Risk Adjustment for a Children's Capitation Rate

Joseph P. Newhouse, Ph.D., Elizabeth M. Sloss, Ph.D., Willard G. Manning, Jr., Ph.D., and Emmett B. Keeler, Ph.D.

Few capitation arrangements vary premiums by a child's health characteristics, yielding an incentive to discriminate against children with predictably high expenditures from chronic diseases. In this article, we explore risk adjusters for the 35 percent of the variance in annual outpatient expenditure we find to be potentially predictable. Demographic factors such as age and gender only explain 5 percent of such variance; health status measures explain 25 percent, prior use and health status measures together explain 65 to 70 percent. The profit from risk selection falls less than proportionately with improved ability to adjust for risk. Partial capitation rates may be necessary to mitigate skimming and dumping.

INTRODUCTION

Enrollment of individuals in organizations such as health maintenance organizations (HMOs) that supply medical care for a fixed periodic premium or a capitated rate continues to grow; by 1991 enrollment in such organizations was about 13 percent of the population (National Center for Health Statistics, 1992). Moreover, the percentage of children enrolled is probably even higher because disproportionately few Medicare enrollees are enrolled in HMOs (2.8 percent), and because HMOs have typically covered maternity and well-child care with less cost sharing than insurance plans in the fee-for-service (FFS) system (McMillan, Lubitz, and Russell, 1987).

With growing HMO enrollment, more attention is being paid to the method for setting the rate at which the government and the private sector pay HMOs. In public-sector programs, most researchers have focused on Medicare's formula, the adjusted average per capita cost (AAPCC), and have consequently analyzed the behavior of those 65 years of age or over, or at least the behavior of adults (Anderson et al., 1986; Anderson and Knickman, 1984a, 1984b; Ash et al., 1989; Beebe, Lubitz, and Eggers, 1985; Gruenberg, Wallack, and Tompkins, 1986; Howland et al., 1987; Lubitz, Beebe, and Riley, 1985; McCall and Wai, 1983; McClure, 1984; Newhouse, 1986; Newhouse et al., 1989; Thomas and Lichtenstein, 1986a and 1986b; Thomas et al., 1983). However, 44 percent of the recipients in the other major public program, Medicaid, were less than 18 years of age in 1989, a figure that is likely to grow in light of the planned eligibility expansion for poor children (Reilly, Clauser, and Baugh, 1990). (Nonetheless, this group accounts for only 13 percent of Medicaid expenditure.) Moreover, many States are attempting to expand their use of capitated systems for the Medicaid population.

This research was supported by Cooperative Agreement Number 99-C-98489/9-04 from the Health Care Financing Administration. Joseph P. Newhouse is with Harvard University, Elizabeth M. Sloss and Emmett B. Keeler are with RAND, and Willard G. Manning, Jr., is with the University of Minnesota. The analysis and conclusions are those of the authors and do not necessarily reflect the views or policy positions of the Health Care Financing Administration, Harvard University, RAND, the University of Minnesota, or their sponsors.

The question naturally arises of how the capitated rate should be set for children.

If each capitated group enrolled a representative mix of health risks among the Medicaid population, the rate could simply be a fraction, or perhaps as much as 100 percent, of per capita FFS costs. However, it is unreasonable to expect that each group will do so. On the one hand, chronically ill patients who are under the care of an FFS physician will have incentives to continue with their physicians rather than to join an HMO; on the other hand, HMOs have incentives to avoid costly patients if they only receive payment that is based on the cost of the average patients. Even if health risks were distributed randomly, chance alone would cause some HMOs to have a mix of enrollees whose health characteristics differed from the population average. If all HMOs received the same payment per patient, there would be windfall profits and losses. Such profits and losses would, however, be more important for small HMOs because of the law of large numbers.

The natural approach to the issue of heterogeneous enrollees is to vary the amount paid an HMO according to an enrollee's expected use of medical services, i.e., to adjust the average rate. Indeed, Medicare makes such adjustments, thus the term adjusted average per capita cost (AAPCC). Specifically, it adjusts for age, gender, welfare status, institutional status, county of residence, and basis for Medicare eligibility (old age, disability, or end stage renal disease).

The Medicare set of adjusters has been widely criticized as inadequate. It is estimated that they may account for only 5 to 10 percent of the variation in expected cost—and much less of actual cost—across individuals (Lubitz, Beebe, and Riley, 1985; Newhouse, 1986; Newhouse et al., 1989). In our 1989 article, we examined how much additional adjusters, specifically measures of health and of use in the prior year, would improve the performance of an AAPCC-type formula for adults under 65 years of age. Using similar methods, we present herein results for children, although our results are limited to outpatient expenditures for children 5-13 years of age because our data set has few children that were hospitalized. Outpatient expenditures, however, account for 55 percent of total expenditures by children, compared with 38 percent for adults (Manning et al., 1987), and we see little reason to believe that the relative performance of various adjusters would change much if we had sufficient inpatient data to analyze (though the absolute amounts of explained variance may decrease).

We begin by estimating how much of the actual variation in expenditure one could potentially explain, and how much instead is because of random or unforeseeable events. Because adjusters cannot predict variation due to future random events, we wish to ignore the influence of such events as we assess the performance of adjusters, and want only to explain variation in expected, not actual, expenditure. Put another way, we estimate what the R^2 would be if we regressed actual expenditures on a set of almost ideal adjusters. We term this value the maximum explainable variance. As the explanatory power of a set of adjusters approaches the maximum explainable variance, the incentives for risk selection fall to negligible levels.

We then take up the relative performance of alternative adjusters. We begin with the analog for children of the demographic types of variables currently used in the AAPCC formula. We then estimate the gain from also using several measures of health status and prior use to adjust the capitation rate.

We analyze outpatient expenditures per child; that is, we analyze individual, not family, expenditure. One might argue that because families typically enroll as a group, we should have analyzed family behavior. The explanatory variables whose importance we assess, however, are at the individual level. Their relative importance would not have changed had we chosen to analyze data at the family level, but we would have had to impose additional assumptions to aggregate individual-level explanatory variables to the family level. Indeed, the appropriate assumptions are not at all obvious. However, because family-member expenditures are not independent, any incentive to skim or dump will be increased if families enroll as a unit. We further discuss this issue later.

MAXIMUM EXPLAINABLE VARIANCE

Expenditure cannot be fully or even largely predicted, so any set of risk adjusters will not explain all variance. Fortunately this does not cause a problem, as long as the HMO is paid the average expenditure for the unpredictable part of the variance. (Indeed, simply paying at the average for all is appropriate if all expenditures are unpredictable.) Problems potentially arise, however, if the HMO can determine that one person's expected expenditure (before the fact) exceeds another's. There is a financial incentive to enroll the low-cost person (skimming) if there is no adjustment in payment. Thus, one criterion for judging a set of adjusters is how well they explain expected expenditure or predictable variance. If all predictable variance is accounted for, there should be no skimming. To judge against this criterion, therefore, we need to know the variance in expected expenditure, which will be less than the actual or observed variance by the variance in unpredictable expenditure.

If children's medical expenditures correspond to the following simple model, it would be straightforward to determine the variance in expected expenditure, which is the amount of variance one could possibly explain in a regression of annual expenditure, using cross-sectional data:

$$\text{Expenditure}_{it} = X_{it}\beta + \mu_i + \varepsilon_{it}, \qquad (1)$$

where

X_{it} is a vector of risk adjusters, β is a vector of weights, i indexes the child, t indexes year, μ_i is an unobserved child-specific, time-invariant (stable) component of variance, and ε_{it} is a child-specific, time varying component of variance.

If the last term ε_{it} is random and cannot be predicted by the HMO or by the family, the maximum variance that could be explained is that accounted for by the first two terms, and we shall make this assumption. In fact, this is a lower boundary on the maximum explainable variance because some elements of ε_{it} might also be predictable. That is, there may be some time-varying variables, omitted from the X vector, that explain a non-trivial amount of variance. An example of a variable usually contained in ε_{it} rather than in X_{it} is an illness that has a partially predictable time pattern, such as leukemia. The spending rate will rise when a crisis occurs, and fall during remission. Nonetheless, it seems plausible to assume that most of the variation in ε_{it} is random. To the extent this is true, our estimate of the maximum explainable variance will be a good one. To the degree that ε_{it} is predictable, however, our estimates understate the amount of variation one can explain.

We have thus estimated the maximum explainable variance by estimating the proportion of variance in the μ_i term of the right-hand side of equation 1, assuming no adjusters (i.e., no X's); this is analogous to the R^2 from using a dummy variable for each person or to the proportion of variance that is between-person variance. To estimate the between-person variance, we subtracted an estimate of within-person variance from total variance, correcting for the bias from estimating within-person variance from a finite time series (Searle, 1971).

An alternative method for computing the maximum R^2 is to specify an X vector and estimate the amount of stable variation in the residuals. In principle, this method should lead to a higher maximum R^2 because the method described in the previous paragraph omits any variation for the $X\beta$ term from covariates that change over time. For adults, however, we found that the estimated maximum R^2 by the alternative method was less than the estimate using the method described in the previous paragraph (Newhouse et al., 1989). Thus, we used the latter method in this article.

We have used R^2 as a criterion variable, but some question the appropriateness of doing so. They argue that R^2 shows the goodness of prediction at the individual level, but that the formula only needs to predict well for groups (Lubitz, 1987). In other words, as long as the HMO receives adequate payment for its entire group of enrollees, the formula does not need to predict well at the individual level. This argument, however, ignores the behavioral incentives of the HMO, which can make more money by discouraging enrollment (or encouraging disenrollment) of any individual or family whose expected cost exceeds revenue. To blunt this incentive requires a premium that matches the expected cost of the HMO for each patient or family that it enrolls. Indeed, that is the reason for considering risk adjustment in the first place.

A different criticism is that R^2 may be on average high, but still perform badly for certain subgroups. That is, the payment formula may not fit well in some regions of the response surface, the functional form of the X vector may be specified incorrectly. This criticism has merit if one's purpose is to develop a specific payment formula, but it is less relevant for our purposes. Our aim is not to develop a specific formula, but only to compare in a gross way the performance of demographic, health status, and prior use variables. For that, the use of R^2 as a criterion seems adequate.

The model shown in equation 1 is not the only model one may investigate. Welch (1985) has proposed a model in which the errors follow a first order autoregressive process:

$$\mu_{it} = \rho\mu_{i,t-1} + u_{it}, \tag{2}$$

where

u_{it} is an independently and identically distributed random term and $-1 < \rho < 1$.

In equation 2, the potential explainable variance is that explained by the adjusters plus that explained by the first term on the right-hand side of the equation (because when one is predicting year t's expenditures, one has an estimate of μ_{it-1}). Thus, in this model the maximum R^2 is approximately the proportion of variance explained by the adjusters plus approximately $[1/(1-\rho^2)]\text{var}(\upsilon)$.

The consistency of equations 1 and 2 with the data can be tested straightforwardly by examining the pattern of correlation of the residuals over time. In the first model, 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 variance $[\mu]$/(variance $[\mu]$ + variance $[\varepsilon]$). In the second model, the correlation should decline geometrically (specifically, it should equal ρ^s). We later present results on the time pattern of the correlations in our data for children. For adults, the effect of regression to the mean, which equation 2 implies, appears modest (Newhouse et al., 1989).

Because our principal interest was to ascertain how useful various adjusters would be in further developing capitation rates rather than a particular payment formula, we have not performed a variety of specification tests for the accuracy of the functional form for the X vector. Thus, variables are simply entered in a linear form, and no tests for interactions have been performed. A more complex specification would probably improve performance, but it seems unlikely to change our qualitative conclusions. Moreover, going beyond a simple linear form risks overfitting our sample data, thereby distorting our results.

Using our results on explainable variance, we compute the HMO's expected profit from skimming or dumping. We consider a child whom the HMO predicts to be one standard deviation below or above the mean for expenditure based on the information available to it. We show how the profit or loss to the HMO diminishes as the payment formula changes to incorporate information from additional risk adjusters. In the limit, the additional risk adjusters would encompass all the information available to the HMO, and the HMO would not gain from selection.

DATA

The data we use come from the RAND Health Insurance Experiment, the design of which has been described in many places (Brook et al., 1983; Manning et al., 1987; Newhouse et al., 1981; Newhouse et al., 1993). This experiment randomly assigned families in six areas of the country—Seattle, Washington; Dayton, Ohio; Charleston, South Carolina; Fitchburg-Leominster, Massachusetts; and two rural areas, Franklin County, Massachusetts, and Georgetown County, South Carolina—to insurance plans that varied the cost sharing they faced. We have removed the effect in the sample of cost sharing from all observations because some of the variation in spending due to cost sharing was induced by the experiment; i.e., we have removed the between-plan variance. In effect, we ask how well various explanatory variables or adjusters account for within-plan variance. Thus, our results apply to an insured group with no variation in cost sharing, which is a good approximation to groups covered by capitation arrangements.

An aspect of the experiment that approximates capitation less well is that the experimental plans employed no utilization management techniques, such as pre-admission certification. This clearly raised the absolute level of spending relative to an HMO (Manning et al., 1984), and the question therefore arises as to whether results from this sample apply to children in a capitated group. Although we cannot be sure they do, it is not clear whether utilization management would much affect the proportion of variance explained by various personal characteristics. Unless utilization management techniques differentially increase the predictable portions of expenditure, the conclusions of this article, with respect to how well one can predict, are unaffected. Even if they were to increase the proportion of predictable variation, they would have to increase it differentially by type of covariate for our conclusions with respect to specific covariates to be changed. (The number of children 5-13 years of age enrolled in the HMO portion of the experiment, a little more than 200, were too few to use in this analysis.)

The families who participated in this experiment were randomly assigned to a 3-year or 5-year participation period, during which time the experiment acted as their insurance company. (They formally assigned to the experiment the benefits of any insurance for which they were eligible.) Independent verification of physician office claims indicates that the families filed claims with the experiment for more than 90 percent of their utilization; thus, we have a nearly complete record of utilization for the period of participation (Rogers and Newhouse, 1985).

The families invited to participate in the experiment were randomly selected, subject to some qualifications that are not important for this article. Specifically, the following groups were excluded: (1) those eligible for Medicare; (2) military personnel in active duty and retired; (3) veterans with service-connected disabilities; and (4) those institutionalized indefinitely (those in prison and in long-term psychiatric hospitals are the principal groups excluded by this criterion). Additionally, in five of the six sites (all but Seattle), low-income individuals were over-sampled to a limited extent.

We enrolled 1,844 individuals 13 years of age or under in the experiment. All those living at a given dwelling unit who met the eligibility requirements were offered enrollment. As a result, the utilization of the 1,844 observations are not all independent because the amount of utilization by children in the same family is positively correlated. Our calculations of the explained proportion of variance do not account for intrafamily correlation, but that should have little effect on our estimates of the proportion of variance that various types of individual characteristics can explain. In addition, as we will show, there is dependence over time within child. Indeed, the essence of the risk-adjustment problem is to account for the dependence in the residuals over time.

The sample used for the regression equation included only those participants who completed the study and the physical examination at exit. The reason for this is discussed later. Although 93 percent of those children who began the study completed it, our analysis sample is considerably smaller. Children who turned 14 years of age during the experiment were excluded from our analysis because they took an adult screening exam at exit, which differed from the children's exam. This criterion excluded about one-fourth of the children. Moreover, children under 5 years of age at exit were not given physiologic tests and were also excluded. This excluded another 10 percent of the children who began the study. In the regression analysis, we did not use children 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 those with any missing data for physiologic variables; these were mainly children who moved out of area during the experiment, and so did not have a hearing test as part of their out-of-area screening examination. In all, our sample consisted of 2,185 person-years. Only 84 of the person-years (3.8 percent) had inpatient use. Because inpatient use was so rare, we chose not to include it. Had we analyzed inpatient expenditure, we may well have overfit the data.

Dependent Variables

Our major interest was to predict annual expenditures per child on medical care services in constant dollars. For purposes of calculating the maximum R^2, we examined expenditure in both raw and trimmed form. The trim point was at the 98th percentile of total medical expenditure. For trimmed data, if an observation was in the upper 2 percent of the relevant distribution, it was set equal to the mean of the upper 2 percent of the observations. This preserved the overall mean.

Potential Adjusters

Because we wished to ignore within-plan variation, we began by regressing expenditures on plan. Plan was defined as the log of the nominal coinsurance rate plus a dummy variable for one particular plan (a plan with outpatient-only cost sharing). By design, the plan is approximately orthogonal to all other covariates (Morris, 1979). We then calculated:

$$(R^2[b] - R^2[a])/(1 - R^2[a]), \qquad (3)$$

where

a indexes the specification with only the plan variables included, and *b* indexes any of the more complete specifications. In fact, the plan variables explained only 1 percent of the total variance, so this correction is in practice unimportant.

We have used the sets of explanatory variables shown in Table 1 as adjusters. First, we included the demographic kinds of variables used by Medicare: age (entered linearly); gender; Aid to Families With Dependent Children (AFDC) status (Supplemental Security Income recipients are not in the sample population); and site. Then we added four different sets of variables to this basic set:

- Dichotomous Physiologic Health. A set of dummy variables that indicate 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 cutting point were dichotomized according to whether the test value was above or below the cutting point. For example, a dummy variable for anemia assumes the value one if a child 5-8 years of age has a hemoglobin below 11.0 g/ml. These physiological measures are derived from data collected at exit from the study rather than at entrance because only a random 60 percent of the children were given an exam at entrance. We felt we would obtain more accurate estimates by using data measured at exit for the entire sample than

Table 1
Definition of Health Status Measures

Measure	Definition
Physiologic Measures	
Anemia	Abnormally low hemoglobin, or current treatment for anemia, or physician diagnosis of anemia during past 12 months; 1 = present, 0 = absent
Low Hemoglobin	Measured automatically by the Coulter Model S machine (For children 5-8 years of age) = 0 if 11.0 g/100 ml or higher = 11.0 − X if less than 11.0 g/100 ml (For boys 9-11 years of age and girls 9-13 years of age) = 0 if 11.5 g/100 ml or higher = 11.5 − X if less than 11.5 g/100 ml (For boys 12 or 13 years of age) = 0 if 12.0 g/100 ml or higher = 12.0 − X if less than 12.0 g/100 ml
Hay Fever	Hay fever at any time since birth; 1 = present, 0 = absent
Hay Fever	Amount of time per year bothered by hay fever on a nautral log scale, ranging from 0 (none) to 6.4 (6 months or more)
Eczema or Chronic Rash	Rash during past 12 months lasting 3 months or longer, or physician diagnosis of eczema at any time since birth; 1 = present, 0 = absent
Asthma	Physician diagnosis of asthma at any time since birth; 1 = present, 0 = absent
Impaired Natural Far and Near Vision	Measured without corrective lenses for worse eye = 0 if between 10/20 and 20/20 = X-20 if 25/20 or higher
Impaired Hearing	Measured as simple average of thresholds at 500, 1,000, and 2,000 Hz, for worse ear = 0 if between 0 and 25 decibels = X-25 if 26 decibels or higher
Utilization Variables	
Whether Any Outpatient Expense in Prior Year	0 = No expense 1 = Positive expense
Whether Any Inpatient Expense in Prior Year	0 = No expense 1 = Positive expense
Logarithm of Outpatient Amount if Positive Outpatient Expenditure; Otherwise Zero	
Logarithm of Inpatient Amount if Positive Inpatient Expenditure; Otherwise Zero	
Subjectively Rated Health Status Measures	
Physical Limitations	Measures the presence of role or physical activity limitations, if any.[1]
Mental Health	Measures mental health at enrollment based on 12 item scale.[2]
General Health	Measures mother's rating of child's health based on 7 item scale.[3]

[1]A limitation was coded as present if APHY or AROLE in Sloss et al. (1986) was non-zero.
[2]See MHI in Sloss et al. (1986).
[3]See GHINDX in Sloss et al. (1986).

SOURCE: Newhouse, J.P., Harvard University; Sloss, E.M., and Keeler, E.B., RAND; Manning, W.G., University of Minnesota, 1993.

the data measured at enrollment for a partial sample despite the possibility that use would affect the observed value. The experimental results showed that plan did not affect these values (Valdez, Brook, and Rogers, 1985), and of course pre-experiment use could have affected the enrolled values.

- Continuous Physiologic Health. A set of variables that indicate the presence or absence of the physiologic conditions shown in Table 1, and for some conditions, a measure of the severity. Variables, again measured at exit, were included in the regression as defined in Table 1. For example, two variables related to anemia were included in the regression: (1) low hemoglobin, coded as the maximum of 0.0 or (for children 5-8 years of age) 11.0—hemoglobin value; and (2) the dummy variable for anemia described in the preceding paragraph.

In principle, the coefficient of the dummy variable measures the fixed costs of treating the condition, and the coefficient of the continuous variable measures the variable cost of increased severity. All variation on one side of a cutting point is suppressed. The cutting points reflect a judgment about values above or below which most physicians would not treat. For example, most physicians would probably not prescribe treatment for anemia with hemoglobin values above 11.0. For values differing from the cutting point in an unhealthful direction, we simply entered the physiologic measure linearly. It is quite possible, indeed probable, that the true functional form is non-linear, but theory does not specify a functional form, and we felt we would likely overfit the data if we experimented with non-linear functional forms. For the same reason, we did not explore interactions. For example, we treated the effect of being anemic and having asthma as additive. Any effort to create an actual payment formula using these variables would need to consider more complicated functional forms, though any such effort should employ a larger data set than the one used in this study.

At this point we note a possible extension, not undertaken in this article. The observed value of the physiologic health variables was used; thus, an individual who had a hemoglobin value of 12.0 g/100 ml achieved through medications was not distinguished from one who had a natural hemoglobin value of 12.0 g/100 ml and who was not under treatment. In effect, our specification implies that the expected treatment cost of a child will increase with less healthy values, but that will not always be true. Specifically, it will not be true if treatment alters the physiologic measure and less healthy patients use more resources (or if not all individuals are under treatment). Consider the above example of two persons with a hemoglobin value of 12. Although a physiologic condition that we measured was in fact responsible for treatment costs of one of the persons (i.e., the medication to raise the hemoglobin level), the physiologic variable we measured would not explain any variation in expenditure because it would be at 12.0 for both persons. Ideally one would measure what the value of the physiologic health measure (e.g., hemoglobin) would be if each child were untreated, but this cannot be observed.

An extension that partially allows for this difficulty is to enter a dichotomous variable for being in treatment. Incorporating such an adjuster has the additional advantage that the relevant information can 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 child may have a hemoglobin value of 10 g/100ml without treatment, whereas another may have a value of 10.5. If medication raises them both to 12.0 but the costs of treating the first person are greater, the cost difference would appear to the analyst as unexplained. This may be one reason that the measures of prior use described later achieve considerably more explanatory power than the measures of health status.

- Subjective Health. A set of measures of functional status or physical health, general health perceptions, and mental health as rated by a parent, usually the mother. Although the use of such variables as adjusters in a payment formula seems problematic because of the possibilities for fraud, we wished to ascertain the possible gains from using them in our data where there were no incentives for fraud. The same difficulty just described for the physiologic variables is present in these variables as well because medical care for a chronic problem can affect these measures, and medical care may be greater the more severe the problem. These variables were collected on all participants at entry into the study.
- Prior Year Use. Four variables measuring use of medical services in the previous year: whether there was any outpatient expenditure; whether there was any inpatient expenditure; and the logarithms of outpatients and inpatient expenditures.

Estimation Methods

To determine the gain from using various adjusters, we use a variant of a two-equation model we have used in other work (Manning et al., 1987; Duan et al., 1983), with the variables in Table 1 as explanatory variables. This variant models the probability of outpatient expenditures and the logarithm of outpatient spending, but then retransforms the logarithm to raw dollars using a non-parametric method (the smearing estimate) described in the articles cited.

Armed with our estimated equations (and estimated retransformation factor), we then compute the amount of explained variation as follows: We first predict the total outpatient expenditure of each person using the two-equation model. We then calculate a measure of R^2 due to Efron (1978) who uses the following formula:

$$\text{Efron's } R^2 = 1 - (\Sigma(y_i - \textit{y-hat}_i)^2)/(\Sigma(y_i - \textit{y-bar})^2 \quad (4)$$

where

y-hat$_i$ is the predicted y_i using the two-equation model with alternative sets of explanatory variables, and *y-bar* is the sample mean of y. Thus, the numerator of the expression in parentheses is the unexplained sum of squares, and the denominator is the total sum of squares. Note that this measure of R^2 can be negative when predicting from a non-linear model such as ours, but in this application it never was. We then compute the ratio of this R^2 to the maximum R^2 defined previously.

We have used the two-part model and Efron's R^2 rather than the more ordinary analysis of covariance because the two-part model has less tendency to overfit, and yields estimates with significantly lower mean square forecast error (Duan et al., 1983). Thus, use of analysis of covariance, which is common in the literature, overstates how well one can do.

RESULTS

The maximum R^2 for children is 37 percent for untrimmed data and 35 percent for trimmed data (Table 2). Because the results are not sensitive to use of trimmed or untrimmed data, we present only the untrimmed results. Age, gender, site, and AFDC status explain only about 6 percent of the explainable variance in expected outpatient costs, i.e., 6 percent of what one could hope to explain.

Measures of subjective health status do not much improve on the demographic variables of age, gender, site, and AFDC status. Only 14 percent of the explainable variance in expected costs is explained if those measures are also included. The physiologic measures of health, however, make a more noticeable impact. The variance in expected costs that is explained rises into the 25 to 30 percent range. Results using the continuous specification of the health variable differ little from those using the dichotomous version.

The four measures of prior use have a large effect compared with the other variables. They can account for more than one-half the explainable variance. When paired with measures of health, this fraction rises into the two-thirds region.

The year-to-year correlations decline as the time period extends (Table 3), but the decline appears to bottom out rather than continue geometrically, especially for ambulatory expenses. If there were simple regression to the mean, the values of the "average of diagonal" column should fall geometrically when reading from bottom to top (one is averaging increasingly longer intervals), but they clearly do not. Thus, there does not appear to be full regression toward the mean.

The expected gain from including measures of prior use is substantial in reducing incentives to select favorable risks.

Table 2

Percentage of Total and Maximum Explainable Variation in Ambulatory Care Expenditures with Alternative Specifications

Specification	Percent of Variation Explained: Total Variation	Percent of Variation Explained: Explainable
Percent of Between-Person Variance to Total Variance, Net of Plan Variance[1] (Maximum R^2)	36.6	100.0
Basic Specification: Age, Gender, Site, AFDC	2.1	5.7
Add to Basic Specification:		
Subjective Health	5.1	13.9
Dichotomous Physiologic Health	9.9	27.0
Continuous Physiologic Health	10.5	28.7
Subjective Health and Continuous Physiologic Health	11.2	30.6
Prior Year Use	20.7	56.6
Dichotomous Physiologic Health and Prior Year Use	23.4	63.9
Continuous Physiologic Health and Prior Year Use	23.4	63.9
Continuous Physiologic Health, Subjective Health, and Prior Year Use	23.9	65.3

[1]Plan explains 0.9 percent of total variance in the untrimmed model and 1.1 percent in the trimmed model.

NOTE: AFDC is Aid to Families with Dependent Children.

SOURCE: Newhouse, J.P., Harvard University; Sloss, E.M., and Keeler, E.B., RAND; and Manning, W.G., University of Minnesota, 1993.

HMOs have available to them information on prior use for all but new enrollees. As Table 4 shows, for a child one standard deviation from the mean, the gain or loss is $672 per child if HMOs know only the variance from prior use and our health measures, and only demographic variables are used to adjust premiums. This is about one-half the comparable figure for adults, but is enhanced by the usual practice of enrolling all children in a family because of the positive correlation among children. For a three-child family with an inter-child correlation of 0.25—the correlation between expenditures of children in the same family in the health insurance experiment data—the profit or loss would be 22 percent again as large as for a 3-child family with no correlation, or $1,426 at one standard deviation from the mean of 3-child families ($1{,}426 = 672 \sqrt{4.5}$).

DISCUSSION

Simple demographic variables such as age, gender, site, and welfare status account for only about 6 percent of the ex-

Table 3

Stability of Interyear Correlations[1]

Description	Year 2	Year 3	Year 4	Year 5	Average of Diagonal[2]
Ambulatory Expenses					
Year 1	0.573	0.373	0.315	0.301	0.301
Year 2		0.413	0.267	0.260	0.288
Year 3			0.330	0.255	0.298
Year 4				0.311	0.407
Probability of Any Ambulatory Expenses					
Year 1	0.328	0.320	0.144	0.170	0.170
Year 2		0.361	0.337	0.244	0.194
Year 3			0.259	0.215	0.291
Year 4				0.292	0.310

[1]Sample size for years 1-3 is approximately 850, and for years 4 and 5, approximately 240.
[2]0.407 in the fourth row, for example, is the average of 0.573, 0.413, 0.330, and 0.311.

SOURCE: Newhouse, J.P., Sloss, E.M., Manning, W.G., and Keeler, E.B., 1993.

Table 4

Annual Profit or Loss Per Child at One Standard Deviation from Mean with Varying Amounts of Differential Information[1]

Percent of Variance Explained in Rating Formula	Percent of Variance HMO Can Explain: 23.4[2]	36.6 (Maximum)
0 (No Adjustment)	[3]$704	$881
2.1 (Demographic Only)[4]	672	856
10.5 (Demographic + Health)[5]	523	745
23.4 (Demographic + Health + Prior Use)[2]	0	529

[1]Based on a standard deviation of expenditure of $1,457 among continuously eligible Medicaid children aged 5-13 years in New York State during a 12-month period between September 1985 and November 1986. The standard deviation among children aged 0-4 years is more than twice as great, $3,303.
[2]$23.4 = 0.639 \times 36.6$; the 0.639 and 36.6 values come from Table 2.
[3]These calculations assume the adjusted residuals have the same standard deviation as the raw expenses. Then the $672 figure in the second row, for example, is computed as follows. The HMO gain in variance explained is $0.234 - 0.021$ or 21.3 percentage points. Assuming a linear model, the HMO's predictions will have a standard deviation of square root $(0.213) \times (1{,}457) = 672$. Because the HMO's estimate is unbiased, rejecting a person who is predicted to be one standard deviation from the mean saves on expectation $672.
[4]$2.1 = 0.057 \times 36.6$; The 0.057 and 36.6 values come from Table 2.
[5]$10.5 = 0.287 \times 36.6$; The 0.287 and 36.6 values come from Table 2.

NOTE: HMO is health maintenance organization.

SOURCES: Newhouse, J.P., Harvard University; Sloss, E.M., and Keeler, E.B., RAND; and Manning, W.G., University of Minnesota, 1993; (Newhouse et al., 1989).

plainable variance in the expected annual ambulatory care expenditure of children. Indeed, they performed even less well than they did for adults, where they accounted for about 15 percent of the explainable variance of outpatient expenditure. It is perhaps not surprising that age would explain more variation in expenditure among those 14-65 years of age than among those 5-13 years of age, simply because of the greater variation in age among the adult group.

Despite their poor performance, this set of demographic variables is an administratively straightforward one to include in a payment formula. Put another way, although there is little reason not to vary capitation rates based on these demographic characteristics, simply adjusting for them will not solve the problem of heterogeneity among enrollees. One is far from having a satisfactory set of adjustments.

A frequently advocated step is to add measures of health status as adjusters (Howland et al., 1987; McClure, 1984; Thomas and Lichtenstein, 1986a, 1986b; Thomas, Lichtenstein, and Wyszewianski, 1983). Those measures of health status that we included did not much help. In the case of physiologic measures, this could be because our list of variables was quite short, limited to six types of medical problems. Nonetheless, most remaining chronic diseases affecting children are not very prevalent. Because the percentage of variance explained is proportional to the percentage of children with the problem, rare problems cannot explain much variance. Furthermore, a much longer list of physiologic variables used for adults did not perform markedly better.

More surprising is the poor performance of subjective measures of health status. Although few children in a general population are physically limited or have serious mental illness, the general health index has been shown to be a reliable and valid measure of child health, and its distribution is not as non-normal as the other measures (Eisen et al., 1980). It is striking that even a quite limited set of six physiologic conditions measured at exit can explain notably more variance in outpatient expenditure than the three subjective measures taken at enrollment. This finding also held for adults; subjective measures performed less well than physiologic measures.

From a practical point of view, it may not be so serious that subjective measures of health status are of little help. Implementing them as part of a payment formula would pose possibly insuperable problems because of the possibility of fraud. For that reason, and because we believe the scope for improving our physiologic measures is considerably greater than the scope for improving our subjective measures of health, we suggest that in the case of children the preponderance of any further effort spent to develop health-status adjusters be devoted to physiologic measures.

Although none of the measures of health status performed very well, measures of prior use did. More than one-half the variation in expected ambulatory care costs (i.e., the stable variance) could be explained by our four measures of prior use. Indeed, prior use did even better for children than it did for adults, where the comparable figure was somewhat less than one-half.

Incorporating measures of prior use into the payment formula poses both a mundane and a philosophical issue. The former is simply the issue of data availability. Many capitated organizations do not have readily available data on outpatient use; thus, there would be a serious implementation problem.

The philosophical issue relates to the appropriateness of paying on the basis of prior use. One commonly heard argument for capitation is that the FFS fee structure is distorted (i.e., it is profitable for physicians to treat more intensively, and they can act on these incentives because of consumer ignorance) and that these distortions can be internalized to the organization by means of capitation.

There are two problems with this argument. Although the existing fee structure may well be distorted (i.e., many fees depart markedly from marginal cost), capitation means that the marginal revenue of an additional service is zero. Thus, the implicit fee structure associated with capitation is also distorted. This produces an incentive to shirk on the contract to deliver the "necessary" services that were contracted for, unless these services attract or retain healthy (low cost) patients who pay an average rate or unless they reduce future expenditure. (Preventive care is an example of a service that may both attract healthy patients and reduce future expenditure.) Against the point-in-time incentive to underserve are the possible consequences of current enrollees' withdrawing and potential damage to the organization's reputation, reducing the number of new enrollees. These threats, however, rely upon information, and information may not be very good. If the consumer is not knowledgeable enough to detect overservice from an FFS physician, will he or she be knowledgeable enough to detect underservice? Indeed, the potential of capitated organizations to underserve has been widely noted (Pauly, 1980).

The second problem with the philosophical argument is that it simply ignores the mismatch between an enrollee's cost to the HMO and the revenue received for that enrollee, if the revenue is not tailored to the individual's characteristics. Thus, it does not address the incentive of the capitated organization to seek good risks and the resulting market failure for the poor risk (Table 4).

Including prior use in the payment formula mitigates both problems—the incentive to underserve an enrollee, and the incentive to select good risks. It mitigates the problem of incentives at the time of use by paying some positive amount, provided the child remains an enrollee in the next year. It also addresses the mismatch between marginal revenue and cost at the individual level, as the results in Table 2 demonstrate; i.e., prior use picks up unmeasured variation across individuals in the amount of current use. Thus, its inclusion reduces the incentives to select patients whose expected costs are less than average. For the same reason, it does not equally pay HMOs whose mixes of health risks differ, whether through deliberate action on the HMO's part or simply through random events.

Including prior use and our measures of health status raises the proportion of explained variance to the 65 to 70 percent range, but, as Table 4 shows, there are nonetheless substantial profits to be made from selective enrollment and disenrollment. If one wants to do better, one will have to include a measure of current use in the payment formula (Newhouse, 1986). For example, payment could be a weighted average of an adjusted capitation rate and a current use payment. Basing the formula on current use rather than prior use also means the HMO does not lose revenue if a high-spending child disenrolls.

The objections many have to including any measure of use in the payment formula are twofold. First, the incentives for efficient production are weakened. This is true, but points up that we are dealing with a second-best solution. Second, many feel FFS is inherently a flawed system that provides incentives for overservice. As Pauly (1980) pointed out, the incentive is not inherent. Ignoring the moral hazard on the patient's side, a fee at the marginal cost of delivering the service would provide the physician (or capitated organization) no incentive for overservice. Paying a partial capitation and a partial fee reduces the likelihood that the fee will induce overservice. The moral hazard can be addressed through the incentives of capitation to the provider. In short, systems that vary payment with the amount of use may provide approximately correct incentives.

Although one can conceptualize fees at marginal cost, in practice a regulator will not know the true marginal cost. Thus, one cannot claim a priori that a mixed payment system, part FFS and part capitation, will definitely improve on either pure system, but our results suggest it might. We conclude that some experimentation with a mixed payment system is indicated (Selden, 1990; Newhouse, 1991).

REFERENCES

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

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, 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, 1984b.

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

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

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* 309:1426-1434, December 8, 1983.

Duan, N., Manning, W.G., Morris, C.N. et al.: A Comparison of Alternative Models of the Demand for Medical Care. *Journal of Business and Economic Statistics* 1(2):115-126, April 1983.

Efron, B.: Regression and ANOVA with Zero-one Data: Measures of Residual Variation. *Journal of the American Statistical Association* 73:113-121, 1978.

Eisen, M., Donald, C.A., Ware, J.E., Jr. et al.: *Conceptualization and Measurement of Health for Children in the Health Insurance Study.* Santa Monica: The RAND Corporation, (Pub. No. R-2313-HEW), 1980.

Gruenberg, L., Wallack, S.S., and Tompkins, C.P.: Pricing Strategies for Capitated Delivery Systems. *Health Care Financing Review* 1986 Annual Supplement. Pp. 35-44, December 1986.

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

Lubitz, J.: Health Status Adjustment for Medicare Capitation. *Inquiry* 24:362-375, 1987.

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, CT. 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):251-277, June 1987.

Manning, W.G., Leibowitz, A., Goldberg, G.A. et al.: A Controlled Trial of the Effect of a Prepaid Group Practice on Use of Services. *New England Journal of Medicine* 310:1505-1510, June 7, 1984.

McCall, N., and Wai, H.: An Analysis of Use of Medicare Services by the Continuously Enrolled Aged. *Medical Care* 21:567-585, 1983.

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

McMillan, A., Lubitz, J, and Russell, D.: Medicare Enrollment in Health Maintenance Organizations. *Health Care Financing Review* 8(3):87-93, 1987.

Morris, C.N.: A Finite Selection Model for Experimental Design of the Health Insurance Study. *Journal of Econometrics* 11:43-61, 1979.

National Center for Health Statistics: *Health and Prevention Profile, United States, 1991.* Washington. U.S. Government Printing Office, 1992.

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

Newhouse, J.P.: Pricing and Imperfections in the Medical Care Marketplace. In Zweifel, P., and Frech, H., eds. *Health Economics Worldwide.* Amsterdam. Kluwer, 1991.

Newhouse, J.P., and The Insurance Experiment Group: *Free For All? Lessons from the RAND Health Insurance Experiment.* Cambridge, MA. Harvard University Press, 1993.

Newhouse, J.P., Manning, W.G., Keeler, E.B. et al.: Objective Measures of Health and Prior Utilization as Adjusters for Capitation Rates. *Health Care Financing Review* 10(3):41-55, Spring 1989.

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* 305:1501-1507, December 7, 1981.

Pauly, M.V.: *Doctors and Their Workshops.* Chicago. University of Chicago Press, 1980.

Reilly, T., Clauser, S., and Baugh, D.: Trends in Medicaid Payments and Utilization, 1975-1989. *Health Care Financing Review* 1990 Annual Supplement. Pp. 15-33, December 1990.

Rogers, W.H., and Newhouse, J.P.: Measuring Unfiled Claims in the Health Insurance Experiment. In Burstein, L., Freeman, H.E., and Rossi, P., eds. *Collecting Evaluation Data: Problems and Solutions.* Beverly Hills. Sage, 1985.

Searle, S.R.: *Linear Models*, New York. John Wiley and Sons, 1971.

Selden, T.: A Model of Capitation. *Journal of Health Economics* 9:397-409, 1990.

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

Thomas, J.W., and Lichtenstein, R.: Functional Health Measure for Adjusting Health Maintenance Organization Capitation Rates. *Health Care Financing Review* 7(3):85-95, Spring 1986a.

Thomas, J.W., and Lichtenstein, R.: Including Health Status in Medicare's Adjusted Average per Capita Cost Capitation Formula. *Medical Care* 24:259-275, 1986b.

Thomas, J.W., Lichtenstein, R., Wyszewianski, L. et al.: Increasing Medicare Enrollment in HMOs: The Need for Capitation Rates Adjusted for Health Status. *Inquiry* 20:227-239, 1983.

Valdez, R.B., Brook, R.H., Rogers, W.H.: The Consequences of Cost Sharing for Children's Health. *Pediatrics* 75:942-951, 1985.

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, CT. JAI Press, 1985.

Reprint requests: Joseph P. Newhouse, Ph.D., 25 Shattuck Street, Parcel B, 1st Floor, Boston, Massachusetts 02115.