. No. 03226. Office of Research and Demonstrations, Health Care Financing Administration. Washington. U.S. Government Printing Office, Winter 1986.
- Beebe, J., Lubitz, J., and Eggers, P.: Using prior utilization to determine payments for Medicare enrollees in health maintenance organizations. *Health Care Financing Review*. Vol. 6, No. 3. HCFA Pub. No. 03198. Office of Research and Demonstrations, Health Care Financing Administration. Washington. U.S. Government Printing Office, Spring 1985.
- Berwick, D., Cretin, S., and Keeler, E.: *Children, Cholesterol, and Heart Disease*. New York. Oxford University Press, 1980.
- Brook, R. H., Ware, J. E., Jr., Rogers, W. H., et al.: Does free care improve adults' health? Results from a randomized controlled trial. *New England Journal of Medicine* 309(23):1426-1434, Dec. 8, 1983.
- Division of National Cost Estimates, Office of the Actuary, Health Care Financing Administration: National health expenditures, 1986-2000. *Health Care Financing Review*. Vol. 8, No. 4. HCFA Pub. No. 03239. Office of Research and Demonstrations, Health Care Financing Administration. Washington. U.S. Government Printing Office, Summer 1987.
- Duan, N.: Smearing estimate: A nonparametric retransformation method. *Journal of the American Statistical Association* 78(3):605-610, Sept. 1983.
- Duan, N., Manning, W. G., Morris, C. N., and Newhouse, J. P.: A comparison of alternative models of the demand for medical care. *Journal of Business and Economic Statistics* 1(2):115-126, Apr. 1983.
- Efron, B.: Regression and ANOVA with zero-one data: Measures of residual variation. *Journal of the American Statistical Association* 73(1):113-121, Mar. 1978.
- Howland, J., Stokes, J., III, Crane, S. C., and Belanger, A. J.: Adjusting capitation using chronic disease risk factors: A preliminary study. *Health Care Financing Review*. Vol. 9, No. 2. HCFA Pub. No. 03260. Office of Research and Demonstrations, Health Care Financing Administration. Washington. U.S. Government Printing Office, Winter 1987.
- Levit, K. R., Lazenby, H., Waldo, D. R., and Davidoff, L. M.: National health expenditures, 1984. *Health Care Financing Review*. Vol. 7, No. 1. HCFA Pub. No. 03206. Office of Research and Demonstrations, Health Care Financing Administration. Washington. U.S. Government Printing Office, Fall 1985.
- Lubitz, J., Beebe, J., and Riley, G.: Improving the Medicare HMO payment formula to deal with biased selection. In Scheffler, R., and Rossiter, L., eds. *Advances in Health Economics and Health Services Research*. Vol. 6. Greenwich, Conn. JAI Press, 1985.
- Manning, W. G., Newhouse, J. P., Duan, N., et al.: Health insurance and the demand for medical care: Results from a randomized experiment. *American Economic Review* 77(3):251-276, June 1987.
- McCall, N., and Wai, H. S.: An analysis of the use of Medicare services by the continuously enrolled aged. *Medical Care* 21(6):567-585, June 1983.
- McClure, W.: On the research status of risk-adjusted capitation rates. *Inquiry* 21(3):205-213, Fall 1984.
- Morris, C. N.: A finite selection model for experimental design of the Health Insurance Study. *Journal of Econometrics* 11(1):43-61, Sept. 1979.
- Newhouse, J. P.: Is competition the answer? *Journal of Health Economics* 1(1):109-115, May 1982.
- Newhouse, J. P.: Rate adjusters for Medicare under capitation. *Health Care Financing Review*. 1986 Annual Supplement. HCFA Pub. No. 03225. Office of Research and Demonstrations, Health Care Financing Administration. Washington. U.S. Government Printing Office, Dec. 1986.
- Newhouse, J. P., Manning, W. G., Morris, C. N., et al.: Some interim results from a controlled trial of cost sharing in health insurance. *New England Journal of Medicine* 305(25):1501-1507, Dec. 17, 1981.
- Pauly, M. V.: *Doctors and Their Workshops*. Chicago. University of Chicago Press, 1980.
- Physician Payment Review Commission: *Annual Report to Congress*. Washington, D.C. 1988.
- Rogers, W. H., and Newhouse, J. P.: Measuring unfiled claims in the Health Insurance Experiment. In Burstein, L., Freeman, H. E., and Rossi, P. H., eds. *Collecting Evaluation Data: Problems and Solutions*. Beverly Hills, Calif. Sage, 1985.
- Searle, S. R.: *Linear Models*. New York. John Wiley and Sons, 1971.
- Thomas, J. W., and Lichtenstein, R.: Functional health measure for adjusting health maintenance organization capitation rates. *Health Care Financing Review*. Vol. 7, No. 3. HCFA Pub. No. 03222. Office of Research and Demonstrations, Health Care Financing Administration. Washington. U.S. Government Printing Office, Spring 1986.

Thomas, J. W., Lichtenstein, R., Wyszewianski, L., et al.: Increasing Medicare enrollment in HMOs: The need for capitation rates adjusted for health status. *Inquiry* 20(3):227-239, Fall 1983.

Welch, W. P.: Medicare capitation payments to HMOs in light of regression toward the mean in health care costs. In Scheffler, R., and Rossiter, L., eds. *Advances in Health Economics and Health Services Research*. Vol. 6. Greenwich, Conn. JAI Press, 1985.