



RISK ADJUSTMENT AND MEDICARE

Joseph P. Newhouse, Melinda Beeuwkes Buntin,
and John D. Chapman
Harvard University

Revised
June 1999

The authors are grateful to The Commonwealth Fund for research support. No endorsement of the findings by the Fund is implied. The authors would like to thank Randy Ellis for helpful comments on a draft of the original version.

Corresponding author:

Joseph P. Newhouse
Division of Health Policy Research and Education
Harvard University
180 Longwood Avenue
Boston, MA 02115
617-432-1325 FAX: 617-432-3503
newhouse@hcp.med.harvard.edu

EXECUTIVE SUMMARY

Almost all observers agree in principle that the Medicare program needs better risk adjusters than the adjusted average per capita cost (AAPCC) methodology it has historically used in calculating payments to at-risk health care plans. Indeed, Congress mandated changes in risk adjustment methods as part of the 1997 Balanced Budget Act.

The lack of adequate risk adjustment results in favorable risk selection of beneficiaries into health plans, which in turn causes three problems: potential barriers to access for beneficiaries seen as bad risks; incentives for plans to spend resources toward ends that are not socially productive; and increased costs to the Medicare program. Strong evidence suggests that, as a group, at-risk plans have benefited from favorable risk selection; that is, they have disproportionately attracted low-cost enrollees. Additionally, payment of a lump sum, independent of use, leaves a financial incentive to stint on services.

Adequate risk adjustment would cause payment to match expected individual spending closely. The adjusters in the current AAPCC formula explain only about 1 percent of the variance in actual spending, while an ideal adjuster might explain 20 times more variance. Fortunately diagnostic-based adjustment techniques such as hierarchical coexisting conditions (HCCs), diagnostic cost groups (DCGs), and ambulatory care groups (ACGs) have advanced to the point of making a substantial improvement over the AAPCC. In terms of ability to explain variance prospectively using Medicare data, the HCC model appears to do better, explaining about 8 to 9 percent (Ellis et al., 1996).¹ By contrast, ADGs explain about 6 percent (Weiner et al., 1996), as do PIP-DCGs. Although these figures are still a fair distance from the 20-25 percent lower bound, each is clearly a substantial improvement on the AAPCC.

In addition to explained variance, both Ellis et al. and Weiner et al. (1996) present data on how these risk adjusters affect profitability, thereby suggesting both the improvement that they make and the magnitude of the remaining problem. For example, Ellis and colleagues calculate that plans would experience substantial average profits and losses in 1992 from enrolling those in the lowest and highest quintiles of 1991 expenditures, assuming that 1992 spending by the plan would be the same as spending in traditional Medicare (table 1).

Table 1. Effect of the AAPCC and HCC Adjusters on Average Estimated Profit and Loss

Subgroup	Profit or Loss in 1992: AAPCC Adjusters	Profit or Loss in 1992: HCC Adjusters
Top 20% of Spenders, 1991	-\$4,425	-\$1,311
Bottom 20% of Spenders, 1991	+\$2,134	+\$424

Source: Ellis et al., 1996. Assumes plan was paid 100% of the AAPCC.

As table 1 shows, the HCC method yields a noteworthy reduction in incentives to cream and dump, cutting profits and losses by more than a factor of three relative to the AAPCC. Nonetheless, HCCs leave nontrivial levels of profits and, especially, losses.² (For purposes of scaling these figures, mean spending in 1992 was \$3,800.)

The Health Care Financing Administration (HCFA) has announced its intention to proceed with a variant of the DCG method beginning January 1, 2000. Even this method, however, leaves two potential problems: there is still an incentive, though substantially reduced, to select good risks; and there is still an incentive to stint on care. Moreover, the incentive to avoid bad risks is increased by HCFA's inability in the short term to include diagnostic information from outpatient settings in the risk adjustment formula.

Recent research has also documented substantial underpayment for frail elderly beneficiaries, even after diagnosis-based risk adjustment. To deal with this problem, the Medicare program would need to adopt risk adjustment based on functional status—which in turn poses difficult implementation problems. Similarly, once Medicare moves to an annual rather than a monthly enrollment period in 2002, payment for new nursing home entrants who are also eligible for Medicaid appears much too low.

All these problems would be mitigated by moving policy away from exclusive reliance on a lump-sum, risk adjusted payment. Partial capitation—under which payment for an individual enrollee would combine capitation methods and some reflection of that person's actual use of services—would address both selection and stinting incentives. A somewhat analogous but less desirable approach would be to require that capitation payments carry a reinsurance or a formal outlier policy, under which very expensive cases would be reimbursed an additional amount. Another proposal, potentially useful for terminally ill patients, would allow plans to name in advance a percentage of high-cost cases for which they would be partially or wholly reimbursed using traditional Medicare.

As the Medicare at-risk program grows, lack of satisfactory risk adjustment is causing serious problems. Better risk adjusters now exist. For valid reasons, HCFA is proposing to introduce them into the program slowly. Even when fully implemented, however, it is highly optimistic to think that current problems of selection will fall to negligible levels. Taking some account of the actual use of services will likely be necessary to keep risk selection negligible.

RISK ADJUSTMENT AND MEDICARE

INTRODUCTION

Almost all observers agree in principle that the Medicare program needs better risk adjusters than the adjusted average per capita cost (AAPCC) methodology it has historically used in calculating payments to at-risk health care plans. Indeed, Congress—the most influential “observer” of all—mandated changes in risk adjustment methods as part of the 1997 Balanced Budget Act (BBA).

In January 2000, the Medicare program will begin to phase in a set of diagnosis-based risk adjusters proposed by the Health Care Financing Administration (HCFA), including an interim adjuster based on certain diagnoses of hospitalized patients. The planned changes should produce some improvements, yet they also raise serious questions of effectiveness and the possible creation of new, unwanted incentives. An alternative method—partial capitation—would address many of the problems of the current approach while increasing the accuracy of the proposed system.

THE PROBLEM OF FAVORABLE SELECTION

The lack of adequate risk adjustment creates incentives for favorable selection of beneficiaries into at-risk plans. Favorable selection, in turn, causes three major problems that affect both individuals and government: potential access barriers for beneficiaries seen as bad risks; incentives to spend resources to attract good risks; and increased costs to the Medicare program.

Under current methods of reimbursing plans, people who are bad risks—especially chronically ill beneficiaries—face potential access problems because health plans are reimbursed the same amount for beneficiaries with very different levels of expected treatment costs. Without adequate risk adjustment, plans have a financial incentive to avoid marketing to patients likely to incur high costs and to induce more expensive patients already enrolled to disenroll. Plans also have reason to configure themselves so as not to appeal to bad risks: under current reimbursement arrangements, a reputation for offering the best cancer service in town is not healthy for the bottom line. Further, in a local market where no plan is discriminating against bad risks, a new entrant may attempt to do so. Then, if many good risks switch to the new entrant, existing plans may have to choose between discriminating against bad risks themselves and losing money.³

Further, although it makes financial sense for a plan to spend up to the amount of profit on a good risk to attract such a person, resources used toward that goal represent an inefficient use of funds.⁴ A subsidized health club membership, for example, is an added benefit that also disproportionately attracts good risks. Yet good risks probably do not value the

benefit as much as it costs the health plan to provide it or they would already have bought a membership. At the same time, plans have an incentive to spend money to avoid bad risks, an expenditure that is not socially productive.

Finally, if Medicare pays a plan more to treat a person than it would have paid had that person remained in traditional Medicare, government spending by definition has increased. Most analysts, including the General Accounting Office (GAO), believe that such overspending has occurred (GAO, June 1999).

Under the AAPCC formula, plans are paid 95 percent of average spending for “comparable” patients in traditional Medicare, defined as people of the same age, sex, welfare status, institutional status, and employment status. Within a broad demographic category, then, plans have an incentive to attract enrollees whose expected spending is less and avoid people whose expected spending is more than the payment established for the group. An incentive may be acted upon to a greater or lesser degree, but the evidence suggests that plans have disproportionately attracted beneficiaries whose expected spending is less than AAPCC-generated payment.

Strong evidence of favorable selection comes from analyses of the Physician Payment Review Commission (PPRC, 1996). The commission analyzed a random sample of more than 40,000 persons who had enrolled in at-risk plans between 1989 and 1994 and compared their spending with that of a control group of individuals who remained in traditional Medicare. The comparison was made for periods when individuals in both groups were enrolled in traditional Medicare. The commission also calculated how much of the spending difference could be accounted for by variables in the AAPCC, a technique analogous to controlling for those variables.

The commission found that, during the six months preceding enrollment, the group that entered at-risk plans spent 63 percent as much as the average Medicare beneficiary. After adjustment by the AAPCC variables, this figure actually fell to 62 percent. People who disenrolled and returned to traditional Medicare spent 160 percent as much as the average beneficiary in the six months following disenrollment. After adjustment using the variables in the AAPCC, the figure was 142 percent. In other words, below-average spenders were enrolling and above average spenders were disenrolling, while the AAPCC risk adjustment categories failed to account for more than a small percentage of the variation.

The commission also looked at mortality in the years following enrollment. After adjusting for age and sex, the group in at-risk plans had lower mortality for the first three years of enrollment. During the first six months after enrollment, the mortality difference between new enrollees and a control group matched by age and sex was a hefty 25 percent. Although

some of the mortality difference could theoretically be due to better care in the at-risk plans, selection seems a more likely explanation for most of the difference.

These large differences in spending and mortality tend to shrink as an individual remains in the at-risk plan, but the evidence is conflicting on whether or not the differences are ever fully eliminated. Our reading, and the analysis of others, suggests that they are not (Welch, 1985; Newhouse et al., 1989; Physician Payment Review Commission, 1996).

Riley and colleagues (1996) generated further evidence through their analysis of self-reported health status from the 1994 Medicare Current Beneficiary Survey. Consistent with favorable selection, they found that self-reported health was notably better among health plan enrollees; for example, 18 percent of health plan enrollees reported their general health to be fair or poor, compared with 26 percent of enrollees in traditional Medicare.

PRACTICAL CHALLENGES IN THE DEVELOPMENT OF RISK ADJUSTERS

An ideal risk adjuster would modify payment to match an individual's expected spending: that is, what a person with particular characteristics and average luck would likely spend in the future. A set of adjusters capable of causing payment to match expected individual spending would need to account for the variance in expected spending across individuals.

The variance in expected spending differs from the variance in actual spending according to the amount of luck. Thus, a person with obstructed coronary arteries is at higher risk of a heart attack than a similar person with no obstructions and so has higher expected spending. If the risk is not much increased—because, for example, the arteries are not much obstructed—expected spending is not much increased. Actual spending, however, will be much increased if a heart attack actually occurs (and the person does not die immediately from the attack). Even so, a good risk adjuster would not need to explain variance in actual spending due to luck; purely random effects on spending, such as being hit by a drunk driver, should not cause plans either to desire or to shun a particular enrollee.⁵

The demographic adjusters in the current AAPCC formula explain about 1 percent of the variance in actual spending. One percent is better than nothing, but not much better. Expected spending levels for two 70-year-old, noninstitutionalized women—one with colon cancer and one with no chronic illness—are very different, yet the present formula pays the same amount for each. A health plan contemplating the enrollment of these two women clearly faces very different financial incentives with respect to each.

If 1 percent does not get the job done, what percentage of the variance in actual spending should an adequate risk adjuster explain? This is a difficult question for three reasons:

- The percentage that is ideally possible to explain cannot be estimated precisely. At best, we can hope to establish a lower bound on the relevant figure.
- It is unclear how accurate a predictor must be to make the rewards from selection less than the costs of engaging in it; in other words, it is not known to what degree plans can or will act on changes in incentives.
- The adequacy of an adjuster presupposes agreement on appropriate treatments and hence the relative expense of various diseases. In many cases, that agreement is lacking.

These three questions represent major technical challenges for the creation of an adequate system of risk adjustment. Readers interested primarily in policy implications may wish to skip the following discussion.

Identifying a Lower Bound for Explaining Variation

What we wish to know—the variance in expected spending—we do not directly observe, which is why the literature analyzes variance in actual spending.⁶ We can, however, try to estimate the variance in expected spending from observations on actual spending. This is commonly done by dividing the variance in actual spending into three components, the first two of which comprise the variance in expected spending.

The first component is termed a fixed effect, or a constant amount that puts a person above or below the community mean indefinitely. An example would be a person with a chronic disease, such as diabetes, who could be expected to incur spending above the mean indefinitely. The second component is a predictable but time-varying effect; that is, a factor that would put the person above or below the community mean in the following month or year but that would not persist indefinitely. An example would be a person who had stopped smoking and whose risk of cancer from smoking declines each year. The remaining variance, the third component, is by definition random.

The first, or fixed, effect can be approximated by estimating the mean spending of a person over a period of years. Thus, using a person's mean spending over the past several years as an estimate of expected spending for the following year, one can estimate what percentage of the actual variance in spending across people can be explained by a fixed person-specific effect. There is reasonable consensus in the literature that the proportion of variance in actual annual spending that could be explained in this fashion is on the order of 15-20 percent (Newhouse, 1996).

In addition, if a person's spending last year was above the person's long-term average, it is more likely than not that this year's spending will also be above average; conversely, if last

year's spending was below average, next year's is likely to be low. Again, there is something of a consensus that if one knows last year's spending, and to a lesser degree spending in the years preceding, one could explain another 3-5 percentage points of variance above the 15-20 percent that the fixed effect can predict. Thus, it appears that anyone observing past spending of a given person could explain around 20-25 percent of the variance in annual spending in a given year.

This value is, however, a lower bound on the ability to predict future spending because other predictive factors, such as a pregnancy or a recent diagnosis of cancer, may be observed but not reflected in past spending. In other words, plans or physicians accepting capitation could potentially predict more than the 20-25 percent, although how much more is unclear. That ambiguity did not matter in appraising the adequacy of the AAPCC; clearly, if at least 20-25 percent of the actual variance is predictable, a formula that explains only around 1 percent leaves substantial incentives for creaming and dumping. As adjusters improve, however, ambiguity will begin to matter more.

Assessing the Impact of Incentives on Plans' Actions

Regardless of how good risk adjusters become, we will never have a perfect formula, only better or worse approximations of it. Thus, the question arises of how close to perfect the formula must be to reduce plans' incentive and ability to seek favorable enrollment to negligible or tolerable levels. The short answer is that we do not know.

Although evidence shows that favorable selection is occurring, actions to enroll better risks can be costly. As risk adjustment formulas improve and the profitability of creaming and dumping correspondingly diminishes, plans may not find it worthwhile to spend resources to take such actions (Newhouse, 1996). Moreover, better risk adjustment formulas are also likely to change the profile of "good" and "bad" risks. Good risks will no longer necessarily be healthier persons in general, but rather those who are healthier within each class recognized in the formula. Such a change may materially alter the cost or difficulty of achieving a given level of favorable selection. Thus, a workable adjuster need not achieve the ideal, but how close to perfection will be sufficient is unknown.⁷

Accounting for Variations in Treatment Methods

Estimates of use at the individual level have not typically been available for enrollees in the at-risk program. As a result, most risk adjustment methods have been calibrated using data from fee-for-service providers, either from traditional Medicare or from nonelderly individuals. The question therefore arises of whether the relative weight accorded to different classes of enrollees by current or suggested adjusters is appropriate; that is, are weights calibrated on indemnity plan patterns suitable for at-risk plans?

To the degree that at-risk programs simply reduce use proportionately across all persons, relative weights calculated from indemnity plan data would be unaffected, but such proportionate reductions are unlikely. Cutler, McClellan, and Newhouse (1997), for example, find that nonelderly persons enrolled in Massachusetts HMOs and who had a heart attack received approximately the same proportion of catheterization and revascularization as those enrolled in an indemnity plan.⁸ That is, use of these invasive procedures seems little affected by HMO enrollment.

By contrast, Frank and colleagues (in press) find that acute depression is treated much less intensively in managed behavioral health care, although, as far as is known, outcomes appear to be similar. Similarly, the RAND Health Insurance Experiment found 40 percent fewer hospital admissions (across all diagnoses) among a group of persons randomly assigned to a staff model HMO when compared with a similar group enrolled in a comparable indemnity plan (Newhouse and the Insurance Experiment Group, 1993). Together, these findings indicate not only that at-risk plans treat conditions differently than indemnity plans, but that the reduction in services will not be spread uniformly across conditions. Likewise, to the degree that organized plans either prevent disease better, detect it at an earlier stage, or better coordinate care, those effects will likely be dissimilar by age group or disease class.

One can draw two conclusions from the disproportionate reductions in use across classes of people. First, most existing risk adjusters, whether based on age, diagnosis, or health status, are distorted in a world of managed care because they have been calibrated on “unmanaged” patterns of treatment. That is, even if 95 percent of the AAPCC were the right aggregate payment level, there would be profits and losses on particular subgroups with resulting incentives to favor or disfavor those groups. (The analog in Medicare’s Prospective Payment System is getting the weights across DRGs right.) Second, and in some ways more important, risk adjustment by definition sets the payments for different classes of people—for example, people with and without certain diagnoses, if adjusters based on diagnoses are used. As such, risk adjustment embodies a judgment about how individuals with a certain diagnosis should be treated if there are treatment options of varying cost.

Most existing research on risk adjusters does not face up to this problem; rather, the work simply assumes that the relative resource allocation across groups of persons implied by existing fee-for-service treatment is appropriate. Of course, it is also unlikely that any two at-risk plans would treat conditions similarly, so simply using data from at-risk plans rather than traditional Medicare to calculate weights is no assurance that the resulting relative weights would be appropriate.

A Practical Standard (and a Technical Caveat)

In sum, an ideal risk adjuster that used no information from current or past utilization to set rates (that is, was prospective in nature) should explain at least 20-25 percent of the variance in spending across beneficiaries. In addition, a plan could observe factors that would predict next year's spending but would not affect this year's spending or would affect it minimally. As a result, the actual portion of variance that might be predictable could be higher than 20-25 percent, perhaps much higher.

It should be pointed out, however, that if the preceding analysis had been based on monthly data, the 20-25 percent lower bound figure would be smaller. This discussion has been couched in terms of annual spending, since that is the time period the literature has used.

Until 2002, Medicare regulations will allow beneficiaries to enroll in and disenroll from at-risk plans monthly.⁹ In general, monthly spending is more difficult to estimate than annual spending because random or unpredictable events play a proportionately larger role in a shorter period: an individual is not likely to have, for example, a serious car accident every month. Although policy conclusions drawn from analyses of annual data would probably hold if monthly data were used instead, demonstrating this empirically remains for future work. Beginning in 2003, however, this technical caveat will be of little importance.

THE STATE OF THE ART OF RISK ADJUSTMENT: RECENT ADVANCES

Five years ago, Newhouse (1994) assessed the state of the art in risk adjustment as primitive. Since that time, HCFA and others have continued to support the development of risk adjustment methods, particularly by incorporating information on diagnosis and self-reported health status. The following selective review covers recent advances, focusing on the Medicare population, and considers problems implicit in current assumptions within the field. (For a more thorough review of the literature, see van de Ven and Ellis [in press].)

Incorporating Information on Diagnosis

Two groups of researchers have done most of the work on incorporating diagnosis in risk adjusters for the Medicare population. A team based at Boston University initially developed the diagnostic cost group (DCG) model, which has now evolved into the hierarchical coexisting conditions (HCC) model and the principal inpatient procedure-diagnostic cost group (PIP-DCG) model (Ash et al., 1989; Ellis and Ash, 1995/96; Ellis et al., 1996). A group at Johns Hopkins University has developed the ambulatory care group (ACG) and ambulatory diagnosis group (ADG) models (Weiner et al., 1996).¹⁰ All the methods attempt to improve on the AAPCC by exploiting diagnostic information to predict expected spending. Thus, to continue the example used above, all the methods would predict that a 70-year-old, non-institutionalized woman with colon cancer would spend more than an otherwise similar woman without colon cancer and would pay more to a plan that enrolled such a woman. The

PIP-DCG approach differs from the HCC approach in that it utilizes diagnostic information on hospitalized patients only.

In addition to the Boston University and Johns Hopkins groups, Richard Kronick and others at the University of California at San Diego have developed a diagnosis-based risk adjuster for the disabled on Medicaid, the disability payment system (Kronick et al., 1996). Kronick's work regarding Medicaid recipients has produced results broadly similar to those of the diagnosis-based systems for Medicare.

The principal criterion for judging the value of new adjusters is the amount of variance in spending they are able to explain. That amount depends upon whether the methods are implemented prospectively or retrospectively. Prospective methods use only diagnostic information from a past period or periods, usually the previous year, then apply it to the year ahead; retrospective methods use diagnostic information from the current period, then apply it to the current year. Thus, if a person is diagnosed with diabetes mellitus in 1997, a prospective adjuster would use that information to set reimbursement in 1998, whereas a retrospective adjuster would use the information for setting 1997 reimbursement.

A retrospective adjuster will explain more variance in actual spending than a prospective adjuster because it explains some random variation or luck. Suppose, for example, that a man with no known diseases has a heart attack in the current year, a random event that could not be predicted in advance. The additional current year spending associated with the heart attack will be partly explained by a retrospective adjuster and will place the man in a higher category, reflecting the diagnosis of acute myocardial infarction. As pointed out above, explaining random variation is not important for purposes of reducing selection behavior.¹¹

The variance that can be explained by prospective rather than retrospective adjusters deserves attention for three reasons. First and most important, HCFA is proposing to implement a prospective adjuster, as described below. Second, there is no standard for how much variance a good retrospective adjuster should explain. Third, as just described, explaining random variation is not important to reducing selection behavior, and at least some of the additional variation explained by a retrospective adjuster is random. Unfortunately, the evidence is conflicting on how much is random.

Ellis et al. (1996) and Dunn et al. (1996) find that prospective and retrospective adjusters do about equally well in explaining next year's spending by this year's high and low spenders and also for persons with certain chronic diseases. Thus, their results imply that much of the gain in explained variation achieved by retrospective adjusters comes from explaining random spending. By contrast, Chapman (1996), using data from individuals under 65, finds that retrospective adjusters do detect some spending that appears predictable, suggesting that

the use of a retrospective adjuster could reduce the incentives to select. For example, a plan may know from the results of clinical tests that a person is at above-average risk for a heart attack within the group of people with identified heart disease; with a retrospective adjuster, the plan can recoup some of the expense of the heart attack if it occurs.

In terms of ability to explain variance prospectively using Medicare data, the HCC model appears to do better, explaining about 8 to 9 percent (Ellis et al., 1996).¹² By contrast, ADGs explain about 6 percent (Weiner et al., 1996), as do PIP-DCGs. Although these figures are still a fair distance from the 20-25 percent lower bound, each is clearly a substantial improvement on the AAPCC.

In addition to explained variance, both Ellis et al. and Weiner et al. (1996) present data on how these risk adjusters affect profitability, thereby suggesting both the improvement that they make and the magnitude of the remaining problem. For example, Ellis and colleagues calculate that plans would experience substantial average profits and losses in 1992 from enrolling those in the lowest and highest quintiles of 1991 expenditures, assuming that 1992 spending by the plan would be the same as spending in traditional Medicare (table 1).

Table 1. Effect of the AAPCC and HCC Adjusters on Average Estimated Profit and Loss

Subgroup	Profit or Loss in 1992: AAPCC Adjusters	Profit or Loss in 1992: HCC Adjusters
Top 20% of Spenders, 1991	-\$4,425	-\$1,311
Bottom 20% of Spenders, 1991	+\$2,134	+\$424

Source: Ellis et al., 1996. Assumes plan was paid 100% of the AAPCC.

As table 1 shows, the HCC method yields a noteworthy reduction in incentives to cream and dump, cutting profits and losses by more than a factor of three relative to the AAPCC. Nonetheless, HCCs leave nontrivial levels of profits and, especially, losses.¹³ (For purposes of scaling these figures, mean spending in 1992 was \$3,800.) Whether the changes in profitability brought about by HCCs or ADGs would reduce selection behavior to negligible levels is an empirical question whose answer cannot be known before the methods are implemented. Unfortunately, implementing HCCs poses serious problems, as described below.

Accounting for Information on Functional Status and Self-Assessed Health

Another strand of risk adjustment literature uses functional status or variants of self-reported health status such as the RAND-36 instead of or in addition to diagnostic information. In general, self-reported health status by itself does not explain as much variation as do the diagnostic-based measures. When used in conjunction with diagnostic measures, self-reported

health status appears to add around one percentage point of explained variation (Newhouse et al., 1989; Fowles et al., 1994; Gruenberg et al., 1996, 1999; Pope et al., 1999).

Although the increment of improvement gained by including self-assessed health status is modest overall, functional status appears to be an important predictor of spending by the frail elderly, those impaired in activities of daily living (ADLs) or instrumental activities of daily living (IADLs) (Gruenberg et al., 1999; Pope et al., 1999).¹⁴ In particular, the diagnostic-based measures underpredict spending for the frail elderly and correspondingly overpredict for elderly who are not frail. In other words, plans are underpaid for the frail elderly and overpaid for the nonfrail elderly.

The Medicare Payment Advisory Commission (MedPAC) (1999) shows that differences in spending by frailty status are large. Using 1995 data from the Medicare Current Beneficiary Survey, MedPAC estimates that the average spending for a frail elderly person in the community was more than three times as great as for the average beneficiary and that average spending for nursing home residents was almost twice as high (table 2). (A frail person in the community was someone age 55 or more who required hands-on assistance with three out of five ADLs or one ADL and four out of five IADLs.¹⁵) Spending in all categories was notably higher for the frail group, but home health spending was particularly higher.

Table 2. Medicare Fee-for-Service Spending, by Beneficiary Frailty and Residence, 1995

Provider Type	All Fee-for-Service	Frail in Community	Nursing Home
PPS Hospital	\$1,720	\$ 5,035	\$3,324
Physician	1,092	1,879	1,793
Home Health	472	3,658	380
Outpatient Dept	377	572	1,152
SNF	201	818	1,375
DME	137	471	464
Rehab Facility	110	515	47
Other Hospital	87	236	403
Hospice	19	162	163
Total	\$4,215	\$13,346	\$9,102

Source: Medicare Current Beneficiary Survey, Cost and Use File 1995, as reported in Medicare Payment Advisory Commission (1999). Frail in the community are those 55 or over who require hands-on assistance with three of five ADLs or one ADL and four of five IADLs.

Adding HCCs and an adjustment for functional status appears to improve payment accuracy considerably for subgroups with ADL limitations (table 3). Table 3 shows how well various adjusters perform for subgroups that differ in functional status. In particular, it shows

the extent to which 1992 spending can be predicted based on 1991 functional status or 1993 spending based on 1992 status (Pope et al., 1999). A value of less than 1 means payment was less than actual spending, while a value of more than 1 indicates that payment exceeded actual spending. The AAPCC method dramatically overpays for elderly with no functional status limitations (by 39 percent) and underpays for the other groups. HCCs substantially reduce overpayment and underpayment. Adding several survey-based measures of self-reported health status reduces the discrepancies even further. Unfortunately, there are also serious problems with implementing measures of functional status and self-rated health status, as discussed below.

Table 3. Effect of Survey-Based Risk Measures on the Ratio of Average Payment to Actual Spending, by Functional Status

Functional Status	AAPCC	HCC	HCC+Survey
5–6 ADLs	0.64	0.88	1.09
3–4 ADLs	0.71	0.86	0.96
1–2 ADLs	0.85	0.93	1.06
No ADLs	1.39	1.16	0.95

Source: Pope et al., 1999. Survey measures include self-reported chronic conditions, functional status, and self-rated health status. Data based on the 1991–93 Medicare Current Beneficiary Survey.

Gruenberg et al. (1999) find similar results for the dually eligible (Medicare and Medicaid) population if prospective adjusters are used. A problem occurs, however, because the rate is set using last year’s information, although spending for new entrants to nursing homes is very high—5.7 times the Medicare average—and HCCs and PIP-DCGs do not begin to compensate for the difference. As a result, payment for the dually eligible is too low even after adjustment with the diagnosis-based methods. The amount of the shortfall varies between 7 percent and 24 percent for HCCs and 20 percent and 33 percent for PIP-DCGs, depending on the data set used in making the calculations.¹⁶

Rethinking Assumptions about Incentives to Overserve

Risk adjustment as usually discussed involves a lump sum payment to a health plan or a provider for a given patient’s care. There seems to be a taboo against paying a health plan anything for providing more services. Indeed, the language of the BBA appears to be very restrictive with respect to such payment.¹⁷

The philosophy of paying nothing for doing more has carried into the construction of the PIP-DCGs and HCCs; they have been deliberately chosen to minimize the effect of reimbursement method on choices regarding the treatment of a particular patient. The creators of the HCC method have taken special care to limit higher payments for diagnoses for which treatment is relatively nondiscretionary. But a higher amount of variance in expected cost

could be explained—and selection incentives correspondingly reduced—if categories were created that paid more for cases in which discretionary treatment was rendered.

Minimizing incentives for discretionary treatment at the possible expense of increasing incentives to select—the choice the creators of the HCCs made—follows a dominant view in the literature that payment to health plans should be independent of actual use, or at least independent of any use that is discretionary. This view reflects an implicit assumption that, given a choice of treatment, the less costly method is preferred. There are, however, important objections to this view.

First and most important, if reimbursement is independent of use, additional revenue from doing more is by definition zero. But if excess services appear to be related to overly generous fees, so does underservice appear to be related to no payment (Pauly, 1980).¹⁸ In short, even apart from issues of selection, the policy of making no payment for additional services raises the issue of possible stinting. For example, Medicare's Prospective Payment System pays more for a patient with heart disease who has a coronary bypass operation than for a similar patient who does not have an operation. Although there is often an element of discretion in having a bypass operation, the planners who created the DRGs presumably deemed that paying nothing more for cases in which a bypass operation was performed would likely lead to too few bypass operations. Although excessive fees can potentially induce unwanted or unnecessary services, the conclusion that marginal revenue should always be zero seems to be an overreaction to the excesses of reimbursement in the indemnity insurance system.¹⁹

Second, a substantial amount of medicine involves learning by doing. When first introduced, open heart surgery took much more time and yielded much poorer results than it does today; over time, however, cardiac surgeons became better at performing the operation. If the lowest cost option is always chosen, the possibility that a more expensive option today will become cheaper tomorrow is lost. A third and related point is that, even putting aside learning by doing, there may be substantial uncertainty about which course of treatment is better. Without some variation in treatment, we will not learn.²⁰

Thus, as described in greater detail below, it seems wise to base some portion of reimbursement to a health plan on actual use. Such a policy could potentially reduce selection even further, while also curbing possible stinting, although overservice could result if payments became too high. Another possible model would use a much finer classification of diagnoses to reimburse plans than the DCGs or HCCs now employ, especially for high cost diagnoses. This approach is desirable in principle, but obtaining stable and appropriate relative weights for more numerous categories is problematic.

POTENTIAL PROBLEMS IN IMPLEMENTING BETTER RISK ADJUSTERS

The Medicare program is poised to proceed with the implementation of diagnostic-based risk adjustment in January 2000, despite one major implementation problem: the absence of reliable diagnosis information on outpatient claims. HCFA therefore proposes that next January's implementation use the PIP-DCG method, which takes account of inpatient diagnostic information only, rather than the HCC method, which is substantially more accurate but requires outpatient diagnostic information. PIP-DCGs explain somewhat under 6 percent of the variance in spending, whereas HCCs explain 8 to 9 percent (Ellis et al., 1996).

Moreover, HCFA proposes to implement the PIP-DCG system gradually, giving health plans time to adjust. Only 10 percent of the payments is to be based on the PIP-DCG adjusted payment in 2000, with the remaining 90 percent based on the current AAPCC method. The percentage based on the PIP-DCG method would steadily increase, reaching 30 percent in 2001, 55 percent in 2002, and 80 percent in 2003. In 2004, payment would be entirely based on diagnostic-based methods, and at that time HCFA would move from the PIP-DCG method and introduce outpatient diagnoses into the adjustment method.

The Unreliability of Outpatient Diagnostic Data

An analysis by MedPAC (1998) amply demonstrates the inaccuracy of current outpatient diagnosis coding. For the purposes of the study, MedPAC selected beneficiaries whose Medicare Part B claims in 1994 showed a diagnosis of one of 11 serious diseases, then checked for claims for the same diagnosis in 1995. As shown in table 4, the likelihood of a claim in 1995 was only about 50-60 percent for each of the 11 diagnoses. One would not expect the likelihood to be 100 percent. A few patients may have been misdiagnosed in 1994, and a few more may have had no physician contact in 1995. It seems likely, however, that most of the 1995 shortfall is attributable to physicians simply neglecting to fill in the diagnosis on the Part B claim or recording some other diagnosis. After all, no payment turned on filling in the diagnosis. Two consequences follow.

Table 4. Persistence in Diagnostic Coding

Diagnosis on 1994 Part B Claim	Of Those Identified in 1994, Percent with Part B Claim in 1995
Hypertension	59%
Coronary Artery Disease	53
COPD	62
Congestive Heart Failure	61
Stroke	51
Dementia	59
Rheumatoid Arthritis	55

High Cost Diabetes	58
Renal Failure	56
Quadriplegia/Paraplegia	52
Dialysis	59

Source: Medicare Payment Advisory Commission, March 1998, p. 17. Excludes those who died in 1994 or 1995.

First, introducing payment based on outpatient diagnosis offers enormous potential for “upcoding”—the equivalent of DRG creep in hospital payment—and resulting increases in overall spending. Because the increment in payment can be several thousand dollars per year depending on the exact diagnosis, there is a powerful incentive to be complete with coding. Increasing the completeness of coding would inappropriately raise payment for the same mix of patients. When faced with similar problems in the past, HCFA has adjusted rates down before the fact on the assumption that more complete coding would occur. For example, when HCFA introduced the Resource Based Relative Value Scale (RBRVS) for physician services, it lowered fees by a few percent before the fact to compensate for anticipated volume changes. Even this level of adjustment caused considerable controversy. It is inconceivable that HCFA could lower payments to plans by the amount suggested by the MedPAC study—say 40 percent. On the other hand, if payments were left unchanged, overall spending would probably increase greatly.

As a result, it seems likely that payment based on outpatient diagnoses will have to be phased in, just as the PIP-DCG system is being phased in. For example, if only 10 percent of the payment were based on outpatient diagnoses in the first year, plans would still have an incentive to code much more completely, but the expenditure risk to the government would be limited. Thus, the transition to fully incorporate outpatient diagnoses may extend well beyond 2004.

A second, more subtle consequence of incomplete coding is that weights in the new system may be seriously in error. The weights determine how much more is paid for a patient with a particular diagnosis, analogous to the weight for each DRG. The HCC weights have necessarily been calculated using existing Part A and Part B data and are approximately proportional to mean Medicare expenditure for patients with given diagnoses. Thus, weights will be relatively more skewed for diagnoses in which a higher proportion of the bill is for Part B spending and for diagnoses that are less likely to be reported.

The Problem of Relying on Inpatient Information

Exclusive reliance on the PIP-DCGs and inpatient diagnosis information raises the issue that plans may find it profitable to hospitalize inappropriately. For example, a plan would receive additional reimbursement for a cancer patient only if the patient received inpatient treatment. HCFA is aware of this problem and has taken several steps to minimize it:

- The increment in payment will be based on average spending in the year following the hospitalization; for example, plans enrolling patients hospitalized for heart attacks will receive a higher reimbursement, but the increment will be based on spending in the year following the heart attack, not the year of the heart attack. As a result, spending for the initial hospitalization will not figure into the incremental amount paid.
- Causes of admission that are discretionary or vague will not trigger additional payment; about 40 percent of the admissions fall into this category.
- One-day stays will not generate additional payment.

Although these steps reduce the incentive to hospitalize for financial reasons, they also blunt the gain from risk adjustment; that is, they do not diminish incentives to select good risks and avoid bad risks as much as HCCs. HCFA estimates that the amount of explained variance from the PIP-DCG method will be about 5.8 percent.²¹

The Shortcomings of Functional Status and Self-Ratings

HCFA is not proposing to implement adjusters based on functional status or self-ratings of health status, which seemed plagued by even more serious implementation issues than the diagnosis-based methods. The problems turn on the feasibility and expense of obtaining reliable information when it will be used for payment purposes.

First, many of those who are functionally impaired are also cognitively impaired or demented, which makes filling out a self-administered form problematic. This is especially true for the frail elderly, where use of self-rated measures seems most important for accurate adjustment. Although another party could in principle rate functional status, it is harder to see how a proxy respondent could be used for self-rated health status. Second, if a third party rater is not used, nonresponse may bias the results. Third, in the case of self-rated health status, providers may “prescribe disability,” leading beneficiaries to report lower health status, or beneficiaries may respond strategically if they realize provider reimbursement depends upon their survey answers. Fourth, self-rated health status, as distinct from functional status, cannot be audited. Fifth, it is not clear how frequently surveys would be carried out, but, if less than annually, results would not be as favorable as those reported above because some beneficiaries would have moved to more impaired categories but reimbursement would not adjust. Sixth, the expense of collecting such information would clearly be substantial, especially if a third party were used. There would have to be an adequate sample of enrollees in each plan; indeed, some might argue that 100 percent samples would be needed in order to set adjustment for those transferring from one plan to another. These considerations make clear that introducing functional status into the risk adjustment system is a daunting task.

Beyond these practical problems are important questions of the incentives that might be produced. Would increased payment for poorer functional status induce worse performance?²² Or perhaps less aggressive efforts at rehabilitation? To what degree would increased payment for entrants to nursing homes induce more nursing home admissions?

CONTENDING WITH THE REALITY OF INADEQUATE RISK ADJUSTMENT

Risk adjustment is not likely to be adequate for many years for three reasons:

- HCFA's current plans do not call for using outpatient diagnostic information until 2004. In fact, full implementation of an outpatient-based method is likely to be even further off because of the potential for upcoding.
- There are no current plans to measure functional status using survey-based risk adjusters. Indeed, it is not clear it would ever be practical to use them for payment purposes.
- Even full implementation of both diagnostic- and survey-based methods is unlikely to reduce selection to negligible levels because the amount of explained variance falls far short of what an ideal adjuster should predict.

Given this reality, several alternatives or adjustments to HCFA's plans deserve attention.

Partial Capitation

Newhouse has elsewhere proposed the adoption of partial capitation, or a blend of adjusted capitated payment and payment under the traditional program (Newhouse, 1994, 1996, 1998). Specifically, for each enrollee, a plan would receive part (perhaps 75 percent) of its payment under a risk-adjusted capitation formula and part (perhaps 25 percent) based on the amount paid through traditional Medicare. This could be done in a budget neutral fashion, and the proper mix could be developed through experimentation.²³ This method would require collecting and coding encounter data for each person, but these tasks are already required by the Balanced Budget Act in order to introduce outpatient-based diagnosis as a risk adjuster.

Under this method, plans would be paid more for enrolling those who use more services and less for those using fewer. Because sicker patients use more services, the losses plans incur from enrolling them would be cut, as would profits from enrolling healthy persons. This would reduce the incentive to cream and dump.²⁴

Outlier Systems

An alternative plan with somewhat similar properties is an outlier or reinsurance scheme. Under this method, payments for additional use would be made only after some threshold spending level had been exceeded, analogous to the outlier payments in the Prospective Payment System. Such a system would target payment toward high spenders, the individuals for whom selection incentives appear to have the worst effect.

Relative to a blend of capitation and traditional Medicare, however, a reinsurance or outlier proposal has three important drawbacks. First, reinsurance for large spenders does nothing about a plan's incentive to spend resources to attract good risks, whereas a blend would reduce payments for a good risk. Second, the ideally designed outlier policy still leaves plans incurring a loss on high cost cases, thereby making such cases unattractive, whereas the ideally designed blend does not. Third, for all non-outlier cases, there is an incentive to stint. Moreover, the apparent advantages of an outlier scheme may not be real. The additional payments would not necessarily be generated by the cases with the largest predictable losses; to the degree that outlier spending arises from random events, there is no selection problem for those cases.

Whether additional payment for additional services should be limited to high-cost cases or should apply to all cases depends on how serious the problem of selection is under each method and how much an actual, as opposed to an ideal, fee-for-service payment might degrade efficient production. Some support for an outlier arrangement, as opposed to a blend, seems to arise from the notion that additional payments for additional services should be minimized, which, as explained above, is too facile a view.

A High-Cost Case Pool

A different, not mutually exclusive reinsurance method, proposed by van Barneveld et al. (1996), would improve targeting by making additional payment to cases that the plan had predicted in advance would incur high costs. Made in the Dutch context, the proposal would have plans cede a certain portion of their cases, specifically those they predict will be the most costly, to a pool. Although the percentage to be ceded is a policy choice, the authors illustrate the likely results of using 2 percent and 4 percent of cases in each plan.

In the Medicare context, there is no common pool to which to cede cases. Even so, one could adapt the idea by stipulating that a plan could continue to treat an enrollee but, instead of receiving a capitated payment, would be reimbursed partly or entirely under traditional Medicare for some prearranged fraction of its enrollees. This transaction could be kept invisible to the enrollee and conducted solely between Medicare and the plan. Van Barneveld and colleagues estimate that, if 4 percent of the Dutch cases were ceded in advance, about 25 percent of the spending in any given year would be covered by the pool, a fraction

that would probably be roughly similar in Medicare.²⁵ In short, plans could effectively place their costliest cases in traditional Medicare, thereby receiving additional reimbursement and mitigating the financial hit from those cases.

It is not hard to see administrative concerns with this proposal. For example, a plan could not cede cases until after a person had enrolled but would obtain information about that person's current use only over time. Nevertheless, the proposal may be useful for dealing with new nursing home entrants and terminally ill patients.

Special Payments for Terminally Ill Patients

It is well known that those who die in any year are disproportionately costly (Lubitz and Riley, 1993). Although the figures vary slightly from year to year, the 5 percent of Medicare beneficiaries who die each year account for 25 to 30 percent of dollars spent. There are thus strong incentives for health plans to avoid the terminally ill and leave this population in traditional Medicare. The distribution of deaths seems consistent with these incentives; as already noted, the death rate for new enrollees in the at-risk program during the first six months is 75 percent of the death rate in the traditional program, after adjusting for age and sex. To the degree that terminally ill patients are disproportionately covered under traditional Medicare, the opportunity for managed care to increase coordination and cut waste in their care is not being fully exploited.

Insofar as better risk adjusters would pay additional amounts for enrollees with diagnoses that carry a high risk of mortality, this problem would be mitigated. For reasons of appearance, however, one would not want to pay more simply because an enrollee died, just as the Prospective Payment System does not pay more for patients who die in the hospital.²⁶ If a plan could cede certain individuals back to traditional Medicare, however, these might well be those identified as terminally ill, a solution that would avoid the appearances problem. If the plan remained partially at risk for patients it ceded, Medicare expenses could potentially be reduced. Even without the blend, however, a managed care plan might be more efficient for terminally ill patients because it might use similar treatment methods for its Medicare and commercial business.

Van Vliet and Lamers (1998) have shown that demographic information, prior spending, and inpatient diagnostic information are not good predictors of those who die, in that expenditures of Dutch decedents were 250 percent above what the model would predict. They conclude that "most of the excess costs were unpredictable." That they were not predictable for the analyst, however, does not mean that they were not predictable for the plan or the physician. Particularly in the case of cancer, physicians may be able to observe a terminal phase that could last several weeks or months (Medicare Payment Advisory Commission, June

1998). Indeed, eligibility for hospice services is premised on the assumption that such a terminal phase can be identified for many diseases.

CONCLUDING COMMENT

As the Medicare at-risk program grows, Congress has recognized that lack of satisfactory risk adjustment is causing serious problems. Better risk adjusters that improve on the AAPCC exist, and HCFA is proposing to introduce them into the program. Even when fully implemented, however, these risk adjusters are not likely to reduce current selection problems to negligible levels. The Medicare program could do better by combining these methods with partial capitation; that is, making some part of the payment to health plans depend on the services they deliver.

Partial capitation would reduce plans' incentives to select good risks—the intent of risk adjustment—and also reduce the financial incentive to underserve or stint. By reducing incentives to underserve, the payment method would reduce the burden that any patient protection legislation would need to carry.

REFERENCES

- Ash, Arlene, Frank Porell, Leonard Gruenberg, et al. 1989. "Adjusting Medicare Capitation Payments Using Prior Hospitalization Data." *Health Care Financing Review* 10(4):17–29.
- Chapman, John. October 1996. "Inter-temporal Patterns of Health Care Expense and the Potential for Biased Enrollment in a Non-Elderly Population." Mimeo.
- Chapman, John. November 1996. "Time Perspective and the Effectiveness of Risk-Adjusted Payment to Health Plans." Mimeo.
- Congressional Budget Office. March 1997. "Predicting How Changes in Medicare's Payment Rates Would Affect Risk Sector Enrollment and Costs."
- Cutler, David M., and Sarah Reber. 1998. "Paying for Health Insurance: The Tradeoff between Competition and Adverse Selection." *Quarterly Journal of Economics* 113(2):433–66.
- Cutler, David M., Mark McClellan, and Joseph P. Newhouse. 1997. "Prices, Output, and Productivity in Managed Care Insurance." Mimeo.
- Dunn, Daniel L., Alice Rosenblatt, Deborah A. Taira, et al. 1996. "A Comparative Analysis of Methods of Health Risk Assessment." SOA monograph M-HB96-1. Schaumburg, IL: Society of Actuaries.
- Ellis, Randall P., and Arlene Ash. 1995/96. "Refinements to the Diagnostic Cost Group (DCG) Model." *Inquiry* 32 (Winter):418–29.
- Ellis, Randall P., Gregory C. Pope, Lisa I. Iezzoni, et al. 1996. "Diagnosis-Based Risk Adjustment for Medicare Capitation Payments." *Health Care Financing Review* 17(3):101–28.
- Fowles, Jinnet, Jonathan Weiner, David Knutson, et al. 1994. "A Comparison of Alternative Approaches to Risk Measurement: Final Report to the Physician Payment Review Commission." Selected external research series, number 1. Physician Payment Review Commission.
- Frank, Richard G., Ernst Berndt, and Susan Busch. In press. "Price Indices for Depression," in *Measuring the Price of Medical Treatments*, ed. Jack Triplett. Washington, D.C.: Brookings Press.
- Gruenberg, Leonard, Eugenia Kaganova, and Mark C. Hornbrook. 1996. "Improving the AAPCC with Health-Status Measures from the MCBS." *Health Care Financing Review* 17(3):59–75.
- Gruenberg, Leonard, Ajith Silva, Kirsten N. Corrazini, and Joelyn K. Malone. 1999. "An Examination of the Impact of the Proposed New Medicare Capitation Methods on Programs for the Frail Elderly." Cambridge, MA: Long-Term Care Data Institute.

- Keeler, Emmett B., Grace Carter, and Joseph P. Newhouse. 1998. "A Model of the Impact of Reimbursement Schemes on Health Plan Choice," *Journal of Health Economics* 17(3):297–320.
- Kronick, Richard, Tony Dreyfus, Lora Lee, and Zhiyuan Zhou. 1996. "Diagnostic Risk Adjustment for Medicaid: The Disability Payment System," *Health Care Financing Review* 17(3):7–33.
- Lubitz, James D., and Gerald F. Riley. 1993. "Trends in Medicare Payments in the Last Year of Life," *New England Journal of Medicine* 328:1092–96.
- Medicare Payment Advisory Commission. March 1998. "Report to the Congress: Medicare Payment Policy," volume 2.
- Medicare Payment Advisory Commission. June 1998. "Report to the Congress: Context for a Changing Medicare Program."
- Medicare Payment Advisory Commission. June 1999. "Report to the Congress."
- Newhouse, Joseph P. 1994. "Patients at Risk: Health Reform and Risk Adjustment." *Health Affairs* 13(1):132–46.
- Newhouse, Joseph P. 1996. "Reimbursing Health Plans and Health Providers: Selection versus Efficiency in Production." *Journal of Economic Literature* 34:1236–63.
- Newhouse, Joseph P. 1998. "Risk Adjustment: Where Are We Now?" *Inquiry* 35:122–31.
- Newhouse, Joseph P., and the Insurance Experiment Group. 1993. *Free for All? Lessons from the RAND Health Insurance Experiment*. Cambridge: Harvard University Press.
- Newhouse, Joseph P., Willard G. Manning, Emmett B. Keeler, et al. 1989. "Adjusting Capitation Rates Using Objective Health Measures and Prior Utilization," *Health Care Financing Review* 10(3):41–54.
- Pauly, Mark V. 1980. *Doctors and Their Workshops*. Chicago: University of Chicago Press.
- Physician Payment Review Commission. 1996. Annual Report to Congress.
- Pope, Greg C., Killard W. Adamache, Raul K. Khandker, and E.G. Walsh. 1999. "Evaluating Alternative Risk Adjusters for Medicare," *Health Care Financing Review* (in press).
- Riley, Gerald, Cynthia Tudor, Yen-pin Chiang, and Melvin Ingber. "Health Status of Medicare Enrollees in HMOs and Fee-for-service in 1994," *Health Care Financing Review* 17(4):65-76.
- Van Barneveld, Erik M., Rene C.J.A. van Vliet, and Wynand P.M.M. van de Ven. 1996. "Mandatory High-Risk Pooling: A Means for Reducing the Incentives for Cream Skimming." *Inquiry* 33:133–43.

- Van de Ven, Wynand, and Randall P. Ellis. In press. "Risk Adjustment in Competitive Health Plan Markets," in *Handbook of Health Economics*, eds. Anthony J. Culyer and Joseph P. Newhouse. Amsterdam: North Holland.
- Weiner, Jonathan, Allen Dobson, Stephanie L. Maxwell, et al. 1996. "Risk-Adjusted Medicare Capitation Rates Using Ambulatory and Inpatient Diagnoses," *Health Care Financing Review* 17(3):77–99.
- Welch, W. P. 1985. "Medicare Capitation Payments to HMOs in Light of Regression to the Mean in Health Care Costs," in *Advances in Health Economics and Health Services Research*, vol. 6, eds. Richard Scheffler and Louis Rossiter. Greenwich, CT: JAI Press.
- Winston, Clifford. 1993. "Economic Deregulation: Days of Reckoning for Microeconomists," *Journal of Economic Literature* 31(3):1263–89.

NOTES

¹ Dunn et al. (1996) find ADGs better than DCGs using under-65 data but do not estimate results for HCCs.

² If payment were 95 percent of the AAPCC, profits on the low-spending group would be less and losses on the high-spending group would be greater. The asymmetry between the profit from creaming and the loss from not dumping reflects the skewed distribution of spending, which is not mirrored in the reimbursement formula.

³ In the commercial market, plans with disproportionate numbers of bad risks tend to charge higher premiums, thereby leading to a potential death spiral (Cutler and Reber, 1998). In the Medicare program, charging enrollees more is possible only to a limited degree. To the degree it is possible, it lessens beneficiaries' ability to insure against the risk of chronic illness because those with chronic illnesses will in effect pay more.

⁴ Airlines before deregulation were in a similar position: fare regulations made long distance routes profitable, while short haul routes, typically between small towns, were unprofitable, but regulators required that airlines serving long distance routes also serve short haul routes. Because the profitable but regulated fares on long distance routes could not be bid down, airlines serving such routes found other ways to compete; notably, they scheduled frequent flights, with the result that load factors on long distance routes were relatively low (Winston, 1993). But half empty planes are an inefficient use of the profit in the fare structure; although an empty adjacent seat may be of some value to the passenger, a reduction in the fares would presumably have been valued more highly, at least by passengers paying their own fares.

⁵ Plans must be reimbursed at least the average amount of such random expense, or they will not choose to participate.

⁶ The proportion of variance explained has become a principal measure of the adequacy of risk adjustment measures. Some literature explains the variance in spending across groups (actual or hypothetical) rather than individuals because averaging random events within the group allows a higher fraction of the spending variation to be explained. In our view, however, group spending is relevant only if the group as a whole is enrolling in a plan (as when employer enrolls all employees in a single group). If individuals are free to choose plans, as they are in Medicare, the variance across individuals is relevant because the group is not fixed.

⁷ We do know that, as explained variation improves, incentives to select do not diminish proportionately (Newhouse et al., 1989, 1996).

⁸ Cost, however, is much lower in the HMOs because prices paid for treatment are lower.

⁹ In 2003 and after, beneficiaries will generally be "locked in" to a plan choice, but may disenroll during the first three months. (In 2002 they may disenroll during the first six months.)

¹⁰ The Ambulatory Care Groups were originally designed as a case-mix measure for non-aged populations. The ACG classification methodology was subsequently augmented with

inpatient ICD-9-CM diagnosis information to form the ADG model risk adjustment model for Medicare beneficiaries.

¹¹ Explaining random variation could be important if plans face a nontrivial risk of bankruptcy from random health events, but this risk should be unimportant for plans with tens of thousands of enrollees.

¹² Dunn et al. (1996) find ADGs better than DCGs using under-65 data but do not estimate results for HCCs.

¹³ If payment were 95 percent of the AAPCC, profits on the low-spending group would be less and losses on the high-spending group would be greater. The asymmetry between the profit from creaming and the loss from not dumping reflects the skewed distribution of spending, which is not mirrored in the reimbursement formula.

¹⁴ Even though the adjustment is quantitatively important for the frail elderly, it does not much affect explained variation because the frail subgroup is relatively small and hence its spending does not contribute much to total variation. This illustrates one weakness of simply using explained variation in comparing adjustment methods.

¹⁵ The six ADLs include: transferring into and out of chairs; dressing; bathing; walking; toileting; eating; and maintaining continence. The five IADLs include housework; meal preparation; managing finances; using the telephone; and shopping for personal items. Hands-on assistance means the person has a "need for assistance" with the ADL rather than simply "has difficulty." The group that has a need for assistance has both much higher Medicare costs and a much greater discrepancy with predicted costs relative to the group that simply reports disability (Gruenberg et al., 1999).

¹⁶ Gruenberg et al., tables 4-10 and 4-12.

¹⁷ Section 1855(b) of the 1997 Balanced Budget Act: "Assumption of Full Financial Risk. The Medicare+Choice organization shall assume full financial risk on a prospective basis for the provision of health care services...except that the organization: 1) may obtain insurance...for the cost of...such services the aggregate value of which exceeds such aggregate level as the Secretary specifies from time to time...3) may obtain insurance...for not more than 90 percent of the amount by which its costs...exceed 115 percent of its income for such fiscal year." In other words, plan payment should be fully prospective except for outlier payments above a HCFA-determined amount and reinsurance for costs exceeding 115 percent of revenue should the plan wish to purchase such reinsurance. Similar language about payment to health plans being fully prospective appeared in the Clinton Health Security Act.

¹⁸ One will not observe underservice if the consumer can monitor the actual delivery of services that should be in the bundle. But information is notoriously asymmetric between providers and patients. Compounding the problem is that a nontrivial percentage of the Medicare population is demented.

¹⁹ The ideal payment would be at the marginal cost that would induce the optimal quantity (Pauly, 1980). Combined with ideal capitation, this would address concerns of both selection and stinting. In practice, we do not know marginal cost and thus will not have this

fee structure. The argument about the tradeoff between selection and efficient production thus reduces to the system that in practice is likely to be closer to the ideal. A different argument is that a manager who keeps the entire dollar from any savings puts the optimal amount of effort into efficient production (Newhouse, 1996). This, however, is an argument against cost-based reimbursement; it does not argue against paying plans a set rate for additional procedures. Related to this point, patients may prefer a higher-cost discretionary option and may be willing to pay an extra premium for it. Willingness to pay is a standard economic test for efficiency.

²⁰ Where possible, this variation should take place in the context of a randomized trial, but for ethical and other reasons a randomized trial is not always possible. It is also true that under fee-for-service there can be too little incentive to find cost-reducing technologies.

²¹ Daniel Zabinski, personal communication.

²² There is a similar problem in principle with respect to prevention for the diagnosis-based information, but it seems much less serious. That is, it does not seem likely that plans would markedly reduce effective preventive efforts for, say, cancer, simply because their payment would not be as great. And of course expenses would rise to treat sick patients, so it is not clear that the plan is any more profitable from failure to prevent.

²³ See Keeler, Carter, and Newhouse (1998) for some evidence that a 75-25 mix may be about right.

²⁴ In terms of the R^2 criterion used above, the portion of payment based on traditional Medicare would exactly match the dependent variable; thus, the overall proportion of explained variation under this system would be $wR_1^2 + (1-w)$, where w is the weight on the capitation method and R_1^2 is the proportion of variance explained by the capitation method. This number, however, cannot be compared with the 20 to 25 percent criterion for a fully prospective adjuster because this method is not fully prospective.

²⁵ Ex post about 1 percent of cases account for a quarter of spending, but ex ante the percentage accounting for a quarter of cases will be larger because of the inability to predict random events.

²⁶ Acute myocardial infarction (AMI) is the one exception, because those dying from AMI are on average cheaper than those who survive.